sensors

Special Issue Reprint

Advanced Sensors for Intelligent Robotic Systems

Vision, Touch, and Dexterous Manipulation

Edited by
Chongkun Xia, Xueqian Wang and Yajing Shen

mdpi.com/journal/sensors

MDPI

Advanced Sensors for Intelligent Robotic Systems: Vision, Touch, and Dexterous Manipulation

Advanced Sensors for Intelligent Robotic Systems: Vision, Touch, and Dexterous Manipulation

Guest Editors

Chongkun Xia
Xueqian Wang
Yajing Shen

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editors

Chongkun Xia
School of Advanced Manufacturing
Sun Yat-Sen University
Shenzhen
China

Xueqian Wang
Shenzhen Graduate School of Tsinghua University
Shenzhen
China

Yajing Shen
Department of Electronic and Computer Engineering
Hong Kong University of Science and Technology
Hong Kong
China

Editorial Office

MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Sensors* (ISSN 1424-8220), freely accessible at: https://www.mdpi.com/journal/sensors/special_issues/T4SV8A8961.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-9223-5 (Hbk)
ISBN 978-3-7258-9224-2 (PDF)
https://doi.org/10.3390/books978-3-7258-9224-2

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

Preface . . . . . **vii**

Yanyan Wu, Xueqian Wang, Yajing Shen and Chongkun Xia
Advanced Sensors for Intelligent Robotic Systems: Vision, Touch, and Dexterous Manipulation
Reprinted from: *Sensors* **2026**, *26*, 2626, https://doi.org/10.3390/s26092626 . . . . . **1**

Yanyan Wu, Hanhan Wu, Yifei Han, Yi Ding, Bosheng Cao and Chongkun Xia
High-Resolution Contact Localization and Three-Axis Force Estimation with a Sparse Strain-Node Tactile Interface Device
Reprinted from: *Sensors* **2026**, *26*, 1378, https://doi.org/10.3390/s26041378 . . . . . **5**

Zhifu Xue, Zhiquan Yang, Junyi Hu, Bin Zhu and Jianqing Peng
A Cable-Driven Hybrid Robot with Series-Parallel Coupling: Design, Modeling, Optimization Analysis, and Trajectory Tracking
Reprinted from: *Sensors* **2026**, *26*, 1147, https://doi.org/10.3390/s26041147 . . . . . **24**

Tongmu Liu, Baocheng Zhou, Xinwen Zhang, Jianqing Peng, Hua Huang and Tianhao Jian
Real-Time Deep-Sea Mooring System with Inductive Telemetry and Multi-Sensor Integration: Deployment and Performance in the South China Sea
Reprinted from: *Sensors* **2026**, *26*, 937, https://doi.org/10.3390/s26030937 . . . . . **63**

Czaryn Diane Salazar Ompico, Julius Noel Banayo, Yamato Mashio, Masato Odagaki, Yutaka Kikuchi, Armyn Chang Sy and Hirofumi Kurosaki
Development of a Robot-Assisted TMS Localization System Using Dual Capacitive Sensors for Coil Tilt Detection
Reprinted from: *Sensors* **2026**, *26*, 693, https://doi.org/10.3390/s26020693 . . . . . **79**

Bruno J. Guerreiro, Francisco Azevedo, Paulo Oliveira and Rita Cunha
Automatic Grasping System and Hybrid Controller Towards Multi-Drone Parcel Delivery
Reprinted from: *Sensors* **2026**, *26*, 653, https://doi.org/10.3390/s26020653 . . . . . **97**

Zhonghua Hu, Jinlong Yang, Wenfu Xu, Hengtai Chen, Longkun Xu and Deshan Meng
Pose Stabilization Control for Base of Combined System Using Feedforward Compensation PD Control During Target Satellite Transposition
Reprinted from: *Sensors* **2026**, *26*, 206, https://doi.org/10.3390/s26010206 . . . . . **120**

Zhonghua Hu, Chaowen Deng, Kai Wang and Jianqing Peng
Kinematic Modeling and Solutions for Cable-Driven Parallel Robots Considering Adaptive Pulley Kinematics
Reprinted from: *Sensors* **2026**, *26*, 39, https://doi.org/10.3390/s26010039 . . . . . **150**

Bo Shan, Donghui Zhao, Ruijin Zhao and Yokoi Hiroshi
Real-Time 6D Pose Estimation and Multi-Target Tracking for Low-Cost Multi-Robot System
Reprinted from: *Sensors* **2025**, *25*, 7130, https://doi.org/10.3390/s25237130 . . . . . **175**

Diyun Wen, Jiangtao Xiao, Yu Xie, Tao Luo, Jinhui Zhang and Wei Zhou
A Universal Tool Interaction Force Estimation Approach for Robotic Tool Manipulation
Reprinted from: *Sensors* **2025**, *25*, 6619, https://doi.org/10.3390/s25216619 . . . . . **200**

Jianing Ding, Fuming Qu, Weihua Zhou, Jiajun Xu, Lingyu Zhao and Yaming Ji
A Sensor Based Waste Rock Detection Method in Copper Mining Under Low Light Environment
Reprinted from: *Sensors* **2025**, *25*, 5961, https://doi.org/10.3390/s25195961 . . . . . **216**

Zirun Liu, Shijie Jiang, Shuhan Feng, Qirui Song and Ji Zhang
Comprehensive Review of Deep Learning Approaches for Single-Image Super-Resolution
Reprinted from: *Sensors* **2025**, *25*, 5768, https://doi.org/10.3390/s25185768 . . . . . **236**

Preface

Intelligent robotic systems increasingly rely on advanced sensing capabilities to perceive, understand, and interact with complex physical environments. Vision provides rich information about objects, scenes, and motion, while tactile and force sensing enable robots to directly perceive contact, interaction forces, and other physical properties that cannot be fully captured through vision alone. The integration of these sensing modalities with artificial intelligence, robotic control, and manipulation is therefore becoming an important foundation for the development of more autonomous, adaptive, and dexterous robotic systems.

This Reprint brings together recent research on advanced sensing and perception technologies for intelligent robotics, with a particular focus on vision, touch, force sensing, sensor fusion, and dexterous manipulation. The contributions span sensing hardware, perception algorithms, interaction force estimation, multimodal sensing, robotic manipulation, and sensor informed control. They reflect the growing trend toward tightly integrating perception and action so that sensing is not treated as an isolated component, but as an essential part of the closed loop interaction between robots and their environments.

The purpose of this Reprint is to provide researchers and engineers with a broad view of current developments in robotic sensing and their role in intelligent manipulation and interaction. We hope that the studies collected here will promote further research on robust sensing technologies, multimodal perception, intelligent control, and their practical implementation in robotic systems. This Reprint is intended for researchers, engineers, and students working in robotics, sensing technologies, automation, artificial intelligence, human robot interaction, and related fields.

Chongkun Xia, Xueqian Wang, and Yajing Shen
Guest Editors

sensors

MDPI

Editorial

Advanced Sensors for Intelligent Robotic Systems: Vision, Touch, and Dexterous Manipulation

Yanyan Wu [1], Xueqian Wang [2], Yajing Shen [3] and Chongkun Xia [1,*]

[1] School of Advanced Manufacturing, Sun Yat-Sen University, Shenzhen 518005, China; wuyy363@mail2.sysu.edu.cn
[2] Shenzhen Graduate School, Tsinghua University, Shenzhen 518055, China; wang.xq@sz.tsinghua.edu.cn
[3] Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong; eeyajing@ust.hk
* Correspondence: xiachk5@mail.sysu.edu.cn

1. Introduction and Current Trends in the Field

Advanced sensing is rapidly reshaping intelligent robotic systems by enabling robots to perceive, reason, and interact with the physical world under real-world uncertainty. Vision sensors provide rich spatial and semantic information, while tactile and force sensing deliver interaction cues that are fundamentally inaccessible to purely optical systems—such as contact state, incipient slip, frictional constraints, and distributed load transfer [1,2]. With the recent progress of AI-driven multimodal learning, perception is no longer an isolated front-end module; instead, it increasingly becomes a co-designed component tightly coupled with control and decision-making, closing the perception–action loop for robust grasping and dexterous manipulation [3,4].

Several trends are particularly evident. First, sensing solutions are moving toward task-oriented co-design, where hardware structures, sensor layouts, and decoding models are jointly optimized for manipulability and cost-effectiveness [5,6]. Second, there is increasing emphasis on robust perception in challenging environments, including low light, occlusions, dynamic disturbances, and long-term deployment constraints [7,8]. Third, fusion and compensation across modalities such as vision and touch, or proprioception and exteroception, as well as compensation for error sources such as zero drift, creep, and parameter uncertainty, are becoming essential for reliable closed-loop interaction [9]. Finally, the community continues to pursue deployable systems: low-cost, real-time, and engineering-validated solutions that can translate from lab prototypes to operational platforms in industrial automation, healthcare, and field robotics [10].

This Special Issue, "Advanced Sensors for Intelligent Robotic Systems: Vision, Touch, and Dexterous Manipulation", was organized to gather innovative sensing technologies and perception–control methodologies that bridge the perception–action gap, with particular interest in vision–touch frameworks, sensor fusion, and practical validation in manipulation and robotics applications.

2. Scope of Special Issue and Contributions

This Special Issue collates 11 accepted papers (10 Articles and 1 Review) after a rigorous peer-review process. The accepted contributions span sensing hardware, perception algorithms, and closed-loop robotic systems, with representative topics including:

- tactile/force sensing and interaction intelligence for manipulation;
- robust vision-based perception (pose estimation, tracking, low-light sensing, and image enhancement);

- sensor-informed modeling, calibration, and control for complex robotic platforms;
- system-level integration and validation in healthcare, industrial, aerial, marine, and space robotics.

These contributions are outlined in more detail below.

In contribution 1, Wu et al. present a sparse strain-node tactile interface device that achieves high-resolution contact localization and three-axis force estimation with a small number of readout channels, combining structural design with learning-based decoding and sim-to-real calibration to improve accuracy while maintaining low hardware cost.

In contribution 2, Wen et al. propose a universal tool-end interaction force estimation approach for robotic tool manipulation. By integrating online-identifiable Newton–Euler tool dynamics, zero-drift identification, and an SNN-based compensation mechanism for wrist sensor creep, the method enables precise force/torque estimation across tool changes and grasp posture variations, supporting compliant manipulation control.

In contribution 3, Shan et al. develop a low-cost, real-time pipeline for multi-robot 6D pose estimation and multi-target tracking using a single RGB-D camera. Their approach enhances robustness under partial occlusion and visually similar targets via confidence filtering and tracking association strategies, enabling deployable state perception for multi-robot coordination.

In contribution 4, Ding et al. address vision sensing degradation under insufficient and uneven illumination by introducing a low-light waste-rock detection method that combines illumination-adaptive enhancement and detection-oriented feature modulation. This work demonstrates how perception models can be redesigned to improve reliability of sensor-based detection in harsh industrial environments.

In contribution 5, Liu et al. provide a comprehensive review of deep learning approaches for single-image super-resolution (SISR). By organizing methods around core components (datasets, upsampling strategies, objectives, and quality assessment), the review offers an accessible knowledge framework and highlights open challenges relevant to robotics perception, where resolution constraints frequently limit downstream recognition and tracking performance.

In contribution 6, Xue et al. report a cable-driven hybrid robot with series–parallel coupling, and develop modeling, workspace analysis, and self-calibration strategies for high-precision operation in large workspaces. The work emphasizes how multi-sensor system design and calibration pipelines underpin robust manipulation and automated grasping in complex hybrid robots.

In contribution 7, Hu et al. study kinematic modeling for cable-driven parallel robots with adaptive pulley kinematics, where pulley motion significantly affects cable length variation. By establishing an improved kinematics model and proposing a hybrid Levenberg–Marquardt and genetic algorithm solver, this work advances accurate modeling and high-precision kinematic solutions for CDPR systems.

In contribution 8, Guerreiro et al. present an automatic grasping system for multi-drone parcel delivery, combining a grasping mechanism, marker-based pose estimation, and a hybrid MPC strategy that coordinates aerial motion and gripper actuation across discrete phases. This paper highlights the role of perception-guided, sensor-informed control in autonomous aerial manipulation.

In contribution 9, Ompico et al. develop a robot-assisted TMS localization system that integrates a 3D camera with dual capacitive sensors for coil tilt detection and contact feedback. The proposed system emphasizes cost-effective sensing and human-interaction safety, illustrating how sensor fusion can enhance reproducibility and reduce operator dependency in clinical robotic assistance.

https://doi.org/10.3390/s26092626

In contribution 10, Hu et al. investigate base pose stabilization for a post-capture space robot–satellite combined system. By integrating dynamic feedforward compensation with PD control and validating via simulation under disturbance and uncertainty, this contribution demonstrates sensor-informed modeling and control for robust operation in space robotic missions.

In contribution 11, Liu et al. present a real-time deep-sea mooring system that integrates inductive telemetry with satellite communication and multi-sensor arrays for long-term deployment. While outside conventional manipulation settings, this work exemplifies robust multi-sensor integration, reliability engineering, and real-world validation—capabilities increasingly essential for field robotics sensing infrastructures.

3. Conclusions

As Guest Editors, we are pleased with the outcome of this Special Issue and the breadth of sensing innovations and system-level validations represented by the accepted papers. Collectively, these contributions illustrate a clear direction for intelligent robotics: sensing solutions must be accurate, robust, and deployable, while being designed together with learning and control to close the perception–action loop in contact-rich tasks. We sincerely thank all authors for their high-quality submissions and the reviewers for their careful evaluations and constructive suggestions. We also acknowledge the Sensors Editorial Office and managing team for their consistent support throughout the preparation, submission, and review processes.

Author Contributions: Conceptualization, Y.W.; methodology, Y.W.; validation, Y.W.; investigation, Y.W.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W., X.W., Y.S. and C.X.; supervision, Y.W. and C.X.; project administration, Y.W. and C.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China (2024YFB3213600), Shenzhen Science and Technology Program (JCYJ20250604175514018, JCYJ20241202123724032).

Conflicts of Interest: The authors declare no conflicts of interest.

List of Contributions:

1. Wu, Y.; Wu, H.; Han, Y.; Ding, Y.; Cao, B.; Xia, C. High-Resolution Contact Localization and Three-Axis Force Estimation with a Sparse Strain-Node Tactile Interface Device. *Sensors* **2026**, *26*, 1378. https://doi.org/10.3390/s26041378.
2. Xue, Z.; Yang, Z.; Hu, J.; Zhu, B.; Peng, J. A Cable-Driven Hybrid Robot with Series-Parallel Coupling: Design, Modeling, Optimization Analysis, and Trajectory Tracking. *Sensors* **2026**, *26*, 1147. https://doi.org/10.3390/s26041147.
3. Liu, T.; Zhou, B.; Zhang, X.; Peng, J.; Huang, H.; Jian, T. Real-Time Deep-Sea Mooring System with Inductive Telemetry and Multi-Sensor Integration: Deployment and Performance in the South China Sea. *Sensors* **2026**, *26*, 937. https://doi.org/10.3390/s26030937.
4. Ompico, C.D.S.; Banayo, J.N.; Mashio, Y.; Odagaki, M.; Kikuchi, Y.; Sy, A.C.; Kurosaki, H. Development of a Robot-Assisted TMS Localization System Using Dual Capacitive Sensors for Coil Tilt Detection. *Sensors* **2026**, *26*, 693. https://doi.org/10.3390/s26020693.
5. Guerreiro, B.J.; Azevedo, F.; Oliveira, P.; Cunha, R. Automatic Grasping System and Hybrid Controller Towards Multi-Drone Parcel Delivery. *Sensors* **2026**, *26*, 653. https://doi.org/10.3390/s26020653.
6. Hu, Z.; Yang, J.; Xu, W.; Chen, H.; Xu, L.; Meng, D. Pose Stabilization Control for Base of Combined System Using Feedforward Compensation PD Control During Target Satellite Transposition. *Sensors* **2026**, *26*, 206. https://doi.org/10.3390/s26010206.
7. Hu, Z.; Deng, C.; Wang, K.; Peng, J. Kinematic Modeling and Solutions for Cable-Driven Parallel Robots Considering Adaptive Pulley Kinematics. *Sensors* **2026**, *26*, 39. https://doi.org/10.3390/s26010039.

8. Shan, B.; Zhao, D.; Zhao, R.; Hiroshi, Y. Real-Time 6D Pose Estimation and Multi-Target Tracking for Low-Cost Multi-Robot System. *Sensors* **2025**, *25*, 7130. https://doi.org/10.3390/s25237130.
9. Wen, D.; Xiao, J.; Xie, Y.; Luo, T.; Zhang, J.; Zhou, W. A Universal Tool Interaction Force Estimation Approach for Robotic Tool Manipulation. *Sensors* **2025**, *25*, 6619. https://doi.org/10.3390/s25216619.
10. Ding, J.; Qu, F.; Zhou, W.; Xu, J.; Zhao, L.; Ji, Y. A Sensor Based Waste Rock Detection Method in Copper Mining Under Low Light Environment. *Sensors* **2025**, *25*, 5961. https://doi.org/10.3390/s25195961.
11. Liu, Z.; Jiang, S.; Feng, S.; Song, Q.; Zhang, J. Comprehensive Review of Deep Learning Approaches for Single-Image Super-Resolution. *Sensors* **2025**, *25*, 5768. https://doi.org/10.3390/s25185768.

References

1. Kappassov, Z.; Corrales, J.-A.; Perdereau, V. Tactile sensing in dexterous robot hands—Review. *Robot. Auton. Syst.* **2015**, *74*, 195–220. [CrossRef]
2. Xia, Z.; Cao, X.; Li, Y. A review on sensory perception for dexterous robotic manipulation. *Int. J. Adv. Robot. Syst.* **2022**, *19*, 1–21. [CrossRef]
3. Shahria, M.T.; Sunny, M.S.H.; Zarif, M.I.I.; Ghommam, J.; Ahamed, S.I.; Rahman, M.H. A Comprehensive Review of Vision-Based Robotic Applications: Current State, Components, Approaches, Barriers, and Potential Solutions. *Robotics* **2022**, *11*, 139. [CrossRef]
4. Han, X.; Chen, S.; Fu, Z.; Feng, Z.; Fan, L.; An, D.; Wang, C.; Guo, L.; Meng, W.; Zhang, X.; et al. Multimodal Fusion and Vision-Language Models: A Survey for Robot Vision. *Inf. Fusion* **2025**, *73*, 103652. [CrossRef]
5. Grushko, S.; Vysocký, A.; Heczko, D.; Bobovský, Z. Intuitive Spatial Tactile Feedback for Better Awareness about Robot Trajectory during Human–Robot Collaboration. *Sensors* **2021**, *21*, 5748. [CrossRef] [PubMed]
6. Kakani, V.; Yang, T.; Tian, Y. Vision-Based Tactile Sensor Mechanism for the Estimation of Contact Position and Force Distribution Using Deep Learning. *Sensors* **2021**, *21*, 1920. [CrossRef] [PubMed]
7. Garcia, G.J.; Corrales, J.A.; Pomares, J.; Torres, F. Survey of Visual and Force/Tactile Control of Robots for Physical Interaction in Spain. *Sensors* **2009**, *9*, 9689–9733. [CrossRef] [PubMed]
8. Zhao, W.; Gangaraju, K.; Yuan, F. Multimodal perception-driven decision-making for human-robot interaction: A survey. *Front. Robot. AI* **2025**, *12*, 1604472. [CrossRef] [PubMed]
9. Kroemer, O.; Niekum, S.; Konidaris, G. A Review of Robot Learning for Manipulation: Challenges, Representations, and Algorithms. *J. Mach. Learn. Res.* **2021**, *22*, 1–82.
10. Zhan, Z.; Zhan, Z.; Yang, Y.; Zuo, W.; Xie, M.; Ning, M. Recent Advances and Challenges of Tactile Sensing for Robotics: From Fundamentals to Applications. *Mater. Today Phys.* **2025**, *54*, 101740. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

sensors

MDPI

Article

High-Resolution Contact Localization and Three-Axis Force Estimation with a Sparse Strain-Node Tactile Interface Device

Yanyan Wu †, Hanhan Wu †, Yifei Han, Yi Ding, Bosheng Cao and Chongkun Xia *

School of Advanced Manufacturing, Sun Yat-sen University, Shenzhen 518005, China;
wuyy363@mail2.sysu.edu.cn (Y.W.); wuhh27@mail2.sysu.edu.cn (H.W.); hanyf33@mail2.sysu.edu.cn (Y.H.);
dingy96@mail2.sysu.edu.cn (Y.D.); caobsh5@mail2.sysu.edu.cn (B.C.)
* Correspondence: xiachk5@mail.sysu.edu.cn
† These authors contributed equally to this work.

Abstract

High-resolution contact localization and three-axis force estimation are crucial for human–robot interaction and precision manipulation, yet the sensing area is limited by channel density and wiring cost. Sparse strain readout makes joint estimation of location and three-axis force challenging due to cross-axis coupling and nonlinear responses, while dense arrays or extensive calibration increase complexity. We present a sparse strain-node tactile interface device (SSTID) whose three-module layout is optimized via particle swarm optimization to maximize informative response overlap, enabling contact localization (x, y) and three-axis force (F_x, F_y, F_z) estimation using only nine strain channels. We further propose a strain-node contact-state decoding framework (SCDF) implemented with a lightweight multilayer perceptron and trained via a two-stage sim-to-real strategy, including FEM pretraining followed by few-shot real-data adaptation. Experiments demonstrate accurate contact-state decoding with full-workspace characterization, supporting low-cost and scalable deployment of sparse tactile interfaces.

Keywords: sparse tactile sensing; contact localization; three-axis force estimation; sim-to-real transfer; few-shot fine-tuning

1. Introduction

High-resolution contact localization and three-axis force estimation are indispensable for precise contact measurement and surface interaction [1–3]. Ideally, a tactile sensing system should not only measure the three-axis contact force vector but also localize the contact point accurately in space to support tasks such as contact-state estimation, force control, and slip feature analysis [4–7]. Conventional high-resolution tactile sensors often rely on dense arrays, which increase channel count, wiring complexity, packaging volume, and cost, and reduce reliability and cumbersome calibration procedures [8–10]. For deployable tactile systems, achieving perception capabilities comparable to high-resolution arrays under a "sparse sensing unit" constraint has become an important research direction in tactile sensing [11–13].

To alleviate the interconnect and packaging burden of high-density arrays, several alternative routes have been explored. Vision-based tactile sensors [14,15] can achieve high spatial resolution by tracking deformations of a compliant medium; however, they require imaging optics and additional packaging volume, which limits form-factor flexibility and robustness in compact deployments [16]. Moreover, learning a reliable mapping from

observed deformations to force/location often demands large datasets or high-fidelity modeling, and performance is often sensitive to the domain gap between simulation and physical hardware [17]. To address this, Sferrazza et al. [18] proposed a sim-to-real strategy that generates training data in simulation and transfers it to real high-resolution optical tactile sensing, showing that "simulation-synthesized data and learning-based inference" can substantially reduce the burden of real data collection, while domain discrepancy remains a key factor affecting final accuracy and robustness.

A scalable alternative is computational tactile sensing, where mechanical transmission yields overlapping receptive fields and learning-based models reconstruct higher-resolution contact states from sparse readout. Recent super-resolution frameworks demonstrate that sparse tactile readout can still support dense virtual outputs when sensor layout and decoding models are co-designed [19]. Complementary theoretical [20,21] analyses further clarify the conditions and trade-offs for super-resolution under multidirectional loading, highlighting that stable multi-axis force and location decoding require careful handling of coupling, nonlinearity, and domain shift. Collectively, these studies suggest that sparse sensing units do not necessarily imply low spatial resolution if the structure and mapping model are properly designed; nevertheless, the stable decoding of the multidirectional force and contact location still require coupling, nonlinearity, and sim-to-real domain discrepancy to be addressed.

Among sparse tactile sensing modalities, strain-based sensing offers advantages such as compact structure, easy integration, fast response, and low cost, making it suitable for sparse deployment in confined spaces [22,23]. However, few-channel strain signals often exhibit pronounced cross-axis coupling and nonlinear responses, and the mapping between the contact location and three-axis force can drift with variations in material properties, assembly tolerances, bonding processes, and circuit gain. Recent studies on modular tactile units based on three-dimensional micro strain gauges have demonstrated the potential of micro-structured strain devices for multi-axis stimulus measurement and integration, while also highlighting the importance of multi-axis decoupling and consistent manufacturing for system performance [24]. On the other hand, low-cost tactile fingertips for robot end-effectors achieve contact localization and multi-DoF force/torque sensing via structural compliance and model-based inference, emphasizing deployability, integrability, and calibratability; yet such approaches still need to handle structural errors and calibration complexity [25,26]. Although finite element simulation can generate large-scale training data covering diverse conditions at relatively low cost, directly transferring the learned mapping from simulation to hardware is often limited by sim-to-real domain shift [27]. Conversely, relying entirely on large amounts of real data for end-to-end training or pointwise calibration significantly increases experimental cost, undermining the advantages of sparse and minimalist designs [28].

While prior sparse/super-resolution tactile studies commonly aim to reconstruct a dense virtual tactile map from sparse measurements (i.e., a super-resolution inversion problem) [19–21], our goal is different: we target contact-state inference under sparse strain readout by directly estimating the physically meaningful state. This leads to three key distinctions. (i) Decoding principle: instead of "sparse-to-dense" reconstruction that may amplify coupling and ambiguity under multi-axis loading, we formulate force–location decoupling as a unified regression-based contact-state inference problem, which directly outputs location and three-axis force without requiring an intermediate dense field. (ii) Training strategy: rather than depending on large-scale real calibration or purely simulation-trained models that degrade under domain shift, we adopt a two-stage sim-to-real pipeline that learns a physics-consistent prior from FEM data and then performs prior-constrained few-shot adaptation using a compact real dataset, explicitly mitigating systematic discrepancies. (iii) Hardware

https://doi.org/10.3390/s26041378

layout optimization: instead of using heuristic sensor placement or fixed layouts, we optimize the minimal three-module architecture and placement via PSO to maximize informative response overlap (i.e., identifiability) across the workspace under multi-directional loading, enabling channel-efficient decoding with only nine strain channels.

To address the strong coupling, nonlinearity, and sim-to-real domain shift that hinder high-resolution contact localization and three-axis force estimation under sparse strain readout, we jointly co-design the hardware structure and the decoding algorithm. Firstly, we design a compact sparse strain-node tactile interface device (SSTID) that adopts a three-module layout optimized via particle swarm optimization (PSO) to maximize effective response overlap across the workspace, enabling contact localization and three-axis force estimation using only nine strain channels. Building on this optimized structure, we first propose a strain-node contact-state decoding framework (SCDF) that formulates contact-state inference as a unified regression problem and is implemented with a lightweight multilayer perceptron (MLP). The framework adopts a two-stage sim-to-real strategy. Stage I pretrains the decoder on 14,400 FEM-simulated single-point contacts to learn a physics-consistent prior mapping of diverse contact locations, directions, and magnitudes. Stage II performs prior-constrained, regularized backpropagation using only 648 real calibration samples to compensate for systematic discrepancies caused by material properties, assembly tolerances, and circuit gain variations, thereby reducing the sim-to-real gap. With this pipeline, the proposed framework supports high-resolution contact-state decoding under sparse readout channels while maintaining low calibration cost and real-time deployability. More broadly, it offers a practical and scalable route toward low-cost, high-resolution three-axis tactile perception for contact-rich robotic sensing and control. The main contributions of this work are as follows:

(1) We establish a minimum three-module architecture and optimize its placement to maximize effective response overlap across the workspace, resulting in a channel-efficient SSTID with only nine strain channels over a 60 mm diameter workspace at a hardware cost below USD 15.
(2) We first propose an SCDF for force–location decoupling that employs a two-stage MLP training strategy. Stage I pretrains the decoder on 14,400 FEM-simulated contacts to learn a physics-consistent prior, and Stage II performs regularized few-shot fine-tuning with 648 real samples to reduce sim-to-real domain shift.
(3) We characterize its full-workspace accuracy and resolution, low force stability, repeatability/hysteresis/drift, and dynamic response, and further demonstrate sparse high-resolution three-axis force sensing via force regulation experiments.

2. Materials and Methods

2.1. Structural Optimization Design and Principal Analysis of the SSTID

To obtain the optimal channel design for the SSTID, we performed structural optimization. Under minimalist, low-cost hardware constraints, a key question is the minimum number of sensing modules required and how to place them to achieve stable and reproducible decoding of both the contact location and three-axis force. With fewer than three modules, the planar geometric reference frame becomes degenerate, making contact localization prone to multiple solutions or instability under noise. Therefore, we adopt a three-module configuration as the minimum setup that satisfies geometric non-degeneracy and provides sufficient information.

Under the three-module configuration, the joint identifiability of the contact location and three-axis force depends on two conditions: (i) the three modules must form a non-collinear geometric reference frame in the plane to avoid degeneracy in localization; and (ii) any contact within the workspace should be able to simultaneously excite at least three

https://doi.org/10.3390/s26041378

modules to produce effective responses—otherwise, the estimation becomes ill-conditioned due to insufficient information or noise amplification. Accordingly, as shown in Figure 1a, we define the "effective response region" of each module in the contact plane as a set. When the contact lies within A_i, the reaction force magnitude F_i of module i exceeds a preset threshold and can be used for inference; outside this region, the response decays below the threshold and contributes negligibly to estimating location and load. The region A_i can be approximated by a circular area with radius r based on a threshold isoline, where r is determined jointly by structural transmission characteristics and the chosen threshold and can be calibrated via experiments or simulations.

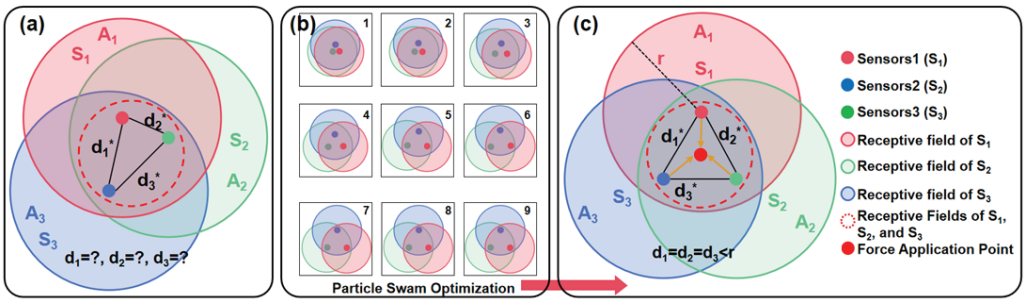


Figure 1. Optimization of the sensing units' layout. (**a**) Problem statement for the optimization process. (**b**) Optimization process using the particle swarm optimization algorithm. (**c**) The optimization result, demonstrating that the optimal layout is achieved when $d_1 = d_2 = d_3 = r$.

Based on the above definition, we assume that the three module centers form a triangle in the plane with pairwise distances d_1, d_2, d_3. To ensure that any workspace point can trigger effective responses from at least three modules, we enforce an overlap constraint that limits inter-module distances to be no larger than the effective radius r, and we maximize the union area of the three circular regions as a layout efficiency metric.

$$(d_1^*, d_2^*, d_3^*) = \arg\max_{\{d_1,d_2,d_3\}} (A_1 \cup A_2 \cup A_3), \; \textit{subject to } d_1 \le r, \; d_2 \le r, \; d_3 \le r \quad (1)$$

As shown in Figure 1b, we solve this optimization using particle swarm optimization (PSO) to obtain the optimal geometry under a given r. Under the general condition $r_1 : r_2 : r_3 = 1 : 1 : 1$, common optimizers converge to the same optimum: the three units lie at the vertices of an equilateral triangle with optimal side length r, yielding $d_1^* = d_2^* = d_3^* = r$, as shown in Figure 1c. This result directly supports the threefold-symmetric (0°, 120°, and 240°) module layout of the SSTID; while keeping the structure minimal, it maximizes the receptive field union and ensures simultaneous availability of informative signals across the workspace, thereby improving identifiability and reducing layout sensitivity for subsequent static decoupling and learning-based regression. This analysis provides a quantitative and reproducible basis for selecting a low-cost three-module layout.

Based on the layout analysis, the device adopts a threefold-symmetric configuration: the sensing plane consists of an SUS304 stainless steel top plate (diameter: 60 mm), three three-axis force sensing modules distributed circumferentially with 120° symmetry, and a rigid bottom plate made of the same material (Figure 2a). Both the top and bottom plates are 1 mm thick, and the top plate–sensor modules–bottom plate are bonded with metal adhesive to form an integrated force transmission chain (Figure 2b). Each sensing module uses a layered structure: a polycarbonate pick layer on top for load bearing (Pick), followed by a flexible printed circuit (FPC), acrylic pressure-sensitive adhesive (ADH), and a polyimide (PI) substrate. Seven metal foil strain gauges are integrated on the underside of the PI film (overall dimensions shown in Figure 2c). The strain gauges are arranged with

https://doi.org/10.3390/s26041378

functional zoning to enhance the separability of the three force components (Figure 2d): tangential gauges enhance the F_x shear response, radial gauges enhance the F_y shear response, and the central region enhances the F_z normal response. When an external load is applied to the pick layer, deformation is transmitted through the FPC–ADH–PI stack to the strain gauge array, inducing resistance changes and forming strain observations correlated with three-axis forces, thereby providing a clear physical basis for cooperative decoding of the three-axis force and contact location. The total bill-of-materials cost is below USD 15, as summarized in Table 1. Importantly, this layout optimization not only maximizes channel efficiency but also improves the conditioning of the subsequent force–location decoupling model, thereby reducing sensitivity to noise and facilitating robust two-stage sim-to-real training.

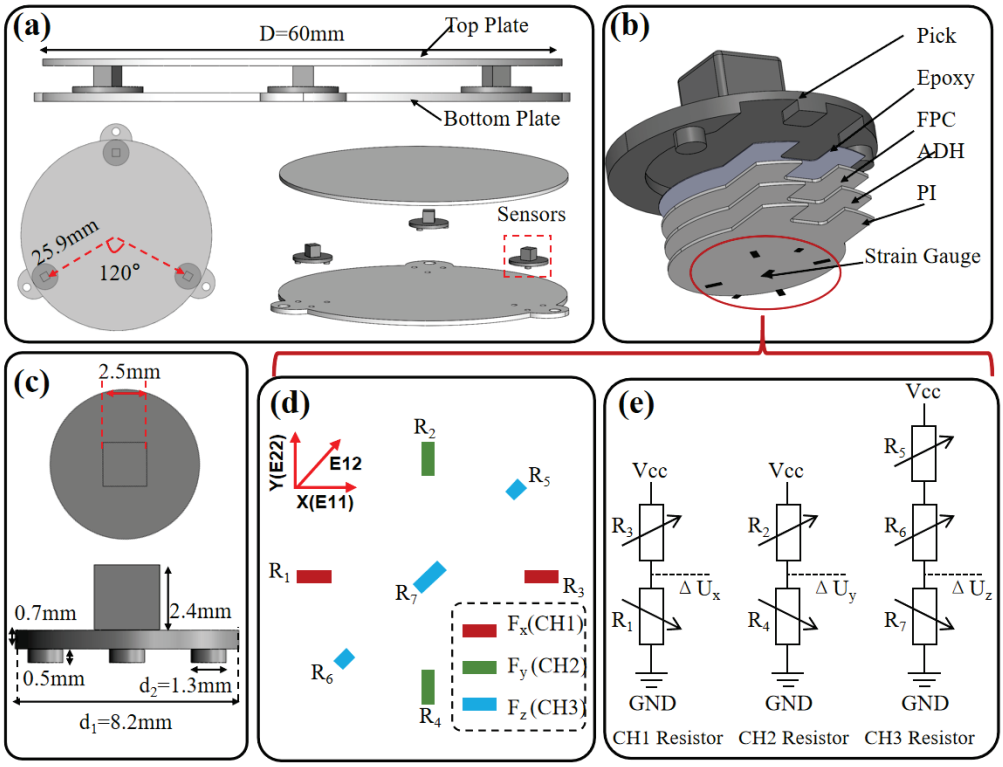


Figure 2. A schematic illustration of the structure and working principle of an SSTID. (**a**) A schematic illustration of the structure. (**b**) Geometric parameters of pick interface. (**c**) Strain gauge mounting architecture. (**d**) Strain gauge layout. (**e**) Wheatstone bridge circuit schematic.

Table 1. The cost of the SSTID.

Part	Quantity	Unit Price (USD)	Total Price (USD)
Sensor Module	3	1.1	3.3
Microcontroller	1	5.2	5.2
Circuit Board	3	1	3
Top Plate	1	1	1
Bottom Plate	1	1	1
Base	1	0.5	0.5
Electrical Wire	1	1	1
Total Cost			15

2.2. Circuit Architecture and Readout Principle of the SSTID

The mechanical readout of the SSTID is based on the Wheatstone bridge principle, which converts strain gauge resistance variations into voltage signals for three-axis force

https://doi.org/10.3390/s26041378

prediction. As shown in Figure 2d,e, three orthogonally arranged strain gauge groups are integrated on the PI film, and each group adopts a half-bridge configuration to save space, corresponding to the measurement of the three force axes. When an external load is applied to the pick layer, mechanical deformation is transmitted through the FPC–ADH–PI laminate, inducing resistance changes ΔR_i. The X- and Y-axis gauges each form a half-bridge, and the Z-axis gauges form a half-bridge with the corresponding resistors, as illustrated in Figure 2e. Here, R_1, R_2, R_3, R_4, R_7 share the same initial resistance R, while R_5 and R_6 are $R/2$. ΔR_i denotes the resistance change under deformation, and $\Delta R_i/R$ is the normalized resistance variation. The X- and Y-axis half-bridge output voltages $V_{\text{out_x}}$ and $V_{\text{out_y}}$ satisfy:

$$V_{\text{out_x}} = V_{\text{cc}}\left(\frac{1+\Delta R_1/R}{2+\Delta R_1/R+\Delta R_3/R}\right),\ V_{\text{out_y}} = V_{\text{cc}}\left(\frac{1+\Delta R_4/R}{2+\Delta R_4/R+\Delta R_2/R}\right), \quad (2)$$

The Z-axis output voltage $V_{\text{out_z}}$ can be expressed as:

$$V_{\text{out_z}} = V_{\text{cc}}\left(\frac{1+\Delta R_7/R}{2+\Delta R_7/R+(\Delta R_5+\Delta R_6)/R}\right), \quad (3)$$

The normalized variations $\Delta R_i/R$ for the X/Y/Z axes are jointly contributed by the averaged orthogonal strain components E_{11} (length direction), E_{22} (width direction), and E_{12} (bisector direction), as shown in Figure 2d. Thus, $\Delta R_i/R$ can be written as:

$$\frac{\Delta R_i}{R} = L \times E_{11}(R_i) + W \times E_{22}(R_i), \quad (4)$$

where L and W are the sensor dimensions along the length and width directions, taking values 9.8 and 5.4, respectively. Finally, according to the strain gauge arrangement in Figure 2e, the three-axis signals corresponding to F_x, F_y, F_z can be expressed as:

$$Signal_F_x = \delta \times (\frac{\Delta R_1}{R} - \frac{\Delta R_2}{R}),\ Signal_F_y = \delta \times (\frac{\Delta R_3}{R} - \frac{\Delta R_4}{R}),\ Signal_F_z = \delta \times (\frac{\Delta R_5+\Delta R_6}{R} - \frac{\Delta R_7}{R}) \quad (5)$$

Here, δ is the bridge excitation voltage and is set to 2.8 V, enabling conversion of the applied load on a single sensor module into voltage signals.

2.3. Design of the Simulation Dataset

To enable Stage I pretraining of the force–location decoder and to provide a physics-consistent prior that is transferable to real hardware, we construct a large-scale FEM simulation dataset covering diverse contact locations, directions, and magnitudes. We perform batch FEM simulations by applying concentrated loads with varying directions and magnitudes on the surface of the tactile interaction platform and obtain the corresponding three-axis force responses of the three sensors. Following the procedure in Figure 3a, we construct a dataset of 14,400 samples: 7200 samples with random three-axis force loading in the range 0–5 N, and another 7200 samples with random normal-force loading in the range 0–5 N, both applied at random locations on the platform surface. Figure 3b–d show the stress and strain distributions of the top plate and individual modules under representative loading cases, supporting the consistency of the load–transmission–strain response chain in the FEM model. Loads in different directions induce distinctive response patterns across the three modules, providing discriminative information for subsequent regression of both (x, y) and (F_x, F_y, F_z) from few-channel signals. We validate fidelity at both the module and system levels. Figure 4 shows a quantitative FEM experiment comparison of the module-level output voltage differences, where the F_z-related (normal) channel and the F_x/F_y related (tangential) channels match closely under the same loads. For the assembled

https://doi.org/10.3390/s26041378

three-module SSTID, we further apply 1 N normal loads at 73 spatial points and obtain an average relative error of 11% for F_z, while the average tangential-force error (F_x / F_y) is 32%, mainly due to assembly tolerances and inter-module coupling not fully captured by FEM.

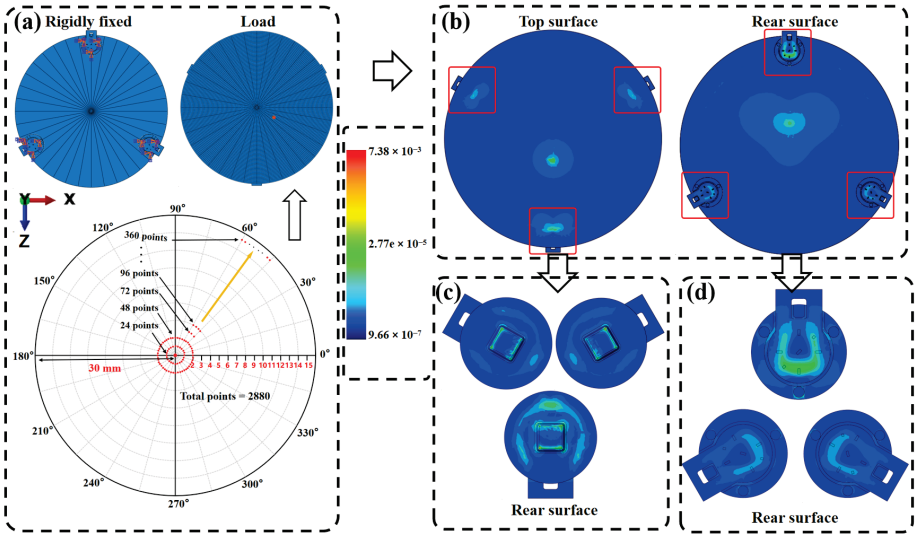


Figure 3. FEM simulation setup and dataset generation for the SSTID. (**a**) Prescribed sampling of single-point contact locations and three-axis forces over the workspace, with magnitudes spanning 0–5 N and directions covering normal and tangential loading. (**b**) Representative FEM-predicted response under applied loading with boundary constraints matching the experimental fixture. (**c**) Top-side loading cases. (**d**) Bottom-side loading cases. The responses at the nine strain readout nodes are extracted to form the Stage I pretraining dataset.

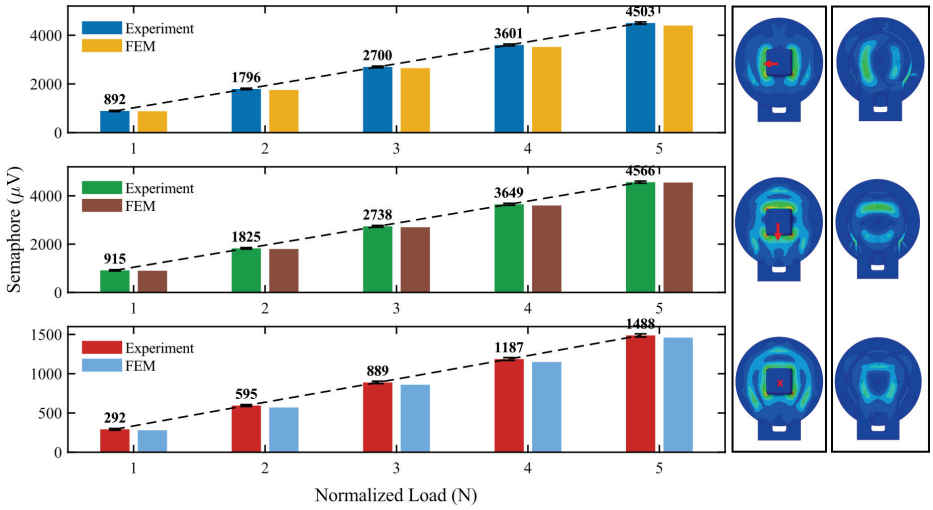


Figure 4. A comparison of the accuracy of the SSTID's (F_x, F_y, F_z) simulation and experimental results in the 0–5 N range.

To ensure physically consistent and transferable simulation data, we used the elastic moduli in Table 2 for the pick, epoxy, ADH, FPC, PI, and strain gauges, while keeping geometry and boundary conditions consistent with the physical assembly. This configuration was used to generate pretraining data and to provide a physics-consistent prior for the SSTID.

https://doi.org/10.3390/s26041378

Table 2. Material properties used in FEM simulation.

Part Name	Value (MPa)
Pick	19,000
Epoxy	250
ADH	600
FPC	4100
PI	5000
Strain gauges	90,000

2.4. Real-World Data Collection Protocol

To acquire real data and validate the performance of the SSTID, we designed and fabricated a measurement setup as shown in Figure 5a. The setup consists of a three-DoF gantry stage and a high-precision base with five degrees of freedom (X, Y, Z, rotation, and tilt), controlled via adjustment knobs. The knob resolution for X/Y/Z translation is 0.02 mm per division, and the tilt knob resolution is 0.1° per division. A force/torque (F/T) sensor (NANO25, ATI Industrial Automation) was used as the reference, with rated loads of F_x, F_y, F_z: 100 N and M_x, M_y, M_z: 4 N·m. The F/T sensor data, together with all other measurement data, were transmitted to a computer for recording.

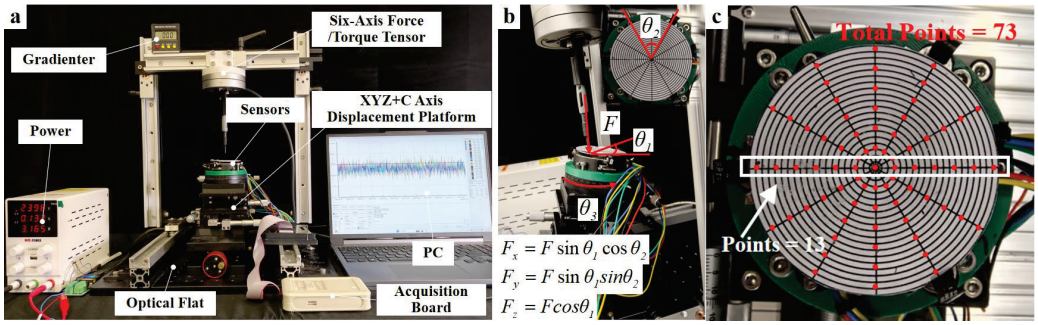


Figure 5. SSTID experimental setup. (**a**) Calibration platform; (**b**) inclined force application platform; (**c**) distribution of data collection points on the top surface of the SSTID.

For low-cost sim-to-real adaptation, we collected $N = 648$ real samples for backpropagation-based adjustment. We selected 73 spatial points on the sensing plane (computed as $13 \times 6 - 5 = 73$), where the center point is counted only once across the six radial groups. Each point included two loading types: normal loading and inclined loading.

(1) Normal loading (216 samples): At each spatial point, a normal force $F_z \in \{1,3,5\}$ N was applied, resulting in 216 normal force samples in total.

(2) Inclined loading (432 samples): At each spatial point, six inclined force samples were applied, resulting in 432 samples in total. The inclined force vector was parameterized by magnitude–tilt angle–azimuth angle: the magnitude $F_{total} \in [0,5]$ N; the tilt angle θ is the angle between the force vector and the normal Z axis; and the azimuth angle ϕ is the in-plane angle of the force vector relative to the X axis. The components are:

$$F_x = F\sin\theta\cos\phi, \quad F_y = F\sin\theta\sin\phi, \quad F_z = F\cos\theta \tag{6}$$

Accordingly, the normal-force component is denoted as F_z (and the force vector as $F = [F_x, F_y, F_z]$) throughout this work. Given the constraint of only six inclined loadings per point, we adopted a uniform coverage strategy with two tilt angles and three azimuth angles to improve generalization across shear directions and shear-to-normal ratios: $\phi \in \{0°,120°,240°\}$, and $\theta \in \{30°,60°\}$. Meanwhile, to avoid spurious correlations caused by fixed binding between the direction and magnitude, the magnitudes {1,3,5} N were

https://doi.org/10.3390/s26041378

cyclically permuted with the azimuth angles across different spatial points (a cyclic right shift of {1,3,5} according to point index), achieving balanced coverage of direction, shear ratio, and magnitude over the full plane.

The real calibration set (648 samples) is designed to be representative rather than exhaustive. This sampling strategy covers combinatorial variations in shear direction, shear-to-normal ratio, and magnitude with minimal trials per point: three azimuth angles cover the principal in-plane shear directions, and two tilt angles cover different shear proportions. The cyclic permutation between magnitudes and azimuth angles further reduces pseudo-correlation due to fixed magnitude–direction pairing, improving model generalization under a limited total sample budget and enhancing the efficiency of real-world adjustment.

2.5. Learning-Based SCDF with Two-Stage Sim-to-Real Adaptation

Based on the simulation and real calibration datasets introduced above, we next present the proposed strain-node contact-state decoding framework (SCDF) for mapping nine-channel strain observations to a continuous contact location and three-axis force. To bridge the sim-to-real gap while keeping calibration cost low, the SCDF is trained using a two-stage strategy: simulation pretraining for a physics-consistent prior followed by regularized few-shot adaptation on real data. We first define the input as a 9-dimensional displacement/strain feature vector $\mathbf{s} \in \mathbb{R}^9$ and the output as a 5-dimensional target vector:

$$\mathbf{y} = [x, y, F_x, F_y, F_z]^{\mathrm{T}} \in \mathbb{R}^5, \tag{7}$$

where (x, y) denotes the contact location and (F_x, F_y, F_z) denotes the three-axis force components. The regression model is $\hat{\mathbf{y}} = f(\mathbf{s}; \theta)$. To mitigate the influence of scale differences across dimensions during optimization, both inputs and outputs are standardized using a Standard Scaler:

$$\bar{\mathbf{s}} = (\mathbf{s} - \mu_s)/\sigma_s, \quad \bar{\mathbf{y}} = (\mathbf{y} - \mu_y)/\sigma_y, \tag{8}$$

the network regresses $\bar{\mathbf{y}}$ in the standardized space, and the outputs are de-standardized during inference to obtain physical quantities.

As illustrated in Figure 5, the SCDF is implemented as a lightweight multilayer perceptron (MLP) decoder composed of stacked fully connected blocks with nonlinear activations and a final linear output layer. This architecture is sufficient to model the strongly coupled and nonlinear mapping from few-channel strain signals to multi-dimensional contact states while remaining computationally efficient for real-time deployment.

To address the sim-to-real domain shift caused by material properties, assembly tolerances, and circuit gain variations, the SCDF adopts a two-stage sim-to-real adaptation strategy. In Stage I (simulation pretraining), we train the decoder on the FEM-based simulation dataset D_sim (14,400 samples) to obtain θ_0 which serves as a physics-consistent prior mapping. In Stage II (few-shot real adaptation), we initialize the model with θ_0 and continue training on the real calibration dataset D_real (N = 648), using prior-constrained and regularized optimization to absorb systematic hardware discrepancies into the model parameters and reduce the sim-to-real gap. The Stage II parameter update can be written as:

$$\theta \leftarrow \theta_0 - \eta \nabla_\theta \mathcal{L}_{ft}, \tag{9}$$

https://doi.org/10.3390/s26041378

thereby absorbing systematic errors caused by material, assembly, and circuit gain discrepancies into the model parameters and reducing the sim-to-real gap. The model output is grouped into location and force terms, and a weighted regression loss is adopted:

$$\mathcal{L}_{task} = \lambda_p \|\widetilde{\mathbf{p}} - \overline{\mathbf{p}}\|_2^2 + \lambda_f \left\|\widetilde{\mathbf{f}} - \overline{\mathbf{f}}\right\|_2^2, \tag{10}$$

where $\mathbf{p} = [x, y]^\top$ and $\mathbf{p} = [x, y]^\top$, $\mathbf{f} = [F_x, F_y, F_z]^\top$. With standardized outputs, we typically set $\lambda_p = \lambda_f = 1$. To suppress overfitting during few-shot adjustment and explicitly preserve the simulation prior, we introduce an L2-SP regularization term during fine-tuning:

$$\mathcal{L}_{ft} = \mathcal{L}_{task} + \beta \|\theta - \theta_0\|_2^2, \tag{11}$$

where β is the regularization weight. We determine β via a log-scale sweep on a held-out validation split of the real calibration set (e.g., $\beta \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$) and select the value that minimizes validation error; a sensitivity check confirms stable performance within a reasonable range, while too small β overfits few-shot real data and too large β under-adapts to the real domain. In this work, we use $\beta = 10^{-3}$ unless otherwise stated.

Both stages employ mini-batch optimization; pretraining uses a larger learning rate for fast convergence, whereas fine-tuning adopts a smaller learning rate and validation-based early stopping to improve generalization. During inference, the SCDF consists only of fully connected operations and element-wise activations, enabling real-time output of (x, y, F_x, F_y, F_z) on conventional edge-computing platforms. Detailed training hyperparameters and implementation settings are provided in Section 2.6.

2.6. Training Procedure

Based on the model definition and two-stage objectives in Section 2.5, we detail the implementation settings for Stage I simulation pretraining and Stage II few-shot real adaptation, including data splits, optimization settings, and training criteria. For Stage I, the FEM-simulated dataset of 14,400 single-point contacts was randomly split into 80% training, 10% validation, and 10% test. For Stage II, the real calibration dataset of 648 samples was split using the same 90%/10% protocol. Unless otherwise stated, the training pipeline and hyperparameters were identical for both stages. To balance loss contributions across dimensions and improve optimization stability, both inputs and outputs were standardized (z-score normalization). During training, the standardized inputs were forwarded through the MLP, and a sequence of fully connected linear mappings with ReLU activations produced predictions $\hat{\mathbf{y}}$ for $[x, y, F_x, F_y, F_z]$. The mean squared error loss $L(\hat{\mathbf{y}}, \mathbf{y})$ was then computed against the ground truth labels. Backpropagation was performed from this loss, and PyTorch automatic differentiation computed gradients of the loss with respect to the weights and biases of each layer via the chain rule (with gradient gating through ReLU derivatives). For each mini-batch, gradients were cleared, recomputed, and parameters were updated using the Adam optimizer (learning rate 10^{-3}), iteratively driving model outputs toward the ground truth. The model was implemented using PyTorch 2.6.0 under Python 3.12 and training converged within 300 epochs. As shown in Figures 6 and 7, the model achieved a mean squared error (MSE) = 0.011, mean absolute error (MAE) = 0.092, and $R^2 = 0.981$ on the test set, indicating good prediction accuracy and generalization. The predicted-versus-ground-truth scatter plots for each component closely align with the diagonal, demonstrating strong overall consistency across multi-dimensional outputs. Meanwhile, the loss curve stabilizes in the later stage of training and the R^2 curve converges close to 1, suggesting no noticeable oscillation or degradation

https://doi.org/10.3390/s26041378

during optimization. These results provide a reliable baseline for subsequent real-world adaptation and metrological experiments.

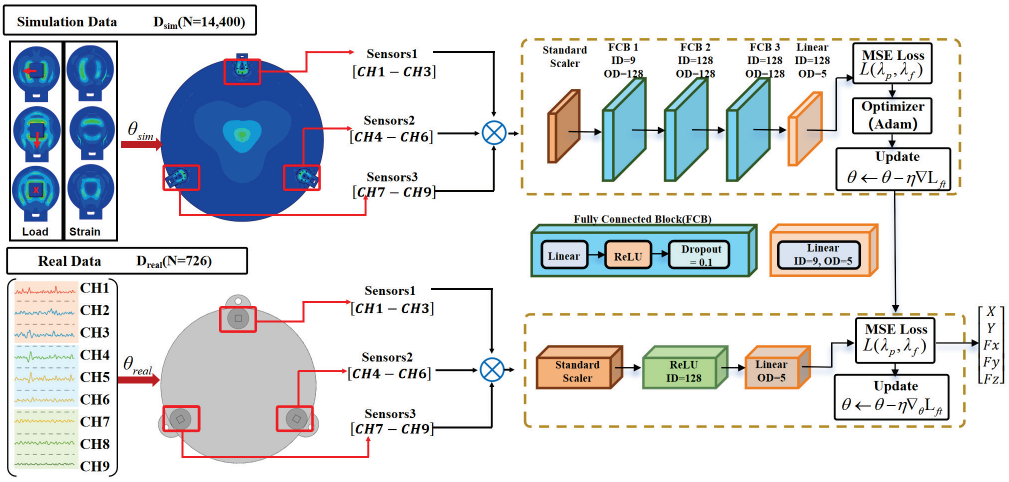


Figure 6. Model training SCDF diagram.

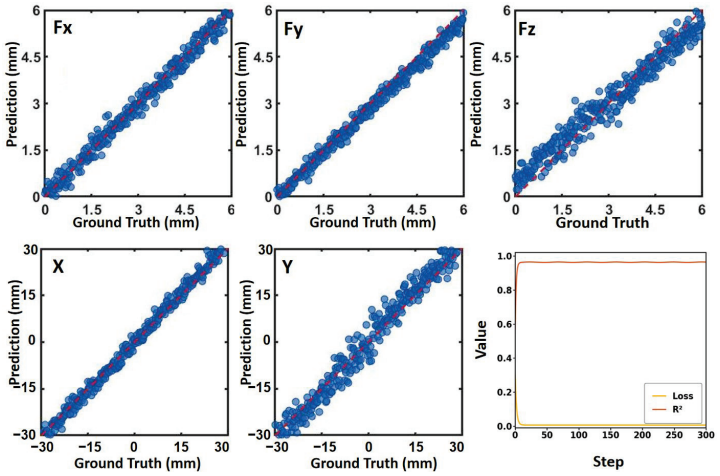


Figure 7. Training results and loss curve.

3. Results

This section evaluates the sensor in terms of: force-estimation accuracy, stability of contact localization, key metrological metrics, and application demonstrations. We first report the evaluation of temperature drift (Figure 8) and then by the mean errors of the three-axis forces and the resultant force (Figure 9), then analyze the distribution stability of estimated contact points under different load levels (Figure 10) for the SSTID through a few-shot adaptation ablation study (Table 3) and demonstrate real-time usability in interaction and game-control scenarios (Figures 11 and 12), and summarize core metrics such as full-scale range, sensitivity, resolution, repeatability, and hysteresis in Table 4. Finally, we evaluate the benefit of real calibration by varying the number of real samples used in Stage II fine-tuning and reporting the corresponding changes in contact localization and force estimation errors.

https://doi.org/10.3390/s26041378

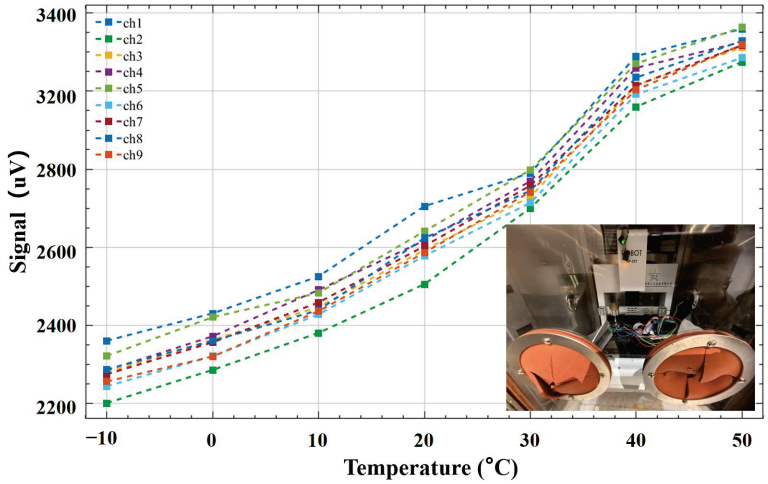


Figure 8. Variation curves of sensor signals from −10 °C to 50 °C.

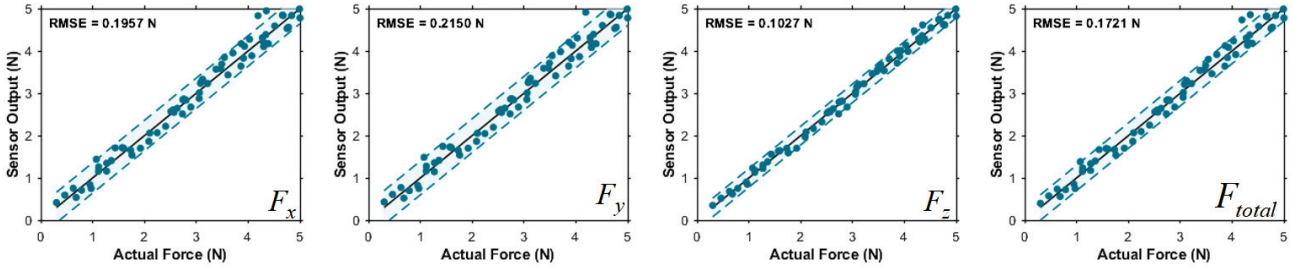


Figure 9. Mean error of the three-axis forces and the resultant force.

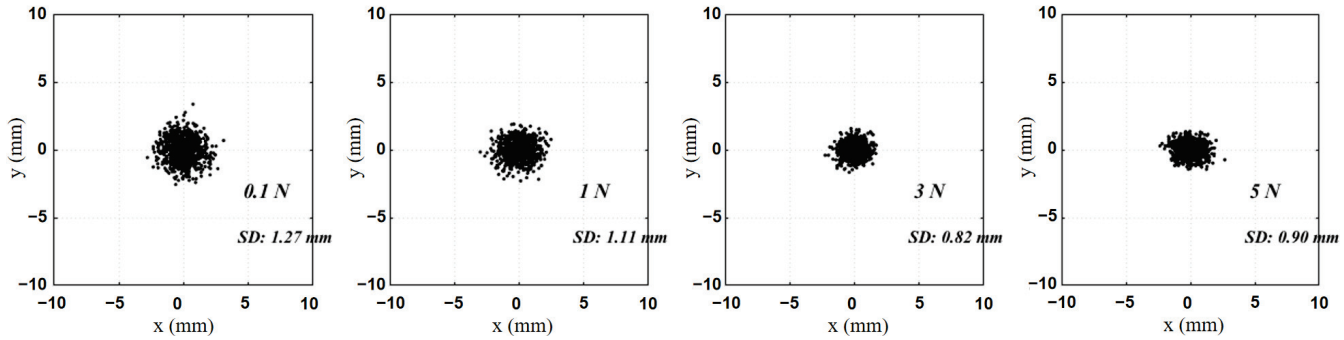


Figure 10. Distribution of measured contact points according to the applied contact force.

Table 3. Ablation study.

Method	Position MAE (mm)	Three-Axis Force MAE (N)
MLP (only simulation)	10.2	0.51
MLP + Backpropagation ($N = 216\ F_z$)	3.51	0.35
MLP + Backpropagation ($N = 216\ F_z + 144\ F_x\ and\ F_y$)	2.87	0.27
MLP + Backpropagation ($N = 216\ F_z + 288\ F_x\ and\ F_y$)	2.38	0.21
MLP + Backpropagation ($N = 648$)	1.97	0.17
KNN (K = 5, only sim data)	12.1	0.69
KNN (K = 5, only real data)	3.01	0.38
Linear Regression (only sim data)	15.3	0.75
Linear Regression (only real data)	3.59	0.43

https://doi.org/10.3390/s26041378

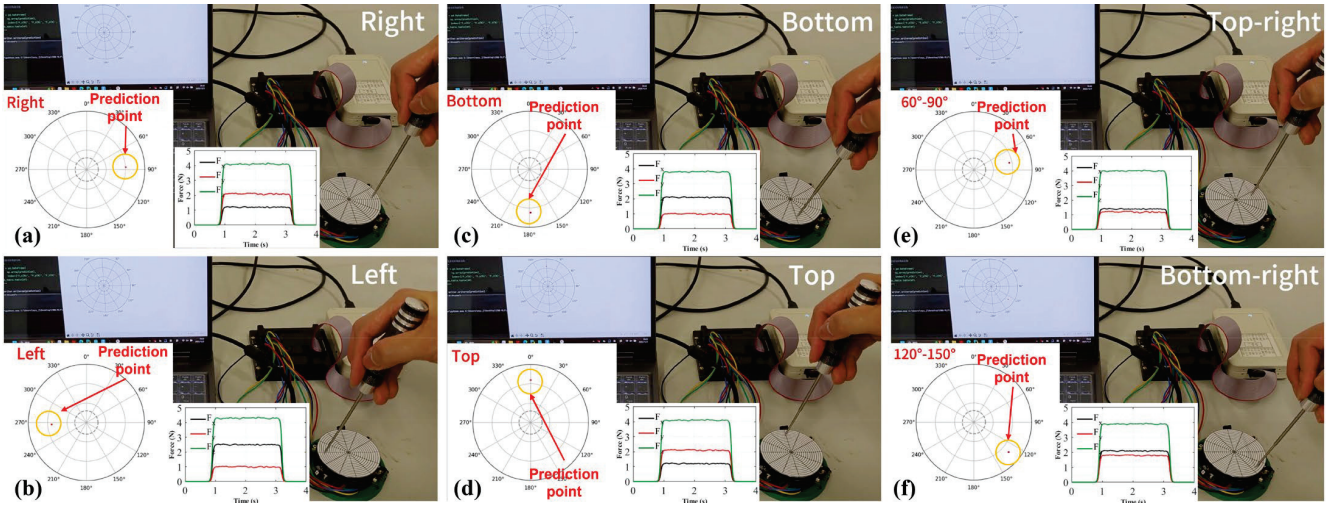


Figure 11. Contact localization experiment of the SSTID. (**a**) Right; (**b**) Left; (**c**) Bottom; (**d**) Top; (**e**) Top-right; (**f**) Bottom-right.

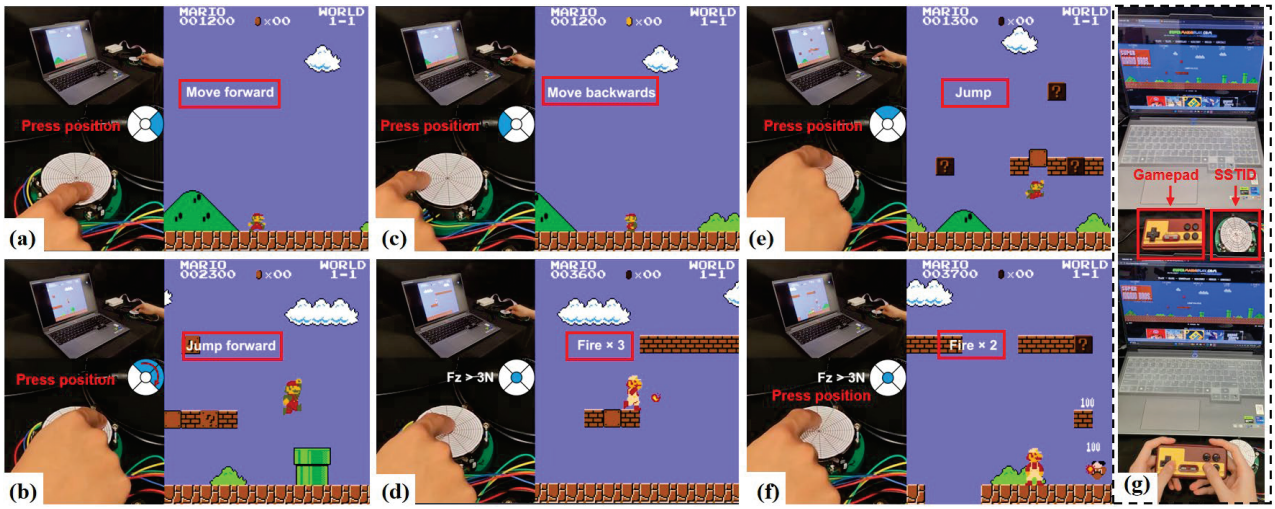


Figure 12. Playing Super Mario using the SSTID. (**a**) Move forward; (**b**) Jump forward; (**c**) Move backwards; (**d**) Release three attacks; (**e**) Jump; (**f**) Release two attacks; (**g**) Comparison diagram of playing games with controller and SSTID.

Table 4. Sensor performance characterization.

Category	Quantity	Value	Unit
Specifications	Force	5	N
	Sensitivity	102	N^{-1}
	Force resolution	5	mN
Performance metrics	Error of average repeatability	1.3	% of FSO
	Average hysteresis	3.1	% of FSO
	F_z axis 50,000 cycles of 1 N load test	19	% initial value error

Note: FSO denotes the full-scale output (5 N). "% initial value error" denotes the relative deviation from the initial reading measured after the 50,000-cycle test under a 1 N load.

3.1. Performance Evaluation of the SSTID

Table 4 summarizes key metrics including full-scale output (FSO = 5 N), sensitivity, force resolution, repeatability, and hysteresis. Repeatability error was computed from the output dispersion under repeated cyclic loading at the same load level, while hysteresis was

https://doi.org/10.3390/s26041378

computed from the output difference between loading and unloading cycles at the same force level. These metrics evaluate the engineering usability of the SSTID from a metrological perspective, beyond pure regression accuracy. As shown in Table 4, repeating a 5 N load over 100 cycles produced consistent signal variations, and the computed average repeatability error was 1.3% of FSO, indicating stable dynamic behavior and strong mechanical robustness. To analyze hysteresis, we applied a constant 5 N load to the sensor for 10 loading/unloading trials; the results show an average hysteresis of approximately 3.1% (Table 4).

Table 3 summarizes the decoder ablation and rationale. Since the SCDF maps a compact strain vector (nine channels) to the continuous contact state, we benchmark the proposed MLP decoder against classical regressors (LR and KNN). All baselines are trained and tuned on the same real calibration split via cross-validation. We note that these classical models cannot directly leverage our Stage I physics-consistent pretraining and Stage II prior-constrained adaptation in a unified differentiable pipeline; when trained on FEM data, they suffer from sim-to-real discrepancy. Table 3 shows that the MLP provides the best accuracy–latency trade-off, supporting its selection for real-time deployment. We additionally evaluate performance versus the number of real samples (e.g., 162/324/486/648) and observe diminishing returns as the sample size increases, supporting the adequacy of 648 samples. Nevertheless, we note that insufficient real samples may underrepresent extreme loads or local regions, which we list as a limitation and will address in future work with broader real data coverage.

Figure 8 presents the temperature drift characteristics of the SSTID. Experiments were conducted in a programmable temperature–humidity chamber (ST-1000L, ZhenHang, Shenzhen, China) with a controllable temperature range of -70 °C to 150 °C and an accuracy of ± 0.5 °C. The SSTID was placed in the chamber and the ambient temperature was swept from -10 °C to 50 °C in 10 °C increments, while the channel outputs were recorded. The nine-channel outputs exhibited an approximately linear dependence on temperature, with an average drift magnitude of about 300 μV per 10 °C. This drift propagated to the decoded outputs: for each 10 °C temperature change, the estimation accuracy of F_x, F_y and F_z deteriorated by approximately 0.117 N, 0.102 N, and 0.191 N, respectively; meanwhile, the contact location error increased by about 0.13 mm per 10 °C. To mitigate the temperature-induced drift observed in Figure 8, we apply a software-based online drift suppression in signal preprocessing. Specifically, the raw strain sequence x_i is smoothed using a sliding-window mean filter with window length W($W = 6$ in this work): for the first samples we use a variable-length window $y_i = \frac{1}{i}\sum_{j=1}^{i} x_j$ when $i < W$, and afterward a fixed window $y_i = \frac{1}{W}\sum_{j=i-W+1}^{i} x_j$ when $i \geq W$. This operation tracks and attenuates the slowly varying baseline component caused by thermal drift, improving output stability during long-duration operation. Note that this is a lightweight drift suppression rather than an explicit temperature model; in future work, we will incorporate on-board temperature sensing and temperature–strain co-calibration (or domain-adaptive learning) for more rigorous compensation.

Figure 9 reports the prediction errors (annotated by RMSE) of the three-axis forces (F_x, F_y, F_z) and the resultant force F_{total}. The errors are of a similar magnitude across axes, and the resultant-force error does not exhibit noticeable amplification, indicating that the model maintains good vector consistency under multi-axis coupled loading. This suggests that the physics-consistent mapping learned via simulation-prior pretraining, together with few-shot real-world adjustment that compensates for systematic biases, effectively suppresses cross-axis error propagation induced by material and assembly variations. We further examined the distribution of estimated contact points under different contact-force levels. As shown in Figure 10, when a constant force of 0.5 N, 1 N, 3 N, and 5 N was applied at a fixed point on the cover plate for 10 s, the standard deviation (SD) of estimated contact

https://doi.org/10.3390/s26041378

points was 1.32 mm at 0.5 N. At 1 N and 3 N, the SD values were 1.06 mm and 0.82 mm, respectively. At 5 N, the estimated contact point remained stable with SD = 0.87 mm.

Figure 11 shows real-time localization and force estimation under different contact regions and loading directions. The correctness of location decoupling is validated by the relative relationship between the predicted point and the ground truth contact region, and the dynamic tracking ability under pressing and direction changes is verified by the temporal responses of the three force components. This experiment highlights that the sensor not only outputs a scalar force but also achieves simultaneous decoding of "contact location + force vector" under sparse-channel constraints. See Supplementary Movie S1 for related experiments.

3.2. Demonstration in Application Scenarios

Figure 12 demonstrates a human–machine interaction application based on real-time tactile decoding. By mapping the predicted contact location to directional control and mapping F_z to a button-trigger threshold, the user can control character movement, jumping, and attacking in the game. This demonstration validates the stability and real-time capability of the decoded outputs from an application perspective and forms a closed loop with the "compatible control types" summarized in Table 5. See Supplementary Movie S2 for details.

Table 5. The sensor is compatible with game controls.

Game Genre	Controls	SSTID Adaptation
 Super Mario	D-pad movement + button jump/attack	$\uparrow\downarrow\leftarrow\rightarrow + F_z$
 Tetris	D-pad rotation + quick drop	$\uparrow\downarrow\leftarrow\rightarrow + F_z$
 Tank Battle	Stick move + fire	$\uparrow\downarrow\leftarrow\rightarrow + F_z$
 Shooter (Contra)	D-pad movement + button jump/attack	$\uparrow\downarrow\leftarrow\rightarrow + F_y + F_z$

In Figure 12, the decoded continuous contact state is mapped to discrete game commands by assigning the predicted contact location to directional inputs and assigning the decoded force component to button-trigger events, forming a closed-loop human-in-the-loop interface. To objectively quantify practicality beyond qualitative visualization, we report the quantitative metrics summarized in Table 6, including task-level performance (success rate and completion time under a fixed level segment), command-level accuracy (direction accuracy and button-trigger F_z, evaluated by time-aligned comparison between the SSTID-generated command stream and a reference stream from a standard gamepad or scripted commands, together with the mis-press rate in events/min), and latency (mean end-to-end latency measured from the decoded trigger threshold crossing t_0 to the OS-level key event timestamp t_1, as well as the in-game latency reported during gameplay). These results provide an objective and reproducible assessment of the game control interface beyond qualitative visualization.

A key advantage of the SSTID lies in its channel efficiency, i.e., how large a tactile workspace can be covered per electrical readout channel while still enabling multi-

https://doi.org/10.3390/s26041378

dimensional inference. As shown in Table 7, conventional strain gauge tactile units typically operate over millimeter-scale active areas, and multi-axis load sensing often requires multiple strain gauges per unit, leading to a small area-per-channel ratio. Here, the "channel efficiency" is used to describe the sensing coverage achieved per electrical readout channel, in order to reflect the coverage–wiring trade-off of sparse tactile interfaces. In this manuscript, it is computed as the effective workspace area divided by the number of independent strain channels. For our SSTID, the workspace is a circular region with a diameter of 60 mm (total area 28.3 cm^2) and the readout uses nine channels, corresponding to 3.14 cm^2 coverage per channel. The same definition is adopted for the entries summarized in Table 7 whenever the workspace size and channel count are available in the cited references. In contrast, our system uses only nine strain channels to achieve tactile perception over a circular workspace with a 60 mm diameter (28.3 cm^2). This corresponds to 3.14 cm^2 per channel, while simultaneously estimating the contact location (x, y) and three-axis force (F_x, F_y, F_z).

Table 6. Quantitative evaluation of the game-control demonstration and press-trigger reliability.

Evaluation Block	Protocol	Trials	Metric	Placeholder Result
Mario task level (SSTID)	Fixed level segment	20	Success rate (%)	90
		20	Completion time (s), mean ± std	41.8 ± 6.5
Mario task level (Gamepad)	Same level segment; standard gamepad (two-hand)	20	Completion time (s), mean ± std	38.5 ± 5.4
Mario command level	Time-aligned command stream comparison (direction + button)	20	Direction accuracy (%)	95.6
Latency	t_0: decoded threshold crossing;	200	Mean latency (ms), mean ± std	60 ± 12
	t_1: OS key event timestamp	200	Games latency (ms)	91
Press-based trigger reliability	Short-press: single-trigger; long-press: hold 3 s	100	Short-press trigger success (%)	98
98		50	Long-press trigger success (%)	100

Table 7. Comparative study.

Work	Sensing Principle	Readout Channels	Active Area (cm^2)	Active Area/Channel (cm^2/ch)	Outputs	Reported Performance
This work	Sparse strain sensing + learning	9	28.3	3.14/channel	(x, y, F_x, F_y, F_z)	Position MAE 1.97 mm; three-axis force MAE 0.17;
Kong et al. [19]	Sparse taxels + deep model	23	47.52	2.066	High-res pressure map/localization	Average localization error 0.73 mm;
Piacenza et al. [28]	Conductive elastomer + multi-electrode	6	1.6	0.024 /line	Localization + depth	median localization accuracy; depth error ~0.74 mm;
Lee et al. [29]	Strain gauges on micro-structure	8	0.027	0.027	Three-axis force	Triaxial sensing (unit-scale)
Mêda et al. [30]	Array + MRE skin + neural network	16	6.5	0.41	(x, y, F_x, F_y, F_z)	Location MAE 0.26 mm Force magnitude MAE 0.17 N;

https://doi.org/10.3390/s26041378

4. Conclusions

High-resolution contact localization and three-axis force estimation are crucial for human–robot interaction and precision manipulation, yet this capability is fundamentally limited by channel density and wiring cost. With sparse strain readout, low-dimensional measurements make decoupling of the location and three-axis force susceptible to cross-axis coupling and nonlinear responses. We address high-resolution contact localization and three-axis force estimation under sparse strain readout by jointly co-designing the hardware structure and decoding the algorithm. Specifically, we perform a layout optimization using particle swarm optimization (PSO) to determine the minimum three-module configuration and its optimal placement, maximizing response overlap and yielding a channel-efficient SSTID with only nine strain channels. Building on this optimized structure, we propose an SCDF that casts force–location decoupling as a unified regression problem and is realized with a lightweight multilayer perceptron (MLP). The decoder is trained with a two-stage sim-to-real strategy, where Stage I learns a physics-consistent prior from 14,400 FEM-simulated single-point contacts and Stage II performs prior-constrained, regularized adaptation using 648 real calibration samples to compensate for systematic discrepancies from material properties, assembly tolerances, and circuit gain variations.

Systematic metrological characterization and ablation studies validate accuracy and resolution across the workspace and quantify low-force stability, repeatability, hysteresis, and dynamic response. Within a 5 N full-scale range, the force resolution reaches 5 mN, with an average repeatability error of 1.3% FSO and an average hysteresis of 3.1% FSO. With $N = 648$ real samples for Stage II, the position MAE decreases to 1.97 mm and the three-axis force MAE decreases to 0.17 N, confirming the effectiveness of combining simulation-based prior learning with few-shot real adaptation. With a hardware cost below USD 15, the SSTID provides a practical reference for cost-sensitive deployment of three-axis force and contact localization in contact-rich robotic sensing and control. Looking forward, we will pursue cross-device generalization and establish an interpretable link between sensor scale and structural parameters and the resulting force–location response characteristics, enabling consistent calibration and rapid transfer across devices with different sizes and manufacturing variations.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/s26041378/s1, Supplementary Materials, Video S1 and Video S2.

Author Contributions: Conceptualization, Y.W.; methodology, YW.; validation, H.W.; investigation, Y.W., H.W. and Y.H.; data curation, Y.W., Y.D. and B.C.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W. and C.X.; supervision, Y.W. and C.X.; project administration, Y.W. and C.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China (2024YFB3213600), Shenzhen Science and Technology Program (JCYJ20250604175514018, JCYJ20241202123724032).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in this article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

https://doi.org/10.3390/s26041378

References

1. Oh, H.; Yi, G.-C.; Yip, M.; Dayeh, S.A.J.S. Scalable tactile sensor arrays on flexible substrates with high spatiotemporal resolution enabling slip and grip for closed-loop robotics. *Sci. Adv.* **2020**, *6*, eabd7795. [CrossRef] [PubMed]
2. Park, J.; Jung, D.; Lee, Y.; Park, D.; Choi, J.; Kim, U. Tactile Sensor Integrated Fingertip Capable of Detecting Precise Contact Force for Robotic Grippers. *IEEE Trans. Ind. Electron.* **2024**, *72*, 5105–5115. [CrossRef]
3. Choi, J.; Han, C.; Lee, D.; Kim, H.; Lee, G.; Ha, J.-H.; Jeong, Y.; Ahn, J.; Park, H.; Han, H.; et al. Thermoforming 2D films into 3D electronics for high-performance, customizable tactile sensing. *Sci. Adv.* **2025**, *11*, eadv0057. [CrossRef] [PubMed]
4. Lee, J.I.; Shin, H.; Shikida, M.; Sato, K.J.M. Structural design of a polymer-based tactile sensor with a table-shaped sensing pad. *Sensors* **2014**, *127*, 1–6. [CrossRef]
5. Das, S.; Thiyagarajan, K.; Kodagoda, S.; Krishnan, A.; Bhattacharjee, M. A Novel Tactile Sensing Skin for Surface Perception. *IEEE Sens. J.* **2025**, *25*, 23155–23162. [CrossRef]
6. Ward-Cherrier, B.; Pestell, N.; Cramphorn, L.; Winstone, B.; Giannaccini, M.E.; Rossiter, J.; Lepora, N.F.J.S. The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies. *Soft Robot.* **2018**, *5*, 216–227. [CrossRef]
7. Trout, M.A.; Mino, F.R.; Olsen, C.D.; Hansen, T.C.; Teramoto, M.; Warren, D.J.; Segil, J.L.; George, J.A. Shared human-machine control of an intelligent bionic hand improves grasping and decreases cognitive burden for transradial amputees. *Nat. Commun.* **2025**, *16*, 10418. [CrossRef]
8. Wang, R.; Hu, S.; Zhu, W.; Huang, Y.; Wang, W.; Li, Y.; Yang, Y.; Yu, J.; Deng, Y. Recent progress in high-resolution tactile sensor array: From sensor fabrication to advanced applications. *Prog. Nat. Sci. Mater. Int.* **2023**, *33*, 55–66. [CrossRef]
9. Li, Q.; Kroemer, O.; Su, Z.; Veiga, F.F.; Kaboli, M.; Ritter, H.J. A Review of Tactile Information: Perception and Action Through Touch. *IEEE Trans. Robot.* **2020**, *36*, 1619–1634. [CrossRef]
10. Schmitz, A.; Maiolino, P.; Maggiali, M.; Natale, L.; Cannata, G.; Metta, G.J.I.T.o.R. Methods and technologies for the implementation of large-scale robot tactile sensors. *IEEE Trans. Robot.* **2011**, *27*, 389–400. [CrossRef]
11. Lin, W.; Wang, B.; Peng, G.; Shan, Y.; Hu, H.; Yang, Z. Skin-inspired piezoelectric tactile sensor array with crosstalk-free row+ column electrodes for spatiotemporally distinguishing diverse stimuli. *Adv. Sci.* **2021**, *8*, 2002817. [CrossRef]
12. Wang, H.; Ma, L.; Nie, Q.; Hu, X.; Li, X.; Min, R.; Wang, Z. Optical tactile sensor based on a flexible optical fiber ring resonator for intelligent braille recognition. *Opt. Express.* **2025**, *33*, 2512–2528. [CrossRef]
13. Jiang, C.; Xu, W.; Li, Y.; Yu, Z.; Wang, L.; Hu, X.; Xie, Z.; Liu, Q.; Yang, B.; Wang, X.J.N.C. Capturing forceful interaction with deformable objects using a deep learning-powered stretchable tactile array. *Nat. Commun.* **2024**, *15*, 9513. [CrossRef] [PubMed]
14. Funk, N.; Helmut, E.; Chalvatzaki, G.; Calandra, R.; Peters, J. Evetac: An event-based optical tactile sensor for robotic manipulation. *IEEE Trans. Robot.* **2024**, *40*, 3812–3832. [CrossRef]
15. Sun, H.; Kuchenbecker, K.J.; Martius, G. A soft thumb-sized vision-based sensor with accurate all-round force perception. *Nat. Mach. Intell.* **2022**, *4*, 135–145. [CrossRef]
16. Yuan, W.; Dong, S.; Adelson, E. GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force. *Sensors* **2017**, *17*, 2762. [CrossRef]
17. Lambeta, M.; Chou, P.-W.; Tian, S.; Yang, B.; Maloon, B.; Most, V.R.; Stroud, D.; Santos, R.; Byagowi, A.; Kammerer, G.J.I.R.; et al. Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3838–3845. [CrossRef]
18. Sferrazza, C.; D'Andrea, R. Sim-to-Real for High-Resolution Optical Tactile Sensing: From Images to Three-Dimensional Contact Force Distributions. *Soft Robot.* **2022**, *9*, 926–937. [CrossRef]
19. Kong, D.; Lu, Y.; Zhou, S.; Wang, M.; Pang, G.; Wang, B.; Chen, L.; Huang, X.; Lyu, H.; Xu, K. Super-resolution tactile sensor arrays with sparse units enabled by deep learning. *Sci. Adv.* **2025**, *11*, eadv2124. [CrossRef]
20. Sun, H.; Spiers, A.; Lee, H.; Fiene, J.; Martius, G. Sensing multi-directional forces at superresolution using taxel value isoline theory. *Nat. Commun.* **2025**, *16*, 8031. [CrossRef]
21. Sun, H.; Martius, G. Guiding the design of superresolution tactile skins with taxel value isolines theory. *Sci. Robot.* **2022**, *7*, eabm0608. [CrossRef]
22. Liu, X.; Yang, W.; Meng, F.; Sun, T. Measurement. Material recognition using robotic hand with capacitive tactile sensor array and machine learning. *IEEE/ASME Trans. Mechatron.* **2024**, *73*, 9508309.
23. Huh, T.M.; Choi, H.; Willcox, S.; Moon, S.; Cutkosky, M.R. Dynamically reconfigurable tactile sensor for robotic manipulation. *IEEE Robot. Autom. Lett.* **2020**, *5*, 2562–2569. [CrossRef]
24. Xu, C.; Wang, Y.; Zhang, J.; Wan, J.; Xiang, Z.; Nie, Z.; Xu, J.; Lin, X.; Zhao, P.; Wang, Y. Three-dimensional micro strain gauges as flexible, modular tactile sensors for versatile integration with micro-and macroelectronics. *Sci. Adv.* **2024**, *10*, eadp6094. [CrossRef] [PubMed]
25. Fernandez, A.J.; Weng, H.; Umbanhowar, P.B.; Lynch, K.M. Visiflex: A Low-Cost Compliant Tactile Fingertip for Force, Torque, and Contact Sensing. *IEEE Robot. Autom. Lett.* **2021**, *6*, 3009–3016. [CrossRef]

https://doi.org/10.3390/s26041378

26. Xin, Y.-H.; Hu, K.-M.; Xiang, R.-J.; Gao, Y.-L.; Zhou, J.-F.; Meng, G.; Zhang, W.-M. Vision-based tactile sensing: From performance parameters to device design. *Appl. Phys. Rev.* **2025**, *12*, 021312. [CrossRef]
27. Lee, S.; Jung, D.; Hwang, J.; Kim, U. Ultracompact Multiaxis Force Sensor Based on Sensitivity Amplification Mechanism for High-Sensitivity and Capacity Measurement. *IEEE/ASME Trans. Mechatron.* **2025**, *30*, 6951–6961. [CrossRef]
28. Piacenza, P.; Sherman, S.; Ciocarlie, M. Data-driven super-resolution on a tactile dome. *IEEE Robot. Autom. Lett.* **2018**, *3*, 1434–1441. [CrossRef]
29. Lee, J.I.; Kim, M.-G.; Shikida, M.; Sato, K. A table-shaped tactile sensor for detecting triaxial force on the basis of strain distribution. *Sensors* **2013**, *13*, 16347–16359. [CrossRef]
30. Mêda, F.; Näf, F.; Fernandes, T.P.; Bernardino, A.; Jamone, L.; Tavares, G.; Cardoso, S. Skin-Inspired Magnetoresistive Tactile Sensor for Force Characterization in Distributed Areas. *Sensors* **2025**, *25*, 3724. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

https://doi.org/10.3390/s26041378

sensors

MDPI

Article

A Cable-Driven Hybrid Robot with Series-Parallel Coupling: Design, Modeling, Optimization Analysis, and Trajectory Tracking

Zhifu Xue [1,†], Zhiquan Yang [1,†], Junyi Hu [1], Bin Zhu [1] and Jianqing Peng [1,2,*]

[1] School of Intelligent Systems Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, China; xuezhf@mail2.sysu.edu.cn (Z.X.); yangzhq33@mail2.sysu.edu.cn (Z.Y.); hujy68@mail2.sysu.edu.cn (J.H.); zhub26@mail.sysu.edu.cn (B.Z.)
[2] Guangdong Provincial Key Laboratory of Fire Science and Technology, Guangzhou 510006, China
* Correspondence: pengjq7@mail.sysu.edu.cn
† These authors contributed equally to this work.

Abstract

Compared to purely serial robots or cable-driven parallel robots (CDPRs), cable-driven hybrid robots (CDHRs) combine the advantages of both, addressing their limitations and enabling the execution of complex tasks. The series-parallel coupling structure increases the complexity of the system, complicating modeling, calibration, and force-closure workspace (FCW) analysis. This study develops a CDHR system equipped with various sensors and proposes methods for series-parallel coupling modeling, workspace analysis, and self-calibration of complex systems. First, the modular design requirements for the CDHR are analyzed, comprising an 8-cable parallel drive and a 4-degree-of-freedom serial manipulator. Second, a kinematic model of the CDHR with series-parallel coupling was derived, and the positions of the dynamic anchor seats were optimized using an optimization algorithm. Based on these optimized results, a modeling and analysis method for the statics and FCW is proposed. Furthermore, considering the complex and interdependent structural parameters of the system, a method for the self-calibration of the system parameters and trajectory planning for the CDHR is presented. Finally, experimental validation on both simulations and a physical prototype confirmed the effectiveness of the proposed methods. The developed prototype and the proposed method provide a basis for high-precision operations in large spaces, operations in dangerous/extreme environments, and automated operations in logistics/warehousing.

Keywords: cable-driven hybrid robot; force-closure workspace; series-parallel coupling; self-calibration; trajectory planning

1. Introduction

The cable-driven parallel robot (CDPR) replaces traditional rigid links with a cable traction mechanism, reduces structural inertia, and achieves a higher payload-to-weight ratio, increased speed, and an expanded workspace [1–3]. By incorporating a serial manipulator (SM) at the end, the degrees of freedom (DOFs) were further increased, allowing adaptation to diverse manipulation tasks. Thus, a cable-driven hybrid robot (CDHR) combines a CDPR with a serial manipulator (SM) [4]. This hybrid design overcomes the limitations of each individual robot and enables more complex tasks, offering a high payload-to-weight ratio, a large workspace, and high operational efficiency. Consequently, CDHRs have broad

application prospects in unstructured and complex environments such as disaster rescue, lifting operations, and large-scale equipment maintenance [5].

Significant progress has been made in the design and modeling of hybrid robots. As shown in Figure 1, a hybrid robot was formed by mounting an SM on the mobile platform of the CDPR [6,7]. This architecture combines the CDPR's large task-space workspace with the SM's high local end-effector dexterity [8–10]. As a result, it improves both global reach and fine motion capability near a given pose. However, this also makes its modeling and control more challenging. The kinematics of the CDHR can be decomposed into those of the CDPR and SM. Because the generalized joint DOFs are significantly higher than the 6-DOFs at the end effector, the kinematic solutions of the CDHR exhibit a high degree of redundancy.

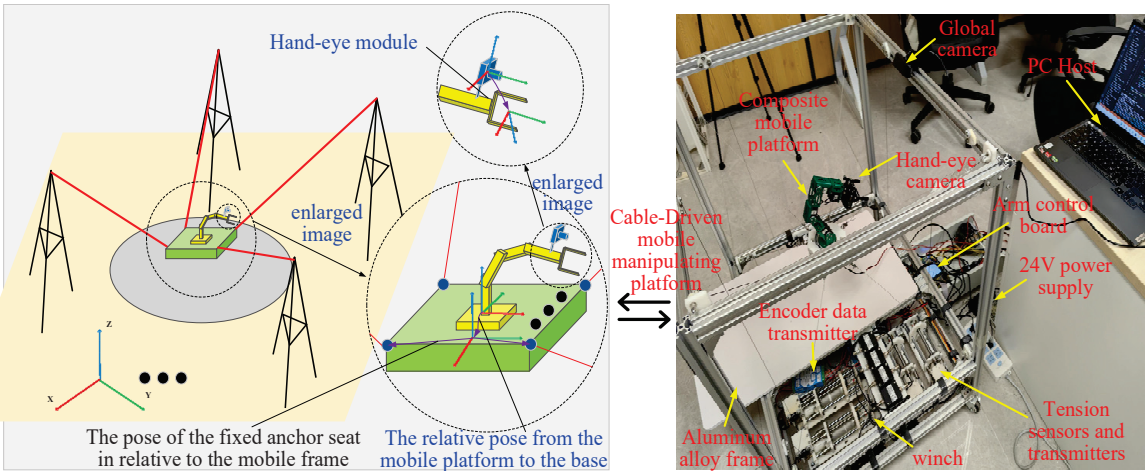


Figure 1. The overall structure of the cable-driven parallel robot.

Pinto et al. [11] designed a CDHR that attaches a rotary gripper as an end-effector tool beneath a mobile platform, thereby enhancing both the spatial range and flexibility of the object manipulation. Osumi et al. [1] proposed a planar CDHR structure that is suspended from a slider and achieves single-DOF reciprocating motion via a linear guide rail. A cable-driven platform was mounted below the slider and carried a 4-DOF SM. For this planar CDHR, an equivalent centroid modeling method was proposed [12]. In that model, cable torques and manipulator moments were neglected. El-Ghazaly et al. [13] designed a hybrid cable-thruster (HCT)-actuated underwater vehicle-manipulator system (UVMS) comprising an underwater operating platform and manipulator, and developed a dynamic modeling method that considers buoyancy and platform weight. Gao et al. [14] analyzed the kinematics of a spring-supported CDPR, integrating statics with inverse kinematics (IKs) to achieve rapid solutions. Lau et al. [15] released CASPR, an open-source software that incorporates various kinematic modeling methods for CDPRs.

To investigate the stability and performance metrics of CDPRs throughout the workspace, scholars have recently proposed methods based on Monte Carlo random sampling, convex hull theory, and other techniques to study force-closure workspaces (FCWs) [16,17]. The FCW of a CDPR is defined as the set of pose points within the kinematically reachable space, where the moving platform can remain stationary under cable tension [18]. In a six-dimensional Cartesian space, the generalized forces exerted by the cables include both forces and torques. Hence, the FCW is also referred to as the wrench-closure workspace (WCW). The FCW is a crucial performance metric for controlling CDHRs [19,20].

Pham et al. [15] proposed an FCW solution method based on the dimensionality reduction in the cable structure matrix using Gaussian elimination. This approach projects

https://doi.org/10.3390/s26041147

high-dimensional vector sets onto lower-dimensional ones and determines whether the moving platform can satisfy static equilibrium conditions by checking the force closure of all lower-dimensional vector sets in the projected space. Diao et al. [17] established the coupling relationship between the structure matrix and cable tension for a 7-DOF CDPR and calculated the FCW by evaluating the existence of solutions at various pose points. However, this method is only applicable to fully constrained 6-DOF CDPRs. Arai et al. [21] studied the workspace of a two-dimensional suspended hybrid manipulator, where the cable-driven platform provides tension to the base of the SM. The experimental results indicated that when the cable lengths were held fixed (i.e., the platform was not actuated), the end-effector's wrench-feasible workspace was reduced compared with the case of active cable adjustment. Yang et al. [18], using a high-dimensional vector decomposition approach, transformed the force-closure determination problem into a two-dimensional vector enclosure problem around the origin and voxelized the sampling space to solve for the FCW volume. They further introduced the Tension Factor (TF) and Global Tension Index (GTI) as indicators for evaluating workspace quality, which can serve as performance metrics for robot design. Osumi et al. [22] developed a planar CDHR system consisting of a suspended 4-DOF manipulator, where joint angles were constrained within certain ranges due to the hybrid mechanical structure, leading to partial workspace loss. From a review of FCWs for CDHRs, it is evident that there are currently few analysis methods that address workspaces with upper and lower bounds of cable tension or mechanical collision constraints [23]. Most previous studies have focused on planar CDHRs. Therefore, research on FCWs of CDHRs under multiple constraints and with higher DOFs is crucial.

After assembling the CDPR, system calibration is required for trajectory planning and control of the CDHR. Owing to the unknown initial cable lengths, the calibration model must incorporate unknown initial joint angles. The forward kinematics (FKs) of the CDHR cannot directly provide explicit results, nor can they be directly converted into a homogeneous transformation matrix (HTM). Therefore, the positions of the anchor seats must first be identified to obtain the HTM between moving and global frames.

Scholars have proposed traditional external instrument calibration and self-calibration methods to identify the position parameters of anchor seats in CDPRs [24]. Traditional calibration methods use high-precision measurement devices, such as inclinometers or laser rangefinders, to measure observable quantities in kinematic equations. The errors were then mapped to the kinematic parameters to be identified using the Jacobian matrix, enabling parameter correction. Self-calibration uses a robot's internal sensors to correct kinematic parameters. Borgstrom et al. [25] defined the anchor seat model as a pinhole model and formulated the identification model as a mathematical optimization problem, where the objective is to minimize the squared difference between the fitted and actual cable lengths based on the estimated coordinates of the anchor seats. However, this method does not consider the physical model of the cable-anchor seat connection. Zhang et al. [26] incorporated the effects of rotating pulleys at the anchor seats in their kinematic modeling of CDPRs and established a mapping between cable length errors and anchor seat position errors. They generated an identification matrix and continuously corrected the position errors of anchor seats using an iterative optimization method. Yuan et al. [25] developed a mapping relationship between motor angle errors and anchor seat errors, using an iterative optimization method to correct the position errors of the anchor seats. Wang et al. [27] abstracted kinematic parameter identification into a general calibration model and analyzed the impact of errors on the coefficient matrix of nonlinear equations based on matrix perturbation theory. They also introduced initial joint angles as unknown parameters in the calibration equations and experimentally validated their method on a 6-DOF CDPR. Ma et al. [28] proposed a probabilistic method for solving the "$\boldsymbol{AXB = YCZ}$"

https://doi.org/10.3390/s26041147

equation. Peng et al. [29] developed a hybrid calibration model for cable-driven hyper-redundant robots and solved it using a particle swarm optimization (PSO) method. If kinematic parameter calibration is conducted separately for each structure of the CDHR, each reconfiguration requires step-by-step calibration, resulting in high time costs. To achieve fast and accurate calibration, the challenge of simultaneously calibrating of multiple unknown mappings must be addressed.

As shown in Table 1, we summarize and compare different modeling, workspace analysis, and control methods. Although CDHRs have made some progress, several challenges remain. First, the strong series-parallel coupling of CDHRs complicates the resolution of their dynamics. Second, kinematic parameter identification for reconfigurable CDHR systems is complex and inefficient. Finally, there is complex force coupling between the CDPR and SM, where changes in the motion state of one affect the other nonlinearly, resulting in poor control accuracy and stability. To address these issues, this study designs a CDHR and derives a generalized pose kinematic and static model that considers the series-parallel coupling relationship. Furthermore, the FCW of the CDHR was solved based on this model and a study on simultaneous calibration and trajectory tracking control considering multiple mapping relationships was conducted.

Table 1. Comparison of Modeling, Calibration, and Control Methods.

Methods	Modeling	Working Space Analysis	Calibration	Control
Harms et al. [30]	Hybrid climbing CDPR; Kinematic & static models; Non-modular.	-	-	-
Michelin et al. [31]	6-DOF CDPR + 7-DOF arm; Redundancy used for obstacle avoidance; Not modular.	-	-	Redundancy-based path following; Coupling-aware tension regulation not considered.
Song et al. [32]	Multi-pattern hybrid CDPR; Geometric and dynamic models; Parallel–serial force coupling neglected.	Geometric workspace analysis for two modes; No movable-anchor optimization.	-	-
An et al. [9]	CDPR with movable anchors (M-CDPR); Kinematics	Movable anchors; ≈456% increase in feasible workspace	-	Obstacle-aware motion planning; Parallel-only, no hybrid coupling control
Borgström et al. [25]	NIMS3D CDPR; Full kinematic	-	In-field self-calibration for Rapid deployment	Feedback tracking with optimized tension distribution

The main contributions of this study are as follows:

(1) A modular, highly integrated, and compact prototype of a CDHR was developed that is capable of rapid self-reconfiguration and cable tension control. Unlike traditional designs, this system utilizes a distributed control architecture (STM32 and Arduino) with integrated tension and angle sensors, enabling effective cable tension management and creating a physical platform for validating complex coupling operations.

https://doi.org/10.3390/s26041147

(2) A kinematic and static modeling method for the series-parallel hybrid structure of the CDHR was proposed. Unlike traditional approaches that often simplify inter-component dynamics, our model meticulously accounts for the strong nonlinear coupling effects between the cable tensions, the platform pose, and the serial manipulator's joint configurations. This involved deriving explicit expressions for both the forward and inverse kinematics that consider the variable anchor points and the cable length constraints through iterative methods, coupled with a static model that accurately predicts the force distribution within the redundant cabling system.

(3) A workspace analysis method and dynamic anchor seat position optimization for the CDHR were introduced. We proposed a coupled workspace analysis method utilizing Gaussian boundary enrichment to precisely determine the Force-Closure Workspace (FCW) under complex series-parallel constraints. Furthermore, a dynamic anchor seat optimization strategy based on the global condition number was implemented using the Particle Swarm Optimization (PSO) algorithm, significantly enhancing the robot's kinematic isotropy and minimizing collision risks.

(4) A self-calibration method for the structural parameters and a trajectory tracking method for the CDHR were presented. We developed a vision-based self-calibration method that simultaneously identifies the fixed anchor seat coordinates, the relative pose of the serial manipulator, and hand-eye parameters. Experimental validation demonstrated that this method reduces the maximum position error to 0.10 mm and the orientation error to 0.07°, significantly improving the initialization efficiency and static accuracy of the system compared to step-by-step calibration approaches.

The remainder of this paper is organized as follows. Section 2 introduces the mechanical design and system setup of the prototype. Section 3 derives the mathematical model of the series-parallel coupling structure. Sections 4 and 5 analyzes the workspace of the CDHR and optimizes the position of dynamic anchor seats using an optimization algorithm. Sections 6 and 7 focuses on the self-calibration of the structural parameters of the system and conducts trajectory-tracking control based on calibration results. Section 8 presents the simulations and experiments to validate the methods and prototypes described in this paper. The final section provides conclusions and outlines future work.

2. Design of the Hybrid Robot

2.1. Function Analysis and Prototype Design

To satisfy the application requirements of large-space movement and flexible operation in local spaces for complex scenarios, a CDHR prototype with an 8-cable parallel drive and a 4-DOF serial joint linkage was designed. To enhance the global flexibility of the moving platform within the workspace and the dynamic collision constraints between the cables and the robot itself, the distribution of the cable connection anchor points on the moving platform was optimized, and a new type of moving platform with an SM mounted at the end effector was designed. The hybrid robot is structured into two spatial layers: the upper and lower layers are used for motion demonstration and hardware placement, respectively, eliminating the need for an additional drive box and improving the space utilization. Each winch module is equipped with a tension sensor and an angle sensor to achieve cable pretension control, overload alarms, and cable position feedback.

This study analyzed the functionality of the prototype across three modules: hardware design, algorithm design, and system design. In the hardware design module, the need for large-range spatial movement of the moving platform is addressed by designing an 8-cable-driven hybrid platform. Simultaneously, to perform complex operations in local spaces, a flexible manipulator was mounted on a moving platform with a vision sensor installed at its end. In the algorithm design module, the coupling forces between the

https://doi.org/10.3390/s26041147

moving platform and manipulator base must be calculated, and a coupled kinematic model is established. In addition, a cable tension control algorithm was designed to maintain the stability of the platform and cables. In the system design module, the lower-level controller reads the sensor data, controls the motors, and communicates with the upper-level controller, which sets up the simulation environment, develops a visual user interface, and communicates with the lower-level controller. Based on these considerations, the designed CDHR prototype is illustrated in Figure 2. The prototype has overall dimensions of 860 mm × 705 mm × 850 mm and consists of a frame, hybrid moving platform, motors (i.e., MAXON RE 25 [M10.1], **maxon motor ag, Sachseln, Switzerland**), drivers (i.e., Robomodule, **Shenzhen RoboModule Technology Co., Ltd., Shenzhen, China**), tension sensors (i.e., JZHL-M1, **Hefei Jinnuo Company, Hefei, China**), angle sensors (i.e., PD-1503-ENC-1024, **Dongguan Pudi Electronic Technology Co., Ltd., Dongguan, China**), transmitters (i.e., IBF166-WIFI-N, **Shenzhen Beifu Technology Co., Ltd., Shenzhen, China**), and winches. The prototype was equipped with eight driving cables and four serial joints to achieve 6-DOF motion by employing a distributed control method to drive the robot. The motors drive cable movements are located at the base of the prototype, whereas the motors that drive the manipulator on the moving platform are housed within the manipulator itself, thereby reducing the load on the moving platform through a separate arrangement. The hybrid moving platform is driven by cables, whereas the motors on the platform control the movement of the manipulator, enabling tasks such as recognition and grasping.

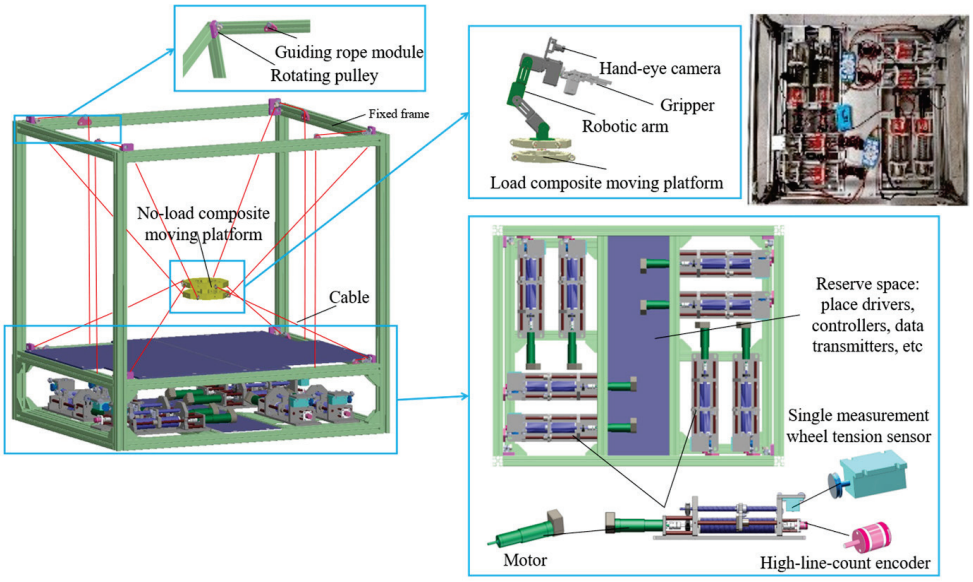


Figure 2. Three-dimensional structure of the prototype.

The moving platform in this prototype was driven using eight cables, with each wound around a winch located on the base. Cables were drawn from the winches and connected to the hybrid moving platform through a cable guide module and rotating pulleys. During the movement of the hybrid platform, the motors on the base drove the rotation of the winch drums, allowing the cables to be retracted or released. The angle sensors installed on the base provide feedback regarding the rotation angles of the drums, which can be used to determine the length of the cable that is retracted or released. Tension sensors provide feedback on the forces exerted on the cables. When the current cable lengths are known, the pose of the moving platform can be calculated using a fixed coordinate system. Conversely, when the desired pose of the moving platform in a fixed coordinate system is known, the corresponding cable length required to reach this pose can be determined. The effects of cable mass, inertia, and damping on solution accuracy are minimal; thus, it is assumed

https://doi.org/10.3390/s26041147

that the cables are subjected only to axial tension. Therefore, the movement of the hybrid platform can be controlled using feedback from cable lengths and tension forces.

2.2. Hardware and Software System Design

The entire hybrid robot system consists of an upper-and lower-level controller with sensor data acquisition based on the high-speed RS485 serial bus standard, ensuring that the data feedback cycle is significantly shorter than the control cycle. The software and hardware design frameworks of the system are shown in Figure 3. The upper-level software, including the simulation and prototype controls, was built using the PyQt5 environment with the control interface illustrated in Figure 3a. The simulation control software communicates with CoppeliaSim to monitor robot states and control movements, whereas the prototype control software communicates with the lower-level controller board to read, display, and detect sensor data. It controls the cable-driven motors of the parallel platform through serial communication from the upper-level controller and the drive motors of the serial manipulator through Ethernet communication, thereby achieving comprehensive control of the prototype. On the hardware side, the cable drives were controlled using an STM32F4 development board, whereas the movement of the manipulator was controlled using an Arduino development board. To control information feedback, tension sensors use transmitters to limit the output voltage signals to a range of 0~3.3 V. These signals were inputted to the ADC channels of STM32F4, where the voltage values were converted into cable tension through a mapping relationship. Angle sensors perform quadrature decoding of their output signals via transmitters, with data transmission between the transmitters and STM32F4 utilizing the RS485 communication protocol.

To analyze the impact of kinematic parameters on the workspace of the CDHR, a modeling and analysis method for kinematics, statics, and workspace was proposed, considering a series-parallel coupling structure. At the kinematic level, the relative position between the base of the SM and moving platform was considered. At the static level, a mechanical model accounting for the force coupling between the moving platform and base was established. Based on this static analysis, a comprehensive examination of the FCW of the CDHR was conducted. To achieve precise movement and autonomous grasping operations, a self-calibration and grasping control strategy based on visual positioning was proposed. After assembling the robot, a vision-based self-calibration method was developed to reduce the complexity of the calibration process during operation. This method simultaneously calibrates the anchor seat positions and the relative position between the moving platform and the base using visual detection data and sensor feedback from the robot. Based on the calibration results, upper-level software was developed to simulate grasping control for typical tasks. The proposed self-calibration and grasping control methods were validated using simulation software, and the experimental results demonstrated their effectiveness and accuracy. Finally, the performances of the proposed methods and prototype were validated through calibration experiments conducted on the CDHR prototype. The experiments included sensor parameter calibration, motor reduction ratio calibration, and system self-calibration, which enhanced the motion accuracy of the prototype. Additionally, cable tension control and complex trajectory tracking experiments were conducted to verify the effectiveness of the control strategies and motion performance of the robot.

https://doi.org/10.3390/s26041147

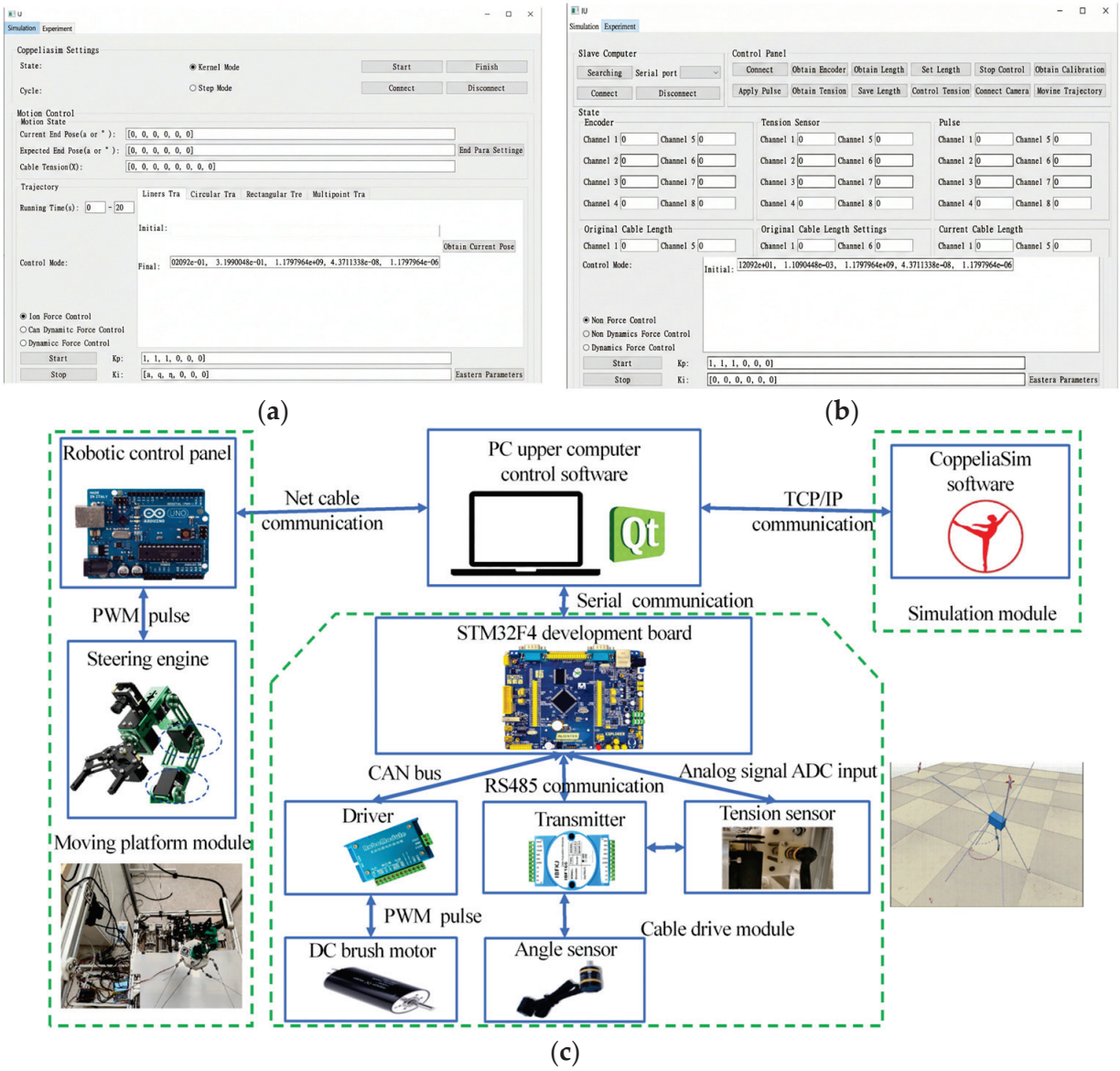


Figure 3. Framework of hardware and software design for hybrid cable-driven robot systems. (**a**) Trajectory tracking control interface of the upper computer, (**b**) communication interface control between the upper and lower computers, and (**c**) interaction interface between the upper and lower computers.

3. Coupled Kinematic and Static Modeling

The model parameters of the CDHR are defined in Table 2, and those of the CDHR with $(D_c + D_s)$-DOFs are shown in Figure 4. It is assumed that all winches of the CDHR are equal.

Table 2. Key symbols and variables.

Variable	Meaning
$\{G\}$, $\{M\}$, $\{S_j\}$, $\{B\}$, $\{E\}$	The global frame of the CDHR, the mobile platform frame, the *j*th joint frame of the SM, the base frame of the SM mounted on the moving platform, the tips frame for the CDHR
O_g, O_m, O_b, O_{sj}, O_e	The origin of the CDHR global frame, the origin of the moving platform frame, the origin of the base frame for the SM mounted on the moving platform, the origin of the *j*th joint frame for the SM, the origin of the tips frame for the CDHR

https://doi.org/10.3390/s26041147

Table 2. *Cont.*

Variable	Meaning
D_c	Number of cables connected to the moving platform
D_s	Number of joints of the SM mounted on the moving platform
D_m	The DOFs of the moving platform
D_h	Generalized DOFs of the CDHR
D_e	The DOFs of the end-effector for the CDHR
$i = 1 \sim D_c$	Index related to the number of cables
$j = 1 \sim D_s$	Index related to the SM
A_i	Anchor seat connecting the moving platform to the ith cable (dynamic anchor seat)
B_i	Anchor seat connecting the frame to the ith cable (fixed anchor seat)
$\boldsymbol{p}_{bi}$	Position vector of the fixed anchor seat relative to the global frame
$\boldsymbol{r}_{ai}$	Position vector of the dynamic anchor seat relative to the mobile frame
l_i	Length of the ith cable
$\boldsymbol{L}_i$	The direction vector of the ith cable, from the dynamic anchor seat to the fixed anchor seat
$\hat{\boldsymbol{L}}_i$	The unit direction vector of the ith cable
$\boldsymbol{l} = [l_1, l_2, \ldots, l_{D_c}]^T$	The vector containing the lengths of all cables
$\boldsymbol{L} = [\boldsymbol{L}_1, \boldsymbol{L}_2, \ldots, \boldsymbol{L}_{D_c}]^T$	The matrix containing the direction vectors of all cables
t_{ci}	Tension magnitude of the ith cable
$\boldsymbol{T}_{ci}$	Tension vector of the ith cable
$\boldsymbol{M}_{ci}$	Moment caused by the tension of the ith cable
$\boldsymbol{p}_m$	Position vector of the origin Om of the moving frame relative to the global frame {G}
θ_j	The joint angle of the jth joint of the SM
$\boldsymbol{\Theta}_s = [\theta_1, \theta_2, \ldots, \theta_{D_S}]^T$	The vector containing the joint angles of all joints in the SM
$\boldsymbol{\Theta}_w = [\theta_{w1}, \theta_{w2}, \ldots, \theta_{wD_c}]^T$	The vector containing the rotation angles of all motors in the CDPR
$\boldsymbol{X}_m = [x_m, y_m, z_m, \alpha_m, \beta_m, \gamma_m]^T$	Pose of the moving platform frame relative to the global frame

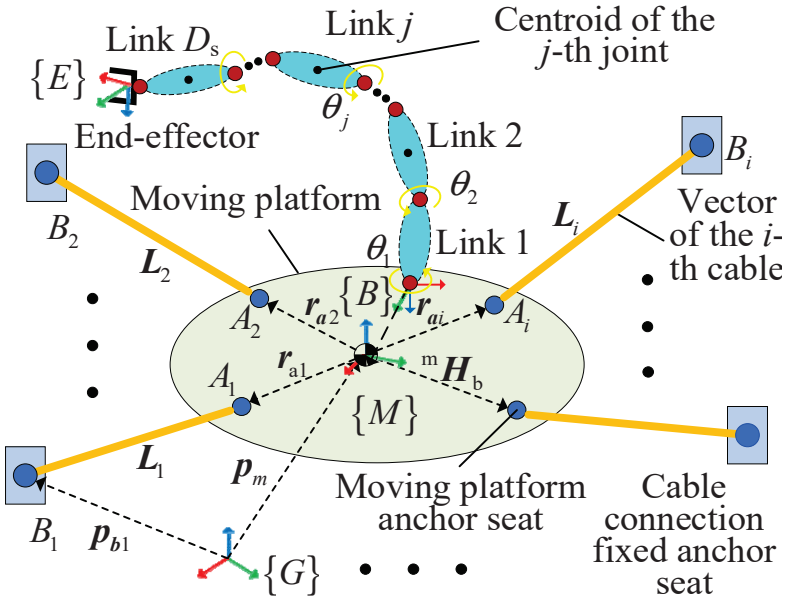


Figure 4. Structure of the CDHR.

3.1. Motor-to-Cable FKs

The relationship between the motor rotation angle and the driver pulse value is:

$$\Delta\theta_w = \frac{p_d}{4l_m r_m} \cdot 2\pi \tag{1}$$

https://doi.org/10.3390/s26041147

where p_{d} is the number of driver pulses, l_{m} is the number of encoder lines, and r_{m} is the motor gearbox-reduction ratio.

Because of the synchronous rotation of the motor output shaft and drum through coupling, the lead of the helical line on the drum, height along the axis of the cylinder, and circumference of the cross-section form a right triangle. Consequently, the relationship between the change in cable length Δl and motor rotation angle can be expressed as:

$$\Delta l = \rho_{\mathrm{w}} \cdot \Delta\theta_{\mathrm{w}} \tag{2}$$

where $\rho_{\mathrm{w}} = \frac{\sqrt{(\pi d_{\mathrm{w}})^2 + p_{\mathrm{w}}^2}}{2\pi}$ is the cable length for one winch rotation, d_{w} is the cross-sectional diameter of the cylindrical drum, p_{w} is the pitch of the helical groove on the cylindrical drum, $\Delta\theta_{\mathrm{w}}$ is the motor rotation angle.

We assume that $\bar{l}_i$ is the initial length of the ith cable, $\overline{\theta}_{\mathrm{w}i}$ is the corresponding motor rotation angle, and $\theta_{\mathrm{w}i}$ is the current motor rotation angle. The cable length between the fixed anchor point at the directional pulley and anchor seat on the mobile platform is determined by:

$$l_i = \bar{l}_i + \left(\theta_{\mathrm{w}i} - \overline{\theta}_{\mathrm{w}i}\right) \cdot \rho_{\mathrm{w}} \tag{3}$$

Equation (3) can be rewritten in vector form as follows:

$$\boldsymbol{l} = \bar{\boldsymbol{l}} + \rho \cdot \left(\boldsymbol{\Theta}_{\mathrm{w}} - \overline{\boldsymbol{\Theta}}_{\mathrm{w}}\right) \tag{4}$$

where $\bar{\boldsymbol{l}} = \left[\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_{D_{\mathrm{c}}}\right]^{\mathrm{T}}$ is the vector of the initial lengths of all cables, and $\overline{\boldsymbol{\Theta}}_{\mathrm{w}} = \left[\overline{\theta}_{\mathrm{w}1}, \overline{\theta}_{\mathrm{w}2}, \ldots, \overline{\theta}_{\mathrm{w}D_{\mathrm{c}}}\right]^{\mathrm{T}}$ represents the initial values of $\boldsymbol{\Theta}_{\mathrm{w}}$.

3.2. Cable-to-End FKs

Owing to the presence of force coupling, changes in the joint angles of the SM alter the interaction forces between its links, leading to changes in the coupling forces between the manipulator and moving platform system. This in turn can cause the moving platform to experience vibrations or oscillations, resulting in changes in the robot's end-effector pose and affecting the robot's motion accuracy. The generalized joints and generalized pose of the CDHR can be defined as:

$$\boldsymbol{Q}_{\mathrm{gj}} = \left[\overbrace{\underbrace{\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_i, \ldots, \bar{l}_{D_{\mathrm{c}}}}_{\bar{\boldsymbol{l}}}}^{\text{cable length}}, \overbrace{\underbrace{\theta_1, \theta_2, \ldots, \theta_j, \ldots, \theta_{D_{\mathrm{s}}}}_{\boldsymbol{\Theta}_{\mathrm{s}}}}^{\text{joint angles}}\right]^{\mathrm{T}} \in \mathbb{R}^{D_{\mathrm{c}}+D_{\mathrm{s}}} \tag{5}$$

$$\boldsymbol{Q}_{\mathrm{gp}} = \left[\overbrace{\underbrace{x_{\mathrm{m}}, y_{\mathrm{m}}, z_{\mathrm{m}}, \alpha_{\mathrm{m}}, \beta_{\mathrm{m}}, \gamma_{\mathrm{m}}}_{\boldsymbol{X}_{\mathrm{m}}}}^{\text{platform pose}}, \overbrace{\underbrace{\theta_1, \theta_2, \ldots, \theta_j, \ldots, \theta_{D_{\mathrm{s}}}}_{\boldsymbol{\Theta}_{\mathrm{s}}}}^{\text{joint angles}}\right]^{\mathrm{T}} \in \mathbb{R}^{6+D_s} \tag{6}$$

Based on the geometric relationship between the pulley and cable, the cable length is divided into two parts: a straight segment and an arc segment. The straight segment is the line between the tangent point at which the cable leaves the rotating pulley and anchor seat on the moving platform. The arc segment is a circular arc in which the cable wraps around the rotating pulley. The lengths of the straight and arc segments can be expressed as:

$$(l_{i,s}, l_{i,a}) = f_{\mathrm{cdpr_ik}}(\boldsymbol{p}_{bi}, \boldsymbol{r}_{ai}, R_r, \boldsymbol{n}_i, \boldsymbol{X}_{\mathrm{m}}) \tag{7}$$

where $l_{i,s}$ is the length of the straight segment, $l_{i,a}$ is the length of the arc segment, $l_i = l_{i,s} + l_{i,a}$, R_r is the radius of the rotating pulley, and $\boldsymbol{n}_i$ is the axis of rotation of the pulley.

https://doi.org/10.3390/s26041147

According to Equation (7), the optimization model for the cable length can be established as:

$$\begin{aligned} \boldsymbol{X}_{\mathrm{m}} &= \underset{\boldsymbol{\Theta}_{\mathrm{s}} \in \Phi}{\operatorname{argmin}} \left(\boldsymbol{l}^{r} - \boldsymbol{l}^{f} \right) \\ &= \underset{\boldsymbol{\Theta}_{\mathrm{s}} \in \Phi}{\operatorname{argmin}} \left(\boldsymbol{l}^{r} - f_{\mathrm{cdpr_fk}} \left(\boldsymbol{p}_{b}, \boldsymbol{r}_{a}, R_{r}, \boldsymbol{n}, \boldsymbol{X}_{\mathrm{m}}^{f} \right) \right) \end{aligned} \tag{8}$$

where $\boldsymbol{l}^{r}$ represents the actual length of the cable, $\boldsymbol{l}^{r} = \left[l_{1}^{r}, l_{2}^{r}, \ldots, l_{D_c}^{r} \right]^{\mathrm{T}}$, $\boldsymbol{l}^{f}$ is the iterative solution value of the cable length, and $\boldsymbol{l}^{r} - \boldsymbol{l}^{f}$ is the residual error. $\boldsymbol{X}_{\mathrm{m}}^{f}$ is the iterative solution value of the pose of the moving platform. Φ is the admissible set of the serial arm joint vector. As shown in Figure 5, using the Jacobian matrix derived from the cable lengths with respect to the pose of the moving platform, the cable length error is mapped to the pose error of the moving platform, thereby determining the actual pose.

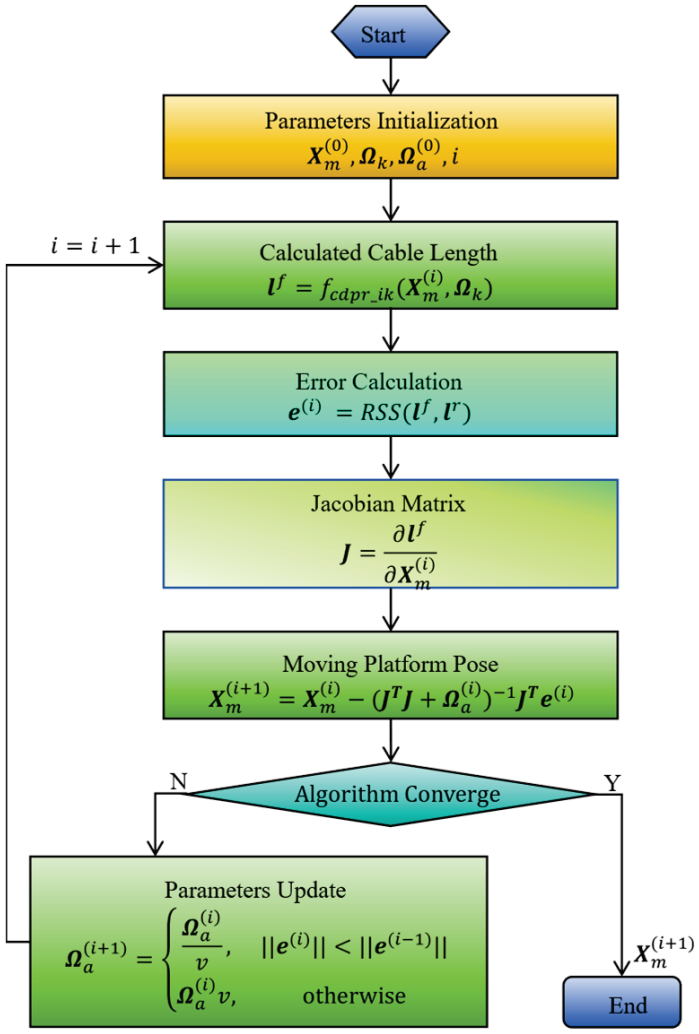


Figure 5. The pose resolving process of the moving platform based on the Levenberg–Marquart (LM) method. RSS(·,·) denotes the residual sum of squares. $\boldsymbol{X}_m^{(i)}$ demotes the pose of the moving platform at step i. $\boldsymbol{\Omega}_a^{(i)}$ demotes the algorithm parameters at step i. $\boldsymbol{e}^{(i)}$ denotes the error at step i. $\boldsymbol{l}^r$ denotes the actual cable length. $\boldsymbol{\Omega}_k$ demotes the kinematic parameters. v denotes the adjustment factor.

Furthermore, the HTM of the CDHR end-effector frame relative to the SM base frame is:

$$^{\mathrm{b}}\boldsymbol{T}_{\mathrm{e}} = {}^{0}\boldsymbol{T}_{D_{\mathrm{s}}} = {}^{0}\boldsymbol{T}_{1} \cdot {}^{1}\boldsymbol{T}_{2} \ldots {}^{D_{\mathrm{s}}-1}\boldsymbol{T}_{D_{\mathrm{s}}} \tag{9}$$

where ${}^{k}T_{k+1}$ is the HTM from the ith joint to the $(i + 1)$-th joint.

https://doi.org/10.3390/s26041147

Based on Equation (9), we obtain:

$$^{\mathrm{g}}\boldsymbol{T}_{\mathrm{e}} = {}^{\mathrm{g}}\boldsymbol{T}_{\mathrm{m}} \cdot {}^{\mathrm{m}}\boldsymbol{T}_{\mathrm{b}} \cdot {}^{\mathrm{b}}\boldsymbol{T}_{\mathrm{e}} \tag{10}$$

where $^{\mathrm{m}}\boldsymbol{T}_{\mathrm{b}}$ is the HTM from the moving platform frame to the base frame of the SM, and $^{\mathrm{g}}\boldsymbol{T}_{\mathrm{m}}$ is the HTM from the global frame to the moving platform frame.

Therefore, the IKs equation of the SM can be expressed as:

$$\dot{\boldsymbol{\Theta}}_{\mathrm{s}} = \boldsymbol{J}_{\mathrm{s}}^{+}(\boldsymbol{\Theta}_{\mathrm{s}}) \cdot \dot{\boldsymbol{X}}_{\mathrm{e}} \tag{11}$$

where $\boldsymbol{J}_{\mathrm{s}}^{+}(\boldsymbol{\Theta}_{\mathrm{s}})$ represents the Moore–Penrose pseudoinverse of $\boldsymbol{J}_{\mathrm{s}}(\boldsymbol{\Theta}_{\mathrm{s}})$, $\boldsymbol{J}_{\mathrm{s}}(\boldsymbol{\Theta}_{\mathrm{s}}) = \left[\boldsymbol{J}_1(\theta_1), \boldsymbol{J}_2(\theta_2), \ldots, \boldsymbol{J}_{D_{\mathrm{s}}}(\theta_{D_{\mathrm{s}}})\right]$, $\boldsymbol{J}_{\mathrm{s}}(\boldsymbol{\Theta}_{\mathrm{s}})$ is the Jacobian matrix of the SM mapping from the joint space to the operational space, and $\dot{\boldsymbol{X}}_{\mathrm{e}}$ is the derivative of the end-effector pose in the global frame.

By combining Equations (8), (9) and (11), the pose $\boldsymbol{X}_{\mathrm{m}}$ of the moving platform and $\boldsymbol{Q}_{\mathrm{gp}} = \begin{bmatrix}\boldsymbol{X}_{\mathrm{m}}^{\mathrm{T}} & \boldsymbol{\Theta}_{\mathrm{s}}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$ can be obtained. Therefore, the FKs equation of the CDHR can be expressed as:

$$\boldsymbol{X}_{\mathrm{e}} = \mathrm{CFK}\left(\boldsymbol{p}_{bi}, \boldsymbol{r}_{ai}, {}^{\mathrm{m}}\boldsymbol{T}_{\mathrm{b}}, \boldsymbol{Q}_{\mathrm{gj}}\right) \tag{12}$$

where $\mathrm{CFK}(\cdot)$ represents the coupled model of the FKs of the CDHR.

3.3. IKs of the CDHR

Similarly, defining the coordinates of the SM base relative to the global frame as $\boldsymbol{X}_{\mathrm{b}} = [x_{\mathrm{b}}, y_{\mathrm{b}}, z_{\mathrm{b}}, \alpha_{\mathrm{b}}, \beta_{\mathrm{b}}, \gamma_{\mathrm{b}}]^{\mathrm{T}}$, and by combining Equations (9) and (10), the HTM of the base relative to the global frame can be written as:

$$^{\mathrm{g}}\boldsymbol{T}_{\mathrm{b}} = {}^{\mathrm{g}}\boldsymbol{T}_{\mathrm{e}} \cdot {}^{0}\boldsymbol{T}_{D_{\mathrm{s}}}^{-1} \tag{13}$$

Because the SM is rigidly connected to the moving platform, after the design, manufacturing, and assembly of the moving platform and the attached manipulator, the relative pose relationship $^{\mathrm{b}}\boldsymbol{T}_{\mathrm{m}}$ remains constant after calibration. Therefore, the HTM of the moving platform relative to the global frame can be expressed as:

$$^{\mathrm{g}}\boldsymbol{T}_{\mathrm{m}} = {}^{\mathrm{g}}\boldsymbol{T}_{\mathrm{b}} \cdot {}^{\mathrm{b}}\boldsymbol{T}_{\mathrm{m}} = {}^{\mathrm{g}}\boldsymbol{T}_{\mathrm{b}} \cdot {}^{\mathrm{m}}\boldsymbol{T}_{\mathrm{b}}^{-1} \tag{14}$$

As observed from Equation (14), the pose of the moving platform depends solely on the base pose, which is determined by the joint angle. Once $^{\mathrm{g}}\boldsymbol{T}_{\mathrm{m}}$ is obtained, the relative pose $\boldsymbol{X}_{\mathrm{m}}$ can be further derived.

Based on the above analysis, the pose estimation of the base and moving platforms can be formulated as an optimization problem. In this study, the objective function of the optimization equation was defined as the weighted sum of the deviations between the actual and desired joint angles of the arm. The deviations between the actual and desired rotational poses of the moving platform can be expressed as:

$$\begin{aligned}&\left(\boldsymbol{X}_{\mathrm{m}}, \boldsymbol{X}_{\mathrm{b}}, {}^{\mathrm{g}}\boldsymbol{T}_{\mathrm{m}}, {}^{\mathrm{g}}\boldsymbol{T}_{\mathrm{b}}, \boldsymbol{\Theta}_{\mathrm{s}}\right)\\ &= \underset{\boldsymbol{X}_{\mathrm{b}} \in \mathbb{X}, \boldsymbol{\Theta}_{\mathrm{s}} \in \Phi}{\operatorname{argmin}} \lambda_1 \sum_{j}^{D_{\mathrm{s}}} \left|\theta_j^{\mathrm{des}} - \theta_j\right| + \lambda_2\left(\left|\alpha_{\mathrm{m}}^{\mathrm{des}} - \alpha_{\mathrm{m}}\right| + \left|\beta_{\mathrm{m}}^{\mathrm{des}} - \beta_{\mathrm{m}}\right|\right)\end{aligned} \tag{15}$$

where $\alpha_{\mathrm{m}}^{\mathrm{des}}$ and $\beta_{\mathrm{m}}^{\mathrm{des}}$ are the desired rotational angles of the moving platform along the x-axis and y-axis, respectively, θ_j^{des} represents the desired joint angle of the jth joint of the manipulator, and λ_1, λ_2 are the weighting coefficients.

https://doi.org/10.3390/s26041147

Furthermore, based on Equation (7), the cable length l can be determined, and the coupled model of IKs for the CDHR can be described as

$$\left(\boldsymbol{Q}_{\mathrm{gj}}, \boldsymbol{Q}_{\mathrm{gp}}\right) = \mathrm{CIK}\left(\boldsymbol{p}_{bi}, \boldsymbol{r}_{ai}, \boldsymbol{\Theta}_{\mathrm{s}}, {}^{\mathrm{b}}\boldsymbol{T}_{\mathrm{m}}, N_{\mathrm{sample}}\right) \tag{16}$$

where $\mathrm{CIK}(\cdot)$ represents the coupled IKs equation.

3.4. Static Modeling

According to the principles of mechanics, the static equilibrium equation of a moving platform can be expressed as:

$$\begin{cases} \sum\limits_{i=1}^{D_{\mathrm{c}}} \boldsymbol{T}_{\mathrm{c}i} + \boldsymbol{G}_{\mathrm{p}} + \boldsymbol{F}_{\mathrm{em}} = 0 \\ \sum\limits_{i=1}^{D_{\mathrm{c}}} \boldsymbol{M}_{\mathrm{c}i} + \boldsymbol{M}_{\mathrm{em}} = 0 \end{cases} \Rightarrow \begin{cases} \sum\limits_{i=1}^{D_{\mathrm{c}}} t_{\mathrm{c}i}\hat{\boldsymbol{L}}_i + \boldsymbol{G}_{\mathrm{p}} + \boldsymbol{F}_{\mathrm{em}} = 0 \\ \sum\limits_{i=1}^{D_{\mathrm{c}}} \boldsymbol{r}_{ai} \times \left(t_{\mathrm{c}i}\hat{\boldsymbol{L}}_i\right) + \boldsymbol{M}_{\mathrm{em}} = 0 \end{cases} \tag{17}$$

where $\boldsymbol{G}_{\mathrm{p}}$ is the gravitational force acting on the moving platform, $\boldsymbol{F}_{\mathrm{em}}$ is the external force applied to the moving platform, $\boldsymbol{M}_i$ is the moment exerted on the moving platform by the tension in the ith cable, and $\boldsymbol{M}_{\mathrm{em}}$ is the external moment acting on the moving platform.

However, because there is no relative motion between the base of the SM and the moving platform, the force exerted by the link one on the base is equivalent to the force exerted by link one on the moving platform. In the moving coordinate system, the force-coupling relationship between the base and moving platform can be expressed as:

$$\begin{aligned} &\boldsymbol{F}_{\mathrm{em}} = \boldsymbol{F}_{\mathrm{s}} = -{}^{m}\boldsymbol{R}_{\mathrm{b}}{}^{0}\boldsymbol{f}_1 \\ &\boldsymbol{M}_{\mathrm{em}} = \boldsymbol{M}_{\mathrm{s}} = {}^{m}\boldsymbol{R}_{\mathrm{b}}{}^{0}\boldsymbol{\tau}_1 - {}^{m}\boldsymbol{t}_{\mathrm{b}} \times ({}^{m}\boldsymbol{R}_{\mathrm{b}}{}^{0}\boldsymbol{f}_1) \\ &\boldsymbol{W}_{\mathrm{em}} = \boldsymbol{W}_{\mathrm{s}} = \begin{bmatrix} \boldsymbol{F}_{\mathrm{em}}^{\mathrm{T}} \\ \boldsymbol{M}_{\mathrm{em}}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F}_{s}^{\mathrm{T}} \\ \boldsymbol{M}_{s}^{\mathrm{T}} \end{bmatrix} \end{aligned} \tag{18}$$

where $\boldsymbol{F}_{\mathrm{s}}$ and $\boldsymbol{M}_{\mathrm{s}}$ represent the force and moment exerted by SM on the moving platform in frame $\{B\}$, respectively. $\boldsymbol{W}_{\mathrm{s}}$ denotes the generalized force applied by SM on the moving platform in frame $\{B\}$. ${}^{\mathrm{m}}\boldsymbol{R}_{\mathrm{b}}$ and ${}^{\mathrm{m}}\boldsymbol{t}_{\mathrm{b}}$ represent the rotation matrix and the position vector of $\{B\}$ relative to $\{M\}$, respectively.

Substituting Equation (17) into Equation (18), we obtain:

$$\begin{cases} \sum\limits_{i=1}^{D_{\mathrm{c}}} t_{\mathrm{c}i}\hat{\boldsymbol{L}}_i + \boldsymbol{G}_{\mathrm{p}} - {}^{\mathrm{m}}\boldsymbol{R}_{\mathrm{b}}{}^{0}\boldsymbol{f}_1 = 0 \\ \sum\limits_{i=1}^{D_{\mathrm{c}}} \boldsymbol{r}_{ai} \times \left(t_{\mathrm{c}i}\hat{\boldsymbol{L}}_i\right) + {}^{\mathrm{m}}\boldsymbol{R}_{\mathrm{b}}{}^{0}\boldsymbol{\tau}_1 - {}^{\mathrm{m}}\boldsymbol{t}_{\mathrm{b}} \times {}^{0}f_1 = 0 \end{cases} \tag{19}$$

Equation (19) represents the static equilibrium condition of the CDHR. Therefore, the optimization model can be established with the objective of minimizing cable tension, thus:

$$\begin{aligned} &f = \tfrac{1}{2}\min \sum_{i=1}^{D_{\mathrm{c}}} {t_{\mathrm{c}i}}^2 \\ &\text{s.t.} \quad \boldsymbol{A}_{\mathrm{p}}\boldsymbol{t}_{\mathrm{c}} + \boldsymbol{G}_{\mathrm{p}} + \boldsymbol{W}_{\mathrm{em}} = \boldsymbol{O}_{3\times 1} \\ &\qquad 0 \le t_{\min} \le t_{\mathrm{c}i} \le t_{\max}, i = 1, 2, \ldots, D_{\mathrm{c}} \end{aligned} \tag{20}$$

where $\boldsymbol{A}_{\mathrm{p}} = \left(\frac{d\boldsymbol{l}}{d\boldsymbol{X}_{\mathrm{m}}}\right)^{\mathrm{T}}$ is the structural matrix of the cables, $t_{\min}$ and $t_{\max}$ represent the lower and upper bounds of the cable tension, respectively.

Equation (20) is a quadratic objective function with the constraints of linear equality and inequality. This problem can be solved using quadratic programming. If a numerical solution for cable tension exists, it indicates that the moving platform satisfies the static

https://doi.org/10.3390/s26041147

equilibrium condition and can provide sufficient support for the SM. If no numerical solution for cable tension is found, the current state cannot maintain a static equilibrium.

4. Coupled Workspace Modeling and Analysis

In general, the FCW of a CDPR is the set of all end-effector pose points that satisfies the static equilibrium condition, reflecting the stability of the robot within its kinematically reachable space. The force closure of a pose point can be determined by the existence of solutions to the static equilibrium equation or by the multidimensional convex hull formed by cable tension [16,19]. The convex hull formed by cable tension indicates the extent to which the moving platform can resist generalized forces in different dimensions. The force closure can be determined by checking the enclosure of each lower-dimensional projection of the convex hull; however, this method is unsuitable for CDHRs. To determine the boundary of the FCW accurately, this study used the Gaussian boundary enrichment method to densify the boundary points of the workspace [33]. Additionally, in CDHRs, the SM may collide with the overall system or the environment, making it necessary to constrain the joint angles of the SM, namely,

$$\boldsymbol{\theta}_{j,\min} \leq \boldsymbol{\theta}_j \leq \boldsymbol{\theta}_{j,\max} \tag{21}$$

where $\boldsymbol{\theta}_{j,\min}$ and $\boldsymbol{\theta}_{j,\max}$ are the lower and upper limits of the angle of the jth joint of the SM, respectively.

4.1. Coupled Workspace Modeling

- Constrained Workspace of the SM

The joint angle space of the SM was fully explored when the pose of the moving platform ${}^{\mathrm{g}}\boldsymbol{X}_{\mathrm{m}}$ remained constant. At this point, the set of end-effector pose points of the CDHR form the constrained SM workspace. This reflects the flexibility of SM spatial movement when the CDHR is in a fixed position. Combined Equation (21) with Equation (13), the FCW can be obtained. As shown in Figure 6, random sampling of the joint angles of the SM is performed while satisfying the constraints of Equation (22), and the corresponding generalized pose $\boldsymbol{Q}_{\mathrm{gp}}$ can be expressed as:

$$\begin{aligned} \boldsymbol{Q}_{\mathrm{gp}}^{(k)} &= f_{\mathrm{random}}\left(\boldsymbol{Q}_{\mathrm{gp,min}}, \boldsymbol{Q}_{\mathrm{gp,max}}\right) \\ \boldsymbol{Q}_{\mathrm{gp}}^{(k)} &= \left[x_{\mathrm{m}}, y_{\mathrm{m}}, z_{\mathrm{m}}, \alpha_{\mathrm{m}}, \beta_{\mathrm{m}}, \gamma_{\mathrm{m}}, \theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_j^{(k)}, \ldots, \theta_{D_{\mathrm{s}}}^{(k)}\right]^{\mathrm{T}} \end{aligned} \tag{22}$$

where $\boldsymbol{Q}_{\mathrm{gp,min}}$, $\boldsymbol{Q}_{\mathrm{gp,max}}$ are the upper and lower bounds of the generalized pose values, respectively, and $f_{\mathrm{random}}(\cdot)$ represents the set of pose points randomly generated within the generalized pose limits by using the Monte Carlo random method.

Thus, the constrained workspace of the SM Ω_{hs} can be described as:

$$\Omega_{\mathrm{hs}} = \left\{ \begin{array}{l} \boldsymbol{X}_{\mathrm{e}}^{(k)} = \left(x_{\mathrm{e}}^{(k)}, y_{\mathrm{e}}^{(k)}, z_{\mathrm{e}}^{(k)}, \alpha_{\mathrm{e}}^{(k)}, \beta_{\mathrm{e}}^{(k)}, \gamma_{\mathrm{e}}^{(k)}\right) \Big| {}^{\mathrm{g}}\boldsymbol{X}_{\mathrm{m}} \\ \boldsymbol{Q}_{\mathrm{gp,min}} \leq \boldsymbol{Q}_{\mathrm{gp}}^{(k)} \leq \boldsymbol{Q}_{\mathrm{gp,max}} \end{array} \right\} \tag{23}$$

https://doi.org/10.3390/s26041147

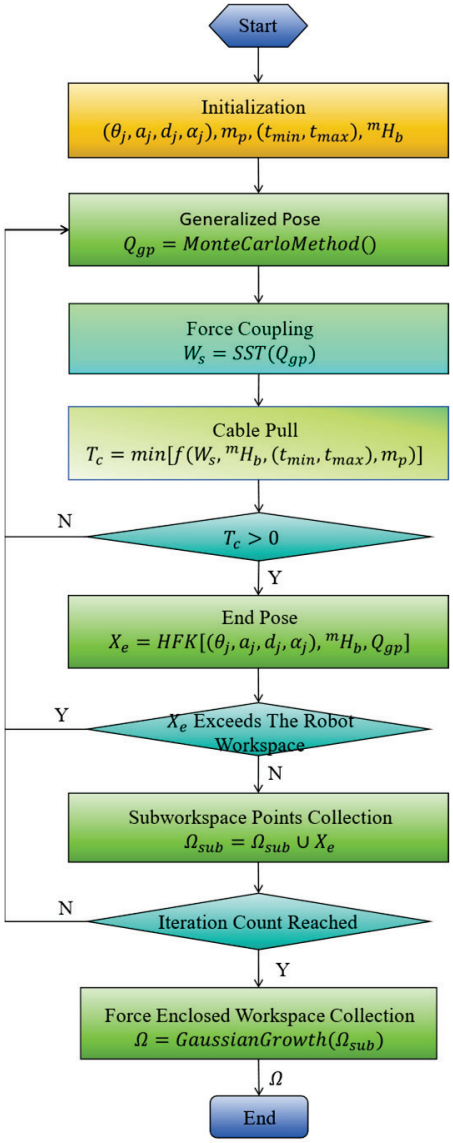


Figure 6. FCW resolving process. SST(·) denotes the recursive Newton–Euler formulation for computing link wrenches of the serial manipulator. HFK[·] denotes the hybrid forward kinematics function of the cable-driven hybrid robot.

- Constrained Workspace of the SM

When the joint angles $\boldsymbol{\Theta}_s$ of the SM remained constant, the workspace of the moving platform was fully explored. The set of end-effector pose points of the CDHR constitutes the constrained workspace of the CDPR, reflecting the flexibility of the moving platform's spatial movement under a fixed load. As shown in Figure 7, by randomly sampling the poses of the moving platform, the corresponding generalized pose $\boldsymbol{Q}_{\mathrm{gp}}$ can be described as:

$$\begin{aligned} \boldsymbol{Q}_{\mathrm{gp}}^{(k)} &= f_{\mathrm{random}}\left(\boldsymbol{Q}_{\mathrm{gp,min}}, \boldsymbol{Q}_{\mathrm{gp,max}}\right) \\ \boldsymbol{Q}_{\mathrm{gp}}^{(k)} &= \left[x_{\mathrm{m}}^{(k)}, y_{\mathrm{m}}^{(k)}, z_{\mathrm{m}}^{(k)}, \alpha_{\mathrm{m}}^{(k)}, \beta_{\mathrm{m}}^{(k)}, \gamma_{\mathrm{m}}^{(k)}, \theta_1, \theta_2, \ldots, \theta_j, \ldots, \theta_{D_{\mathrm{s}}}\right]^{\mathrm{T}} \end{aligned} \tag{24}$$

https://doi.org/10.3390/s26041147

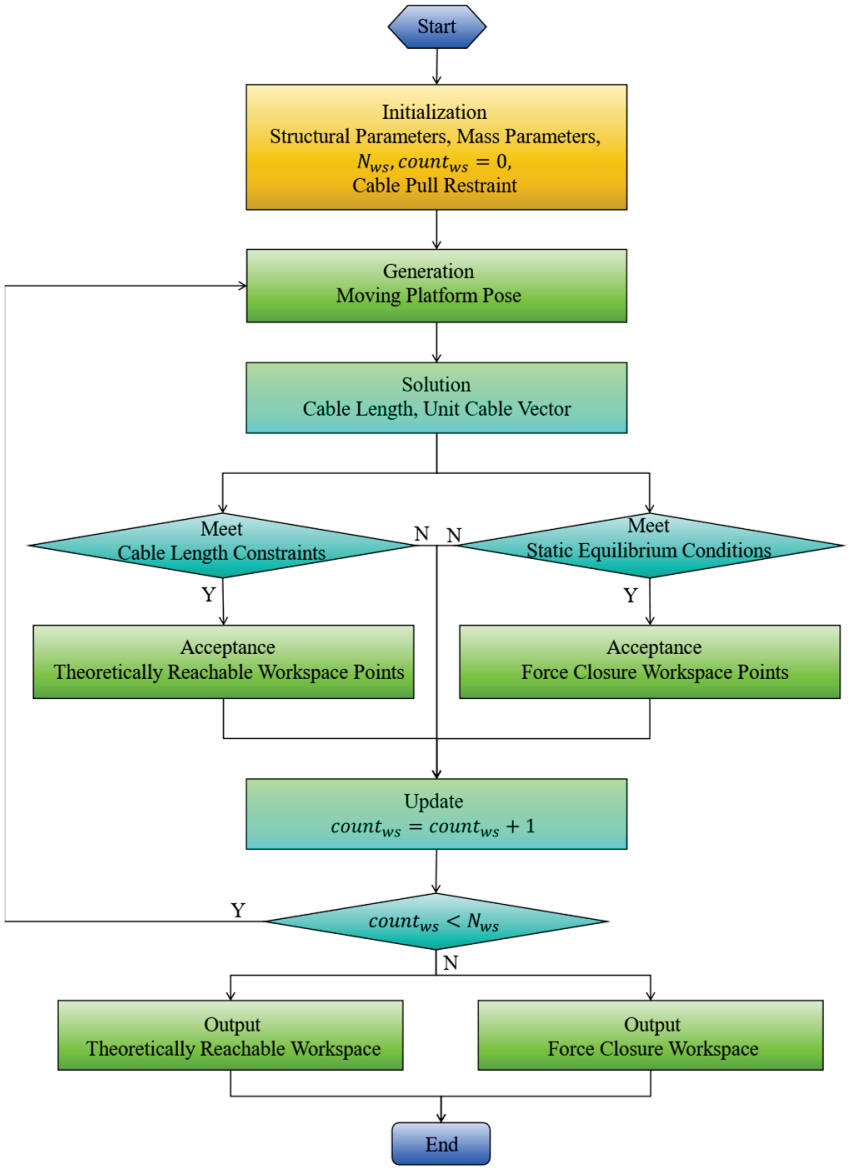


Figure 7. Workspace resolving process of the CDPR.

Similarly, the constrained workspace of the CDPR Ω_{hp} can be written as:

$$\Omega_{\mathrm{hp}} = \left\{ \begin{array}{l} \boldsymbol{X}_{\mathrm{e}}^{(k)} = \left(x_{\mathrm{e}}^{(k)}, y_{\mathrm{e}}^{(k)}, z_{\mathrm{e}}^{(k)}, \alpha_{\mathrm{e}}^{(k)}, \beta_{\mathrm{e}}^{(k)}, \gamma_{\mathrm{e}}^{(k)} \right) \Big| \boldsymbol{\Theta}_{\mathrm{s}}, \\ \boldsymbol{Q}_{\mathrm{gp,min}} \le \boldsymbol{Q}_{\mathrm{gp}}^{(k)} \le \boldsymbol{Q}_{\mathrm{gp,max}}, \boldsymbol{\Theta}_{\mathrm{min}} \le \boldsymbol{\Theta}_{\mathrm{s}} \le \boldsymbol{\Theta}_{\mathrm{max}} \end{array} \right\} \tag{25}$$

where $\boldsymbol{\Theta}_{\mathrm{min}} = [\theta_{1,\mathrm{min}}, \cdots, \theta_{D_{\mathrm{c}},\mathrm{min}}]^{\mathrm{T}}$, $\boldsymbol{\Theta}_{\mathrm{max}} = [\theta_{1,\mathrm{max}}, \cdots, \theta_{D_{\mathrm{c}},\mathrm{max}}]^{\mathrm{T}}$.

- Complete FCW

When all generalized joints of the CDHR change, the set of end-effector pose points constitutes the complete FCW. This reflects the flexibility of CDHR when performing tasks in unstructured and constrained environments. As shown in Figure 6, the generalized pose $\boldsymbol{Q}_{\mathrm{gp}}$ is randomly sampled along all dimensions, that is:

$$\begin{array}{c} \boldsymbol{Q}_{\mathrm{gp}}^{(k)} = f_{\mathrm{random}} \left(\boldsymbol{Q}_{\mathrm{gp,min}}, \boldsymbol{Q}_{\mathrm{gp,max}} \right) \\ \boldsymbol{Q}_{\mathrm{gp}}^{(k)} = \left[x_{\mathrm{m}}^{(k)}, y_{\mathrm{m}}^{(k)}, z_{\mathrm{m}}^{(k)}, \alpha_{\mathrm{m}}^{(k)}, \beta_{\mathrm{m}}^{(k)}, \gamma_{\mathrm{m}}^{(k)}, \theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_j^{(k)}, \ldots, \theta_{D_{\mathrm{s}}}^{(k)} \right]^{\mathrm{T}} \end{array} \tag{26}$$

https://doi.org/10.3390/s26041147

The complete FCW set Ω_{h} is specifically expressed as:

$$\Omega_{\mathrm{h}} = \left\{ \begin{array}{c} \boldsymbol{X}_{\mathrm{e}}^{(k)} = \left(x_{\mathrm{e}}^{(k)}, y_{\mathrm{e}}^{(k)}, z_{\mathrm{e}}^{(k)}, \alpha_{\mathrm{e}}^{(k)}, \beta_{\mathrm{e}}^{(k)}, \gamma_{\mathrm{e}}^{(k)} \right) \Big| \boldsymbol{\Theta}_{\min} \leq \boldsymbol{\Theta}_{\mathrm{s}} \leq \boldsymbol{\Theta}_{\max}, \\ \boldsymbol{Q}_{\mathrm{gp,min}} \leq \boldsymbol{Q}_{\mathrm{gp}}^{(k)} \leq \boldsymbol{Q}_{\mathrm{gp,max}} \end{array} \right\} \tag{27}$$

4.2. Workspace Analysis

- Coupled Workspace Analysis

To describe the workspace boundaries more accurately, the Gaussian enrichment method is used to improve boundary precision. The corresponding simulation parameters are shown in Table 3. (n_x, n_y, n_z) is the grid resolution. σ^{ini} is the initial standard deviation; ω_k is the variance-decay factor; $n_f^{\max}$ is the maximum number of grid-fill attempts; N_c is the maximum number of grid points. The range of joint angles for the SM is provided in Table 4.

Table 3. Range of joint angles for the SM.

Name	θ (°)	d (m)	a (m)	α (°)
Joint 1	0	0.0500	1.5708	0
Joint 2	90	0	0.0829	0
Joint 3	0	0	0	0
Joint 4	0	0	0.1658	0

Table 4. Uniformly discretized sampled pose points.

Name	x-Axis	y-Axis	z-Axis
Minimum	0.06	−0.36	−0.62
Maximum	0.62	0.36	0
Spacing	0.02	0.02	0.02

In Figures 8–11, the blue point cloud denotes the sampled feasible position set (units in meters) that satisfies the corresponding constraints; points outside the feasible set are not plotted. Each figure provides a 3D distribution and its 2D projections on the (XY), (XZ), and (YZ) planes to visualize the boundary. Comparing Figure 8 with Figure 9, the feasible set becomes larger when the serial manipulator is included, especially along the arm extension direction, and feasibility is recovered in regions where the CDPR-only case violates the force-closure condition. The CDPR does not satisfy the force-closure condition at some points near the center, whereas the CDHR remains reachable in these regions.

When the moving platform pose is set as $\boldsymbol{X}_{\mathrm{m}} = [0.32, 0, -0.32, 0, 0, 0]^{\mathrm{T}}$, the generated constrained workspace of the SM is shown in Figure 10. The pose of the moving platform was near the center of the space formed by the aluminum frame of the robot. In each dimension, the geometric shape of the robot end-effector workspace boundary formed a slightly imperfect ellipse. The missing portions were caused by the joint angle limitations of the SM, indicating that the workspace loss owing to structural coupling was minimal.

https://doi.org/10.3390/s26041147

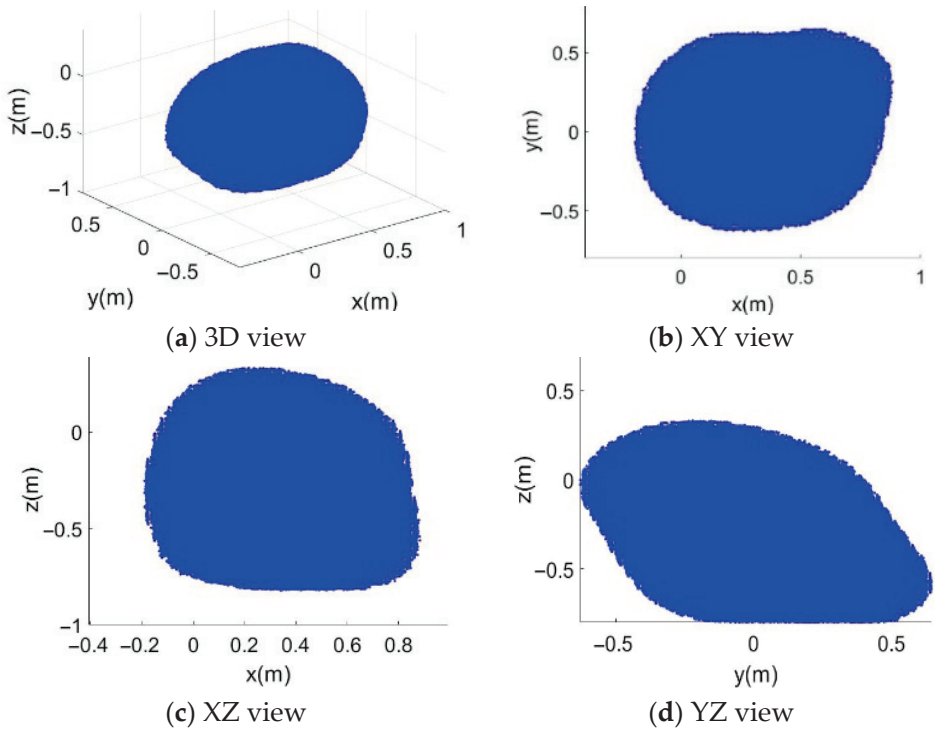


Figure 8. Complete FCW of the CDHR. (**a**) 3D view in (x,y,z); (**b**) projection on the (x,y)-plane; (**c**) projection on the (x,z)-plane; (**d**) projection on the (y,z)-plane. Blue points represent sampled feasible positions satisfying the force-closure condition under the considered constraints.

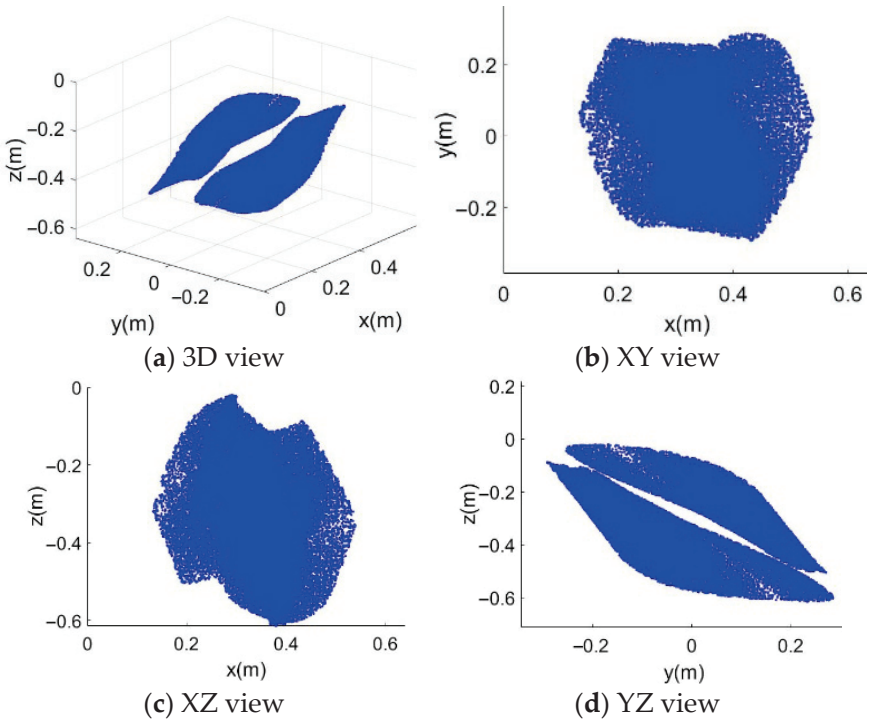


Figure 9. FCW of the CDPR without the SM. (**a**) 3D view in (x,y,z); (**b**) projection on the (x,y)-plane; (**c**) projection on the (x,z)-plane; (**d**) projection on the (y,z)-plane. Blue points represent sampled feasible positions under the same force-closure evaluation.

https://doi.org/10.3390/s26041147

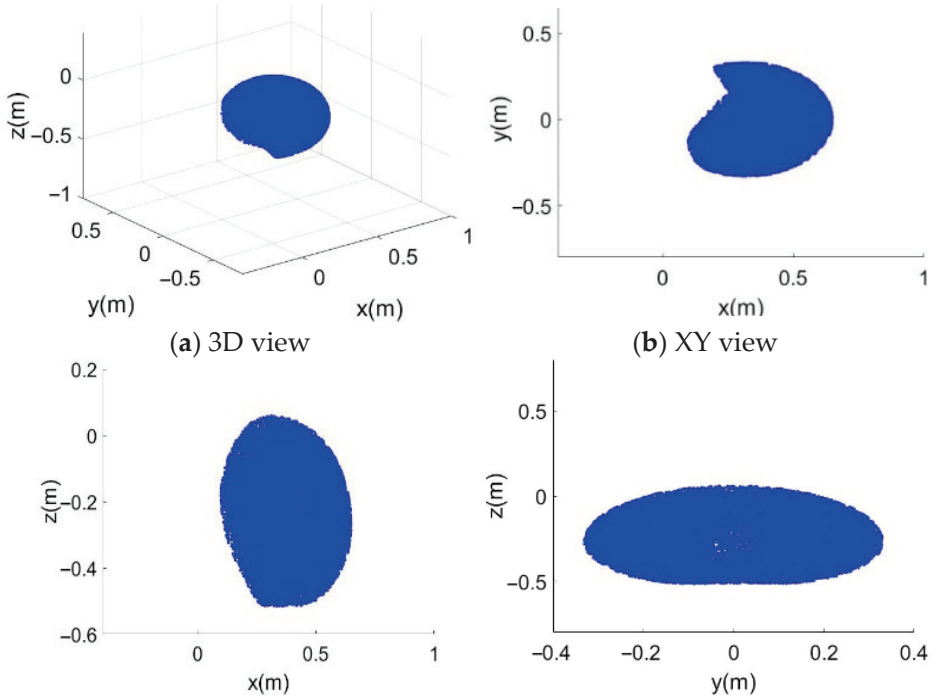


Figure 10. Constrained workspace of the SM with the fixed moving platform. (**a**) 3D view in (x,y,z); (**b**) projection on the (x,y)-plane; (**c**) projection on the (x,z)-plane; (**d**) projection on the (y,z)-plane. Blue points represent reachable end-effector positions under the SM joint-limit constraints.

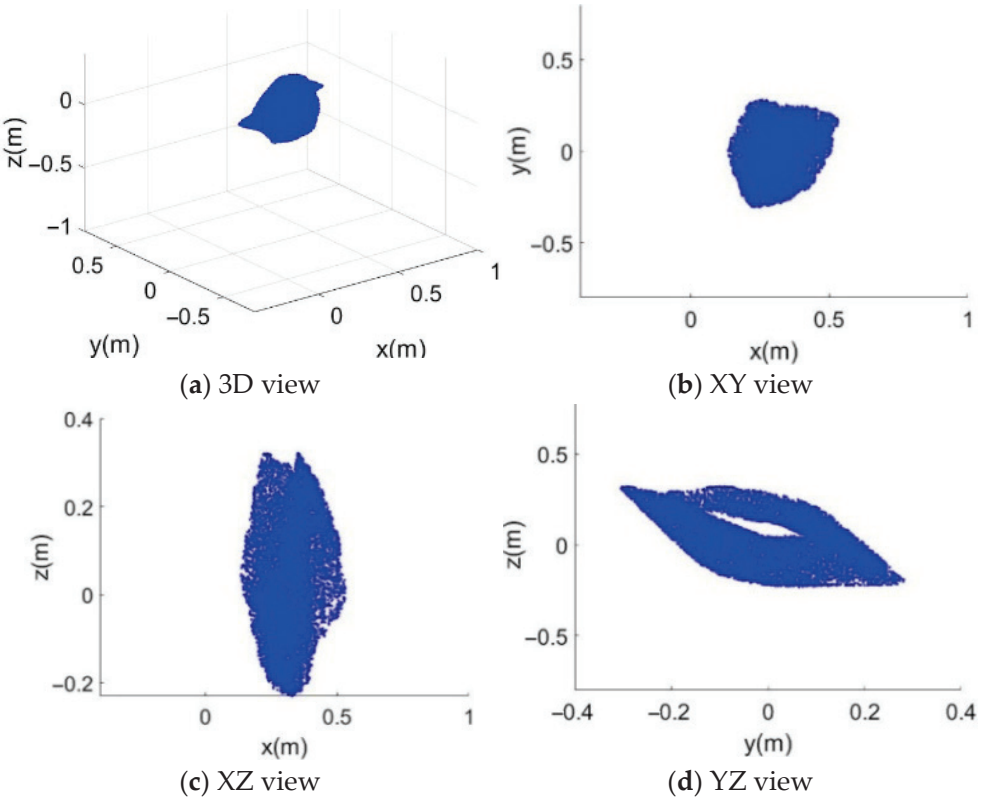


Figure 11. The constrained workspace of the CDPR with the SM pose is fixed. (**a**) 3D view in (x,y,z); (**b**) projection on the (x,y)-plane; (**c**) projection on the (x,z)-plane; (**d**) projection on the (y,z)-plane. Blue points represent sampled feasible platform positions satisfying the force-closure condition with the SM configuration fixed.

https://doi.org/10.3390/s26041147

With the initial joint angles of SM set to $\boldsymbol{\Theta}_{\mathrm{s}} = [0, \pi/2, 0, 0]^{\mathrm{T}}$, the constrained workspace of the CDPR is shown in Figure 11. From Figure 11c,d, it can be observed that the reachable positions along the z-axis are located outside the convex hull formed by the fixed anchor points. This is because the manipulator is mounted on the upper surface of the moving platform, causing the constrained workspace of the CDPR to shift positively along the z-axis. Figure 11b,c show that along the x- and y-axes, the reachable workspace is significantly reduced compared with the complete FCW.

- Workspace Analysis of Arm Flexibility

According to Equation (18), when the joint angles of the SM change, the force exerted by the SM on the moving platform remains constant but the applied moment changes. Therefore, when the pose of the moving platform remains unchanged and the configuration of the SM changes, the static equilibrium equation of the moving platform can be expressed by Equation (20) may not yield a numerical solution. Thus, it is necessary to determine the arm flexibility of the CDHR and evaluate the quality of the robot workspace. Using the Monte Carlo method, the joint angles $\boldsymbol{\Theta}_{\mathrm{s}}$ of the SM can be randomly generated According to Equation (9), the relative pose ${}^{\mathrm{b}}\boldsymbol{X}_{\mathrm{e}}$ from the SM base to the end effector can be calculated. By substituting ${}^{\mathrm{b}}\boldsymbol{X}_{\mathrm{e}}$ and $\boldsymbol{\Theta}_{\mathrm{s}}$ into Equation (28) and repeating the process times yields the corresponding subspace as follows:

$$\begin{aligned} \Omega_{\mathrm{m}} &: \left\{ \begin{array}{l} \boldsymbol{X}_{\mathrm{e}}^{(\kappa)} \left| {}^{\mathrm{b}}\boldsymbol{X}_{\mathrm{e}}^{(\kappa)} = \mathrm{CFK}\left(\boldsymbol{p}_{bi}, \boldsymbol{r}_{ai}, {}^{\mathrm{m}}\boldsymbol{T}_{\mathrm{b}}, \boldsymbol{Q}_{\mathrm{gj}}{}^{(\kappa)}\right), \right. \\ \kappa \in \{1, 2, \ldots, N_a\} \end{array} \right\} \\ \Omega_{\mathrm{s}} &: \left\{ \begin{array}{l} \boldsymbol{Q}_{\mathrm{gj}}{}^{(\kappa)} = \left[\theta_1^{(\kappa)}, \theta_2^{(\kappa)}, \ldots, \theta_{D_{\mathrm{s}}}^{(\kappa)}\right], \\ u \in \{1, 2, \ldots, N_a\} \end{array} \right\} \end{aligned} \tag{28}$$

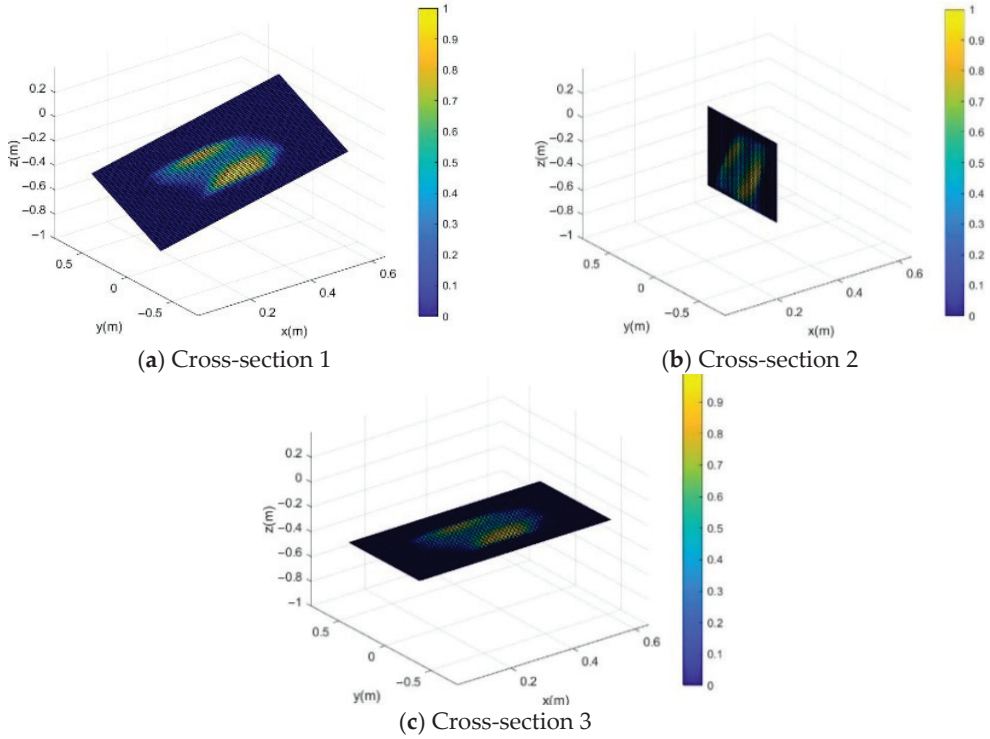
According to Equation (28), three different sectional equations are selected to compute the flexibility space:

$$\begin{aligned} \Phi_1 &: \left\{ \begin{array}{l} x = x_{\min} + i_x \cdot \frac{(x_{\max} - x_{\min})}{50} \\ y = y_{\min} + i_y \cdot \frac{(y_{\max} - y_{\min})}{50} \\ z = z_{\min} + \frac{0.5 i_x}{(z_{\max} - z_{\min})} + \frac{0.5 i_y}{(z_{\max} - z_{\min})} \end{array} \right. , \Phi_2 : \left\{ \begin{array}{l} x = 0.32 \\ y = y_{\min} + i_y \cdot \frac{(y_{\max} - y_{\min})}{50} \\ z = z_{\min} + i_z \cdot \frac{(z_{\max} - z_{\min})}{50} \end{array} \right. \\ \Phi_3 &: \left\{ \begin{array}{l} x = x_{\min} + i_x \cdot \frac{(x_{\max} - x_{\min})}{50} \\ y = y_{\min} + i_y \cdot \frac{(y_{\max} - y_{\min})}{50} \\ z = -0.32 \end{array} \right. \end{aligned} \tag{29}$$

where $z_{\max} = 0.024$ m, $z_{\min} = -0.626$ m, $y_{\max} = 0.385$ m, $y_{\min} = -0.385$ m, $x_{\max} = 0.645$ m, $x_{\min} = 0.025$ m, the values for i_x, i_y and i_z range from $0, 1, 2, 3, \cdots, 50$.

As shown in Figure 12, the CDHR satisfies the force-closure condition near the central region of the bounding box. However, at the edges of the workspace, the flexibility of the CDHR decreases significantly.

https://doi.org/10.3390/s26041147



(**a**) Cross-section 1 (**b**) Cross-section 2

(**c**) Cross-section 3

Figure 12. Heat maps of the CDHR's flexibility at three different cross-sections. (**a**) represents cross-section Φ_1, (**b**) represents cross-section Φ_2, and (**c**) represents cross-section Φ_3. The color spectrum represents the distribution of flexibility values, where yellow indicates greater flexibility, and blue indicates lower flexibility.

5. Dynamic Anchor Seat Position Optimization Method Based on the Global Condition Number

The direction of cable tension on the moving platform of the CDHR is determined by the positions of the anchor seats on the housing and moving platform. To ensure movement flexibility, the positions of the anchor seats on the moving platform need to be optimized. Owing to the unidirectional nature of cable forces, the cable tension applied to the moving platform can form a convex hull surrounding the platform's center of mass, providing force/torque in any direction force [17]. The global condition number is an important indicator for evaluating the flexibility and performance of the robot [34]. Based on this, this study optimizes the shape and installation pose of the dynamic anchor seats on the prototype using the global condition number.

The static equilibrium equation of the moving platform in the CDHR can be expressed as a system of linear equations, i.e., $\boldsymbol{Ax} = \boldsymbol{b}$, $\boldsymbol{A} = \boldsymbol{A}_{\mathrm{p}}$, $\boldsymbol{x} = \boldsymbol{T}_{\mathrm{c}}$, $\boldsymbol{b} = -\boldsymbol{G}_{\mathrm{p}} - \boldsymbol{W}_{\mathrm{em}}$. The solution to this system of linear equations represents the magnitude of cable tension. When the external forces change, the generalized force $\boldsymbol{W}_{\mathrm{em}}$ applied by the SM on the moving platform will also change. The condition number is used to measure the sensitivity of x to changes in b in the linear system $\boldsymbol{Ax} = \boldsymbol{b}$. The smaller the condition number, the less impact perturbations in b have on x. Therefore, the condition number can be described as:

$$\kappa(\boldsymbol{A}) = \mathrm{cond}(\boldsymbol{A}) = \|\boldsymbol{A}\|_2 \cdot \left\|\boldsymbol{A}^{-1}\right\|_2 = \frac{\sigma_{\max}(\boldsymbol{A})}{\sigma_{\min}(\boldsymbol{A})} \tag{30}$$

where $\kappa(\boldsymbol{A}) \in [1, \infty]$ is the condition number of matrix A, $\|\boldsymbol{A}\|_2$ is the 2-norm of matrix A, $\sigma_{\max}(\boldsymbol{A})$ is the largest singular value of matrix A, and $\sigma_{\min}(\boldsymbol{A})$ is the smallest singular value of matrix A.

https://doi.org/10.3390/s26041147

The global condition numbers defined in Equation (30) can be used to describe the degree to which the motion performance of the hybrid robot approaches isotropy at different poses [22], i.e.,

$$\mathrm{GCI} = \frac{\int_U \kappa(\boldsymbol{A})^{-1} du}{\int_U du} \tag{31}$$

where n_κ is the size of the discrete set of workspace points for the CDHR.

The influence of the cable forces on the moving platform in each motion dimension can be represented by the structure matrix $\boldsymbol{A}_\mathrm{p}$ of the CDHR. As $\boldsymbol{A}_\mathrm{p}$ contains both translational and rotational components, the structure matrix $\boldsymbol{A}_\mathrm{p}$ can be normalized to maintain consistent motion performance along the principal axes of each dimension [24]. The normalized structure matrix can be expressed as:

$$\hat{\boldsymbol{A}}_\mathrm{p} = \begin{bmatrix} \hat{\boldsymbol{L}}_1 & \hat{\boldsymbol{L}}_2 & \cdots & \hat{\boldsymbol{L}}_{D_\mathrm{c}} \\ \frac{\boldsymbol{r}_{a1} \times \hat{\boldsymbol{L}}_1}{L_\mathrm{ch}} & \frac{\boldsymbol{r}_{a2} \times \hat{\boldsymbol{L}}_2}{L_\mathrm{ch}} & \cdots & \frac{\boldsymbol{r}_{aD_\mathrm{c}} \times \hat{\boldsymbol{L}}_{D_\mathrm{c}}}{L_\mathrm{ch}} \end{bmatrix} \tag{32}$$

where L_ch is the characteristic length of the matrix $\boldsymbol{A}_\mathrm{p}$.

Based on Equation (33), the corresponding motion flexibility in this state can be obtained as:

$$\begin{aligned} \boldsymbol{M}_q &= \mathrm{diag}\left(\hat{\boldsymbol{A}}_\mathrm{p} \cdot \hat{\boldsymbol{A}}_\mathrm{p}^\mathrm{T}\right) \\ &= \mathrm{diag}\left(\begin{bmatrix} \sum_{i=1}^{D_\mathrm{c}} \hat{\boldsymbol{L}}_i \cdot \hat{\boldsymbol{L}}_i^\mathrm{T} & \sum_{i=1}^{D_\mathrm{c}} \frac{\hat{\boldsymbol{L}}_i \cdot \left(\boldsymbol{r}_{ai} \times \hat{\boldsymbol{L}}_i\right)^\mathrm{T}}{L_\mathrm{ch}} \\ \sum_{i=1}^{D_\mathrm{c}} \frac{\left(\boldsymbol{r}_{ai} \times \hat{\boldsymbol{L}}_i\right) \cdot \hat{\boldsymbol{L}}_i^\mathrm{T}}{L_\mathrm{ch}} & \sum_{i=1}^{D_\mathrm{c}} \frac{\left(\boldsymbol{r}_{ai} \times \hat{\boldsymbol{L}}_i\right) \cdot \left(\boldsymbol{r}_{ai} \times \hat{\boldsymbol{L}}_i\right)^\mathrm{T}}{L_\mathrm{ch}} \end{bmatrix}\right) \end{aligned} \tag{33}$$

Assuming that the CDHR satisfies the condition of isotropic motion, i.e., $\boldsymbol{M}_q = \sigma^2 \boldsymbol{E}_6$, where $\boldsymbol{E}_6$ is the 6-dimensional identity matrix, according to Equation (34), we obtain:

$$\begin{cases} \sum_{i=1}^{D_\mathrm{c}} \hat{L}_{i,x}^2 = \sum_{i=1}^{D_\mathrm{c}} \hat{L}_{i,y}^2 = \sum_{i=1}^{D_\mathrm{c}} \hat{L}_{i,z}^2 = \sigma^2 \\ \sum_{i=1}^{D_\mathrm{c}} \frac{\left(r_{ai,y}^2 \cdot \hat{L}_{i,z} - r_{ai,z}^2 \cdot \hat{L}_{i,y}\right)^2}{L_\mathrm{ch}^2} = \sigma^2 \\ \sum_{i=1}^{D_\mathrm{c}} \frac{\left(r_{ai,z}^2 \cdot \hat{L}_{i,x} - r_{ai,x}^2 \cdot \hat{L}_{i,z}\right)^2}{L_\mathrm{ch}^2} = \sigma^2 \\ \sum_{i=1}^{D_\mathrm{c}} \frac{\left(r_{ai,x}^2 \cdot \hat{L}_{i,y} - r_{ai,y}^2 \cdot \hat{L}_{i,z}\right)^2}{L_\mathrm{ch}^2} = \sigma^2 \end{cases} \tag{34}$$

where $\left(\hat{L}_{i,x}, \hat{L}_{i,y}, \hat{L}_{i,z}\right)$ are the components of $\hat{\boldsymbol{L}}_i$ along each coordinate axis, and $\left(r_{ai,x}, r_{ai,y}, r_{ai,z}\right)$ are the components of $\boldsymbol{r}_{ai}$ along each coordinate axis. By combining the subexpressions in Equation (35), the characteristic length of matrix $\hat{\boldsymbol{A}}_\mathrm{p}$ can be obtained as:

$$L_\mathrm{ch} = \sqrt{\frac{(L_{chx})^2 + \left(L_{chy}\right)^2 + (L_{chz})^2}{D_\mathrm{c}}} = \frac{\left\|\boldsymbol{r}_{ai} \times \hat{\boldsymbol{L}}_i\right\|}{D_\mathrm{c}} \tag{35}$$

where $L_{chx} = r_{ai,y}^2 \cdot \hat{L}_{i,z} - r_{ai,z}^2 \cdot \hat{L}_{i,y}$, $L_{chy} = r_{ai,z}^2 \cdot \hat{L}_{i,x} - r_{ai,x}^2 \cdot \hat{L}_{i,z}$, $L_{chz} = r_{ai,x}^2 \cdot \hat{L}_{i,y} - r_{ai,y}^2 \cdot \hat{L}_{i,z}$.

https://doi.org/10.3390/s26041147

By substituting Equation (36) into Equation (33), the structural matrix of the CDHR can be obtained. Combining this with Equations (30) and (32), the condition number and global condition number of $\hat{\boldsymbol{A}}_{\mathrm{p}}$ are given as:

$$\kappa\left(\hat{\boldsymbol{A}}_{\mathrm{p}}\right)=\frac{\sigma_{\max}\left(\hat{\boldsymbol{A}}_{\mathrm{p}}\right)}{\sigma_{\min}\left(\hat{\boldsymbol{A}}_{\mathrm{p}}\right)} \tag{36}$$

where N_{κ} represents the number of discrete pose points in the set Ω_{m}.

To ensure that the global condition number in the discrete case closely approximated the actual global condition number, the reachable workspace Ω_{m} of the moving platform was uniformly discretized. Thus, the position optimization model for the moving platform anchor seats can be expressed as:

$$\boldsymbol{X}_{ra}=\arg\min\left(1-\sum_{k=1}^{n_{\kappa}}\kappa\left(\hat{\boldsymbol{A}}_{\mathrm{p}}^{(k)}\right)^{-1}/n_{\kappa}\right) \tag{37}$$

Considering that a spherical distribution of anchor seats balances forces in all directions and reduces structural deformation of the moving platform caused by uneven cable tensions, we adopt a moving platform with anchors arranged on a sphere in this study. Let d_a be the radius of the sphere with d_a = 75 mm. Let θ_{ai} be the angle between the line segment connecting the origin of the moving coordinate system to the anchor seat and the z-axis of the moving coordinate system and be the angle between the projection of this line segment onto the XY plane and the x-axis of the moving coordinate system. Based on the correspondence between the spherical and Cartesian coordinates, the spherical coordinates of the anchor seat can be expressed as

$$\begin{cases} r_{ai,x}=d_a\cdot\sin\theta_{ai}\cdot\cos\varphi_{ai} \\ r_{ai,y}=d_a\cdot\sin\theta_{ai}\cdot\sin\varphi_{ai} \\ r_{ai,z}=d_a\cdot\cos\theta_{ai} \end{cases} \tag{38}$$

Furthermore, the optimization variables can be designed to $\boldsymbol{X}_{ra}=[d_a,\theta_{a1},\varphi_{a1},\ldots,\theta_{aD_{\mathrm{c}}},\varphi_{aD_{\mathrm{c}}}]^{\mathrm{T}}$, reduce the dimensionality from $1+2D_{\mathrm{c}}$ by $D_{\mathrm{c}}-1$. The optimization objective corresponding to the global condition number is:

$$\begin{aligned} &\min f(\boldsymbol{X}_{ra}) \\ &\text{s.t. } \boldsymbol{X}_{\mathrm{e}}\in\Omega_{\mathrm{h}} \end{aligned} \tag{39}$$

In this study, the Particle Swarm Optimization (PSO) algorithm is used to solve Equation (41). Because the line segments connecting the fixed anchor points and the moving anchor points may collide with the moving platform, it is necessary to penalize the fitness of particles to remove these unreasonable moving anchor points. As shown in Figure 13, a collision analysis is conducted on the plane μ formed by the sphere center, a single moving anchor seat, and a single fixed anchor seat. Let θ_t represent the reserved tolerance rotation angle, and A_i'' represent the tolerance deviation angle mapped to the tangent. A_i'' represents the position of the moving platform anchor seat during optimization, and represents the position of the moving anchor seat corresponding to the maximum rotation angle of the moving platform around the normal vector of the plane μ. $A_i^{\max}$ represents the position boundary of the moving platform anchor seat during optimization. Based on geometric relationships, $\alpha_t=\theta_t/2$. The red arc represents the safe region of anchor seat distribution in the tangent plane. The safe region in the three-dimensional space for the moving platform anchor seat is the surface generated by rotating the red arc around the plane μ, which connects the fixed anchor seat $O_{\mathrm{m}}B_i$ by 360°.

https://doi.org/10.3390/s26041147

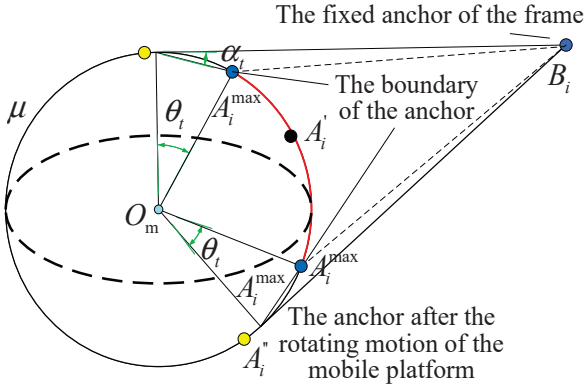


Figure 13. Intersection collision analysis of the sphere center–moving anchor–fixed anchor plane.

According to the above analysis, the evaluation score for intersection collision s_1 can be expressed as:

$$s_1 = \frac{A_i' B_i - A_i^{\max} B_i}{O_m B_i} \tag{40}$$

where $A_i'' A_i^{\max} = \sqrt{2 \cdot d_a^2 (1 - \cos(\theta_t))}$, $A_i^{\max} B_i = \sqrt{\left(A_i'' B_i\right)^2 + \left(A_i'' A_i^{\max}\right)^2 - 2 A_i'' B_i \cdot A_i'' A_i^{\max} \cdot \cos(\alpha_t)}$.

Thus, the particle fitness v_f in the PSO algorithm can be expressed as:

$$v_f = \begin{cases} 1 + s_1, \text{ If there is a cross collision} \\ s_2, \text{ else} \end{cases} \tag{41}$$

where the evaluation score of the particle $s_2 = 1 - \sum_{k=1}^{n_\kappa} \kappa\left(\hat{\boldsymbol{A}}_p\right)^{-1} / n_\kappa$.

6. Vision-Based Self-Calibration Method for Kinematic Parameters

6.1. Multi-Mapping Calibration Problem of the Hybrid Robot

As shown in Figure 14, when deploying the reconfigurable CDHR, it is necessary to calibrate the positions of the fixed anchor seats, installation position of the SM, and the hand-eye relationship. Typically, the coordinates of moving anchor seats relative to the moving coordinate system $\boldsymbol{r}_{ai}$ are known. The global frame {G} is established on the April QR code [35] and aligned with the April frame. {V} represents the camera frame mounted on the robot, {E} is the coordinate system of the SM end-effector, and {T} is the frame of the collaboration marker. After installing the CDHR, the motor encoder count is reset to zero at the initial moment. During the actual operation, the change in the cable length was measured using angle sensors and converted (as in Equation (4)), yielding the current cable length $\mathbf{l}$. Next, the April QR code recognition algorithm calculates the HTM (i.e., ${}^{\mathrm{v}}\boldsymbol{T}_{\mathrm{g}.}$) from the QR code frame to the camera frame. The joint angles $\boldsymbol{\Theta}_s$ of the SM are obtained via feedback from the joint servomotors, allowing the calculation of $\boldsymbol{Q}_{gj}$. Thus, given N_{calib} sets of generalized joint measurements $\boldsymbol{Q}_{gj}$, it is possible to simultaneously calibrate the position of the fixed anchor seats $\boldsymbol{p}_{bi}$, the relative pose between the moving platform and base, and the hand-eye external parameter matrix of the end-effector.

https://doi.org/10.3390/s26041147

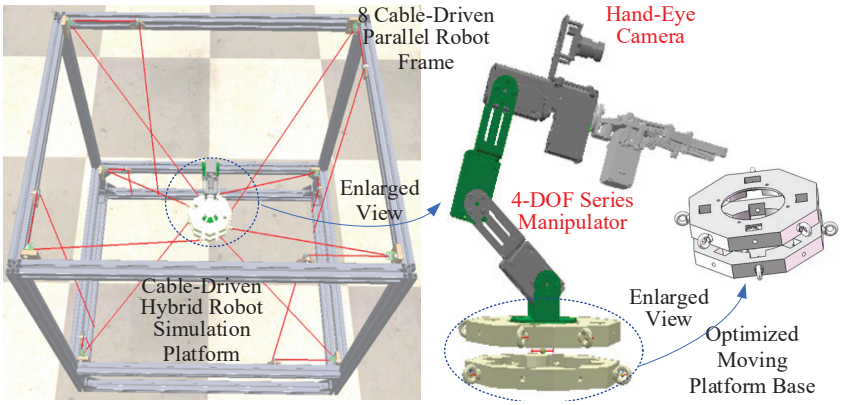


Figure 14. Simulation model of the CDHR.

6.2. Resolving of the Self-Calibration Equation for the CDHR

In the global frame, the position vector of the moving anchor seat can be expressed as:

$$\begin{aligned} {}^{\mathrm{g}}\boldsymbol{r}_{ai} &= {}^{\mathrm{g}}\boldsymbol{t}_{\mathrm{v}} + {}^{\mathrm{g}}\boldsymbol{R}_{\mathrm{v}} \cdot {}^{\mathrm{v}}\boldsymbol{t}_{\mathrm{e}} + {}^{\mathrm{g}}\boldsymbol{R}_{\mathrm{e}} \cdot {}^{\mathrm{e}}\boldsymbol{t}_{\mathrm{b}} + {}^{\mathrm{g}}\boldsymbol{R}_{\mathrm{b}} \cdot {}^{\mathrm{b}}\boldsymbol{t}_{\mathrm{m}} + {}^{\mathrm{g}}\boldsymbol{R}_{\mathrm{m}} \cdot {}^{\mathrm{m}}\boldsymbol{r}_{i} \\ &= {}^{\mathrm{g}}\boldsymbol{t}_{\mathrm{v}} + {}^{\mathrm{g}}\boldsymbol{R}_{\mathrm{v}} \cdot {}^{\mathrm{v}}\boldsymbol{t}_{\mathrm{e}} + {}^{\mathrm{g}}\boldsymbol{R}_{\mathrm{v}} \cdot {}^{\mathrm{v}}\boldsymbol{R}_{\mathrm{e}} \cdot {}^{\mathrm{e}}\boldsymbol{t}_{\mathrm{b}} + {}^{\mathrm{g}}\boldsymbol{R}_{\mathrm{v}} \cdot {}^{\mathrm{v}}\boldsymbol{R}_{\mathrm{e}} \cdot {}^{\mathrm{e}}\boldsymbol{R}_{\mathrm{b}} \cdot {}^{\mathrm{b}}\boldsymbol{t}_{\mathrm{p}} + {}^{\mathrm{g}}\boldsymbol{R}_{\mathrm{v}} \cdot {}^{\mathrm{v}}\boldsymbol{R}_{\mathrm{e}} \cdot {}^{\mathrm{e}}\boldsymbol{R}_{\mathrm{b}} \cdot {}^{\mathrm{b}}\boldsymbol{R}_{\mathrm{m}} \cdot {}^{\mathrm{m}}\boldsymbol{r}_{ai} \end{aligned} \tag{42}$$

where ${}^{\mathrm{c}}\boldsymbol{T}_{\mathrm{d}}$ represents the HTM from frame {c} to frame {d}, ${}^{\mathrm{c}}\boldsymbol{R}_{\mathrm{d}}$ and ${}^{\mathrm{c}}\boldsymbol{t}_{\mathrm{d}}$ represent the rotation matrix and translation vector of ${}^{\mathrm{c}}\boldsymbol{T}_{\mathrm{d}}$, respectively. These satisfy the conditions ${}^{\mathrm{c}}\boldsymbol{R}_{\mathrm{d}} = {}^{\mathrm{c}}\boldsymbol{T}_{\mathrm{d}}(1:3,1:3) = \left[{}^{\mathrm{d}}\boldsymbol{T}_{\mathrm{c}}(1:3,1:3)\right]^{-1}$, ${}^{\mathrm{c}}\boldsymbol{t}_{\mathrm{d}} = {}^{\mathrm{c}}\boldsymbol{T}_{\mathrm{d}}(1:3,4) = \left[{}^{\mathrm{d}}\boldsymbol{T}_{\mathrm{c}}(1:3,4)\right]^{-1}$.

Furthermore, the actual length of the ith cable l_i can be expressed as:

$$l_i = f_{\mathrm{calib}}\left({}^{\mathrm{g}}\boldsymbol{p}_{bi}, {}^{\mathrm{m}}\boldsymbol{r}_{ai}, {}^{\mathrm{b}}\boldsymbol{X}_{\mathrm{m}}, {}^{\mathrm{v}}\boldsymbol{X}_{\mathrm{e}}, {}^{\mathrm{v}}\boldsymbol{T}_{\mathrm{g}}, {}^{\mathrm{b}}\boldsymbol{T}_{\mathrm{e}}\right) = \left\| {}^{\mathrm{g}}\boldsymbol{p}_{bi} - {}^{\mathrm{g}}\boldsymbol{r}_{ai} \right\|_2 \tag{43}$$

where ${}^{\mathrm{c}}\boldsymbol{X}_{\mathrm{d}}$ represents the relative pose of frame {d} with respect to frame {c}. ${}^{b}\boldsymbol{X}_{m}$ represents the relative pose of moving platform {m} with respect to base {b}. ${}^{\mathrm{b}}\widetilde{\boldsymbol{X}}_{\mathrm{m}}^{(0)}$ is the initial value of ${}^{b}\boldsymbol{X}_{m}$.

The solution to Equation (45) can be obtained using the least squares method. Let the iterative values of ${}^{\mathrm{b}}\boldsymbol{X}_{\mathrm{m}}$, ${}^{\mathrm{v}}\boldsymbol{X}_{\mathrm{e}}$, and ${}^{\mathrm{g}}\boldsymbol{p}_{bi}$ be ${}^{\mathrm{b}}\widetilde{\boldsymbol{X}}_{\mathrm{m}}$, ${}^{\mathrm{v}}\widetilde{\boldsymbol{X}}_{\mathrm{e}}$, and ${}^{\mathrm{g}}\widetilde{\boldsymbol{p}}_{bi}$, respectively. Then, the error model for the i-th cable is:

$$\begin{aligned} &\left({}^{\mathrm{b}}\boldsymbol{X}_{\mathrm{m}}, {}^{\mathrm{v}}\boldsymbol{X}_{\mathrm{e}}, {}^{\mathrm{g}}\boldsymbol{p}_{bi}\right) = \operatorname{argmin} \sum_{i}^{D_{\mathrm{c}}} \left\| \hat{l}_i - \widetilde{l}_i \right\|_2 \\ &\text{s.t. } \widetilde{l}_i = f_{\mathrm{calib}}\left({}^{\mathrm{g}}\widetilde{\boldsymbol{p}}_{bi}, {}^{\mathrm{m}}\boldsymbol{r}_{ai}, {}^{\mathrm{b}}\widetilde{\boldsymbol{X}}_{\mathrm{m}}, {}^{\mathrm{v}}\widetilde{\boldsymbol{X}}_{\mathrm{e}}, {}^{\mathrm{v}}\boldsymbol{T}_{\mathrm{g}}, {}^{\mathrm{b}}\boldsymbol{T}_{\mathrm{e}}\right) \\ &\quad i = 1, 2, \ldots, D_{\mathrm{c}} \end{aligned} \tag{44}$$

where $\hat{l}_i$ is the actual value measured by the drum rotation angle sensor.

For a CDHR with D_c cables, each set of poses can establish D_c equations as shown in Equation (45). The unknown calibration variables, ${}^{\mathrm{b}}\boldsymbol{X}_{\mathrm{m}}$, ${}^{\mathrm{v}}\boldsymbol{X}_{\mathrm{e}}$ and ${}^{\mathrm{g}}\boldsymbol{p}_{bi}$, include a total of $12 + 3D_{\mathrm{c}}$ unknowns. To fully solve this calibration problem, data from at least ceil$\left(\frac{12+3D_{\mathrm{c}}}{D_{\mathrm{c}}}\right)$ different poses is required, where ceil$(\cdot)$ represents the ceiling function. Here, we select $N_{\mathrm{calib}} >$ ceil$\left(\frac{12+3D_{\mathrm{c}}}{D_{\mathrm{c}}}\right)$.

Let the superscript k denote the group index of the calibration data. The synchronous calibration optimization function for the N_{calib} sets of data is:

$$\begin{aligned} &\left({}^{\mathrm{b}}\boldsymbol{X}_{\mathrm{m}}, {}^{\mathrm{v}}\boldsymbol{X}_{\mathrm{e}}, {}^{\mathrm{g}}\boldsymbol{p}_{bi}\right) = \arg\min \sum_{k}^{N_{\mathrm{calib}}} \sum_{i}^{D_{\mathrm{c}}} \left\| \hat{l}_i^{(k)} - \widetilde{l_i^{(k)}} \right\|_2 \\ &\text{s.t.} \begin{cases} \widetilde{l_i^{(k)}} = f_{\mathrm{calib}}\left({}^{\mathrm{g}}\widetilde{\boldsymbol{p}}_{bi}, {}^{\mathrm{m}}\boldsymbol{r}_{ai}, {}^{\mathrm{b}}\widetilde{\boldsymbol{X}}_{\mathrm{m}}, {}^{\mathrm{v}}\widetilde{\boldsymbol{X}}_{\mathrm{e}}, {}^{\mathrm{v}}\boldsymbol{T}_{\mathrm{g}}^{(k)}, {}^{\mathrm{b}}\boldsymbol{T}_{\mathrm{e}}^{(k)}\right) \\ i = 1, 2, \cdots, D_{\mathrm{c}},\ k = 1, 2, \ldots, N_{\mathrm{calib}} \end{cases} \end{aligned} \tag{45}$$

https://doi.org/10.3390/s26041147

where $\hat{l}_i^{(k)}$, $\widetilde{l_i^{(k)}}$ are measured and calculated lengths of the ith cable in the kth set of data, respectively.

The overall algorithm process is shown in Algorithm 1. By selecting appropriate initial values for the iterations, the convergence speed of the optimization algorithm can be improved. When the SM is installed at the center of the moving platform and the base frame aligns with the axes of the moving coordinate system, the initial values ${}^{\mathrm{b}}\widetilde{\mathbf{X}}_{\mathrm{m}}^{(0)}$ can be chosen as $\left[0, 0, \frac{h_{\mathrm{m}}}{2}, 0, 0, 0\right]^{\mathrm{T}}$ or $\left[0, 0, -\frac{h_{\mathrm{m}}}{2}, 0, 0, 0\right]^{\mathrm{T}}$, where h_{m} represents the height of the moving platform.

Algorithm 1: Kinematic Parameter Self-Calibration Algorithm for Cable-Driven Hybrid Robots

Input: Structural height h_m, measured cable length set $\left\{\hat{l}_i^{(k)}\right\}$, convergence threshold ϵ, N_c

Output: Calibrated parameters ${}^{b}\mathbf{X}_m$, ${}^{v}\mathbf{X}_e$, ${}^{g}\mathbf{p}_{bi}$.

Initialization: Initialize parameter vector $\boldsymbol{\Theta}^{(0)} = [{}^{b}\widetilde{\mathbf{X}}_m^{(0)}, {}^{v}\widetilde{\mathbf{X}}_e^{(0)}, {}^{g}\widetilde{\mathbf{p}}_{bi}^{(0)}]$, with ${}^{b}\widetilde{\mathbf{X}}_m^{(0)}$ set to $[0, 0, \pm h_m/2, 0, 0, 0]^{\mathrm{T}}$.

Data Collection: Collect cable length data for N_{calib} poses, satisfying $N_{calib} > \mathrm{floor}((12 + 3D_c)/D_c)$.

While $t < N_{max}$ and error $> \epsilon$ do:

Calculate theoretical cable lengths $\widetilde{l}_i^{(k)}$ using Equation (43) with current parameters.

Construct the least-squares objective function: $\min \sum_k \sum_i \|\hat{l}_i^{(k)} - \widetilde{l}_i^{(k)}\|$.

Update parameter vector $\boldsymbol{\Theta}^{(t+1)}$ using the Levenberg–Marquardt (LM) algorithm.

Calculate residuals and check for convergence.

End

Return Optimized parameter set $\boldsymbol{\Theta}^{*}$.

7. Model Predictive Control-Based Trajectory Tracking Method

The model of the CDHR in the generalized discrete joint space at time step k can be expressed as:

$$\boldsymbol{q}_{k+1} = \boldsymbol{A}\boldsymbol{q}_k + \boldsymbol{B}\boldsymbol{u}_k \tag{46}$$

Let

$$\begin{aligned} \delta\boldsymbol{q}_k &= \boldsymbol{q}_k - \boldsymbol{q}_{k-1} \\ \delta\boldsymbol{u}_k &= \boldsymbol{u}_k - \boldsymbol{u}_{k-1} \end{aligned} \tag{47}$$

Let $\hat{\boldsymbol{q}}_k^d$ be the reference generalized joint position and $\boldsymbol{q}_k^d$ the desired generalized joint position. The tracking error between the actual generalized joint position and the reference generalized joint position at time step k is defined as $\boldsymbol{e}_k = \boldsymbol{q}_k - \hat{\boldsymbol{q}}_k^d$. The system state variable is $\boldsymbol{\xi}_k = \left[\delta\boldsymbol{q}_k{}^{\mathrm{T}}, \boldsymbol{e}_k{}^{\mathrm{T}}\right]^{\mathrm{T}}$. Thus, the dynamic augmented system equation can be expressed as:

$$\boldsymbol{\xi}_{k+1} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{O}_{D_{\mathrm{h}}} \\ \boldsymbol{A} & \boldsymbol{I}_{D_{\mathrm{h}}} \end{bmatrix} \boldsymbol{\xi}_k + \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{B} \end{bmatrix} \delta\boldsymbol{u}_k \tag{48}$$

where $\boldsymbol{q} = \mathbf{Q}_{gj} \in \mathbb{R}^{D_{\mathrm{h}}}$ represents the generalized joint positions of the CDHR, $\boldsymbol{A} = \boldsymbol{I}_{D_{\mathrm{h}}} \in \mathbb{R}^{D_{\mathrm{h}} \times D_{\mathrm{h}}}$ is the identity state matrix, $\boldsymbol{B} = \boldsymbol{I}_{D_{\mathrm{h}}} \Delta\boldsymbol{t} \in \mathbb{R}^{D_{\mathrm{h}} \times D_{\mathrm{h}}}$ is the control matrix, and $\Delta\boldsymbol{t}$ represents the discrete time interval. The control input $\boldsymbol{u} = \dot{\boldsymbol{q}}$ denotes the generalized joint velocities of the CDHR.

https://doi.org/10.3390/s26041147

Let $\hat{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{O}_{D_\mathrm{h}} \\ \boldsymbol{A} & \boldsymbol{I}_{D_\mathrm{h}} \end{bmatrix}$, $\hat{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{B} \end{bmatrix}$, the tracking model for the reference trajectory can be designed as:

$$V_M^{O*}(\boldsymbol{\xi}_k, \boldsymbol{q}_k^d) = \min_{\delta \boldsymbol{U}_k, \hat{\boldsymbol{q}}_k} L_p$$
$$\text{s.t.} \begin{cases} \boldsymbol{\xi}_{k+i+1} = \hat{\boldsymbol{A}}\boldsymbol{\xi}_{k+i} + \hat{\boldsymbol{B}}\delta \boldsymbol{u}_{k+i} \\ \delta \boldsymbol{u}_{k+i} \in \mathbb{U} \\ \begin{bmatrix} \boldsymbol{q}_k \\ \boldsymbol{u}_{k-1} \end{bmatrix} \in \mathbb{Q} \times \mathbb{U} \\ \boldsymbol{\zeta}_{k+M} \in \mathbb{O} \end{cases} \tag{49}$$

where $L_p = \sum_{i=0}^{M-1} \left(\left\| \boldsymbol{\xi}_{k+i} \right\|_{\boldsymbol{Q}_1}^2 + \left\| \delta \boldsymbol{u}_{k+i} \right\|_{\boldsymbol{Q}_2}^2 \right) + \left\| \boldsymbol{\xi}_{k+M} \right\|_{\boldsymbol{Q}_3}^2 + \left\| \hat{q}_k - q_k^d \right\|_{\boldsymbol{Q}_4}^2$, $\delta \boldsymbol{U}_k = \left[\delta \boldsymbol{u}_k^\mathrm{T}, \delta \boldsymbol{u}_{k+1}^\mathrm{T}, \ldots, \delta \boldsymbol{u}_{k+M-1}^\mathrm{T} \right]^\mathrm{T}$ represents the incremental control sequence defined over the horizon M. $\boldsymbol{Q}_1 \in \mathbb{R}^{2D_\mathrm{h} \times 2D_\mathrm{h}}$, $\boldsymbol{Q}_2 \in \mathbb{R}^{2D_\mathrm{h} \times 2D_\mathrm{h}}$, $\boldsymbol{Q}_3 \in \mathbb{R}^{2D_\mathrm{h} \times 2D_\mathrm{h}}$ and $\boldsymbol{Q}_4 \in \mathbb{R}^{2D_\mathrm{h} \times 2D_\mathrm{h}}$ are positive definite weighting matrices. $\boldsymbol{\zeta}_k$ is the convex polyhedron constraint set, and $\boldsymbol{\zeta}_k = \left[\boldsymbol{\xi}_k^\mathrm{T}, \hat{\boldsymbol{q}}_k^\mathrm{T} \right]^\mathrm{T}$, where $\mathbb{O}$ represents the Maximal Output Admissible Set (MOAS).

8. Simulation and Experiment Study

8.1. Simulation Study

8.1.1. Simulation System Setup

To validate the methods proposed earlier, a simulation system for the CDHR was built. In the simulation, the parameters are set as follows: R_r = 6 mm, m_p = 0.259 kg, t_{min} = 5 N and t_{max} = 50 N. The DH parameters of the SM are listed in Table 5, where the masses of links 1, 2, 3, and 4 were 0.1290, 0.2240, 0.2240, and 0.4695 kg, respectively.).

Table 5. PSO parameters for optimizing the anchor seat distribution.

Maximum Iterations	Population Size	Personal Learning Factor	Global Learning Factor	Initial Inertia Weight	Final Inertia Weight
2000	100	2	2	0.9	0.4

8.1.2. Numerical Optimization of the Dynamic Anchor Seat Position

The reachable workspace of the CDHR is uniformly discretized. The parameter values are listed in Table 6, and the scatter plot after discretization is shown in Figure 15.

Table 6. Optimized dynamic anchor seat position distribution.

Name	A_1 (m)	A_2 (m)	A_3 (m)	A_4 (m)	A_5 (m)	A_6 (m)	A_7 (m)	A_8 (m)
Position	$\begin{bmatrix} -0.0623 \\ -0.0258 \\ -0.0150 \end{bmatrix}$	$\begin{bmatrix} -0.0258 \\ 0.0623 \\ -0.0150 \end{bmatrix}$	$\begin{bmatrix} 0.0258 \\ 0.0623 \\ -0.0150 \end{bmatrix}$	$\begin{bmatrix} 0.0623 \\ -0.0258 \\ -0.0150 \end{bmatrix}$	$\begin{bmatrix} -0.0258 \\ -0.0623 \\ 0.0150 \end{bmatrix}$	$\begin{bmatrix} -0.0623 \\ 0.0258 \\ 0.0150 \end{bmatrix}$	$\begin{bmatrix} 0.0623 \\ 0.0258 \\ 0.0150 \end{bmatrix}$	$\begin{bmatrix} 0.0258 \\ -0.0623 \\ 0.0150 \end{bmatrix}$

The specific parameters used in the PSO iterations are listed in Table 7. The optimization results are shown in Figure 16 based on the model described in Section 5. After 2000 iterations, the objective function value reached 0.8075, corresponding to a global condition number of 0.1925, indicating a significant improvement in the flexibility of the CDHR. The distribution of the optimized dynamic platform anchor seats is listed in Table 8, and the optimized point distribution is shown in Figure 17.

https://doi.org/10.3390/s26041147

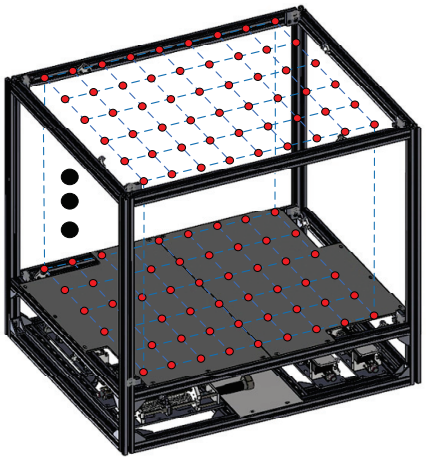


Figure 15. Illustration of uniformly discretized sampled pose points.

Table 7. Simulation results of self-calibration.

Variable	True Value		Measured Average Value		Error ΔX	
	Position (mm)	Orientation (°)	Position (mm)	Orientation (°)	Position (mm)	Orientation (°)
${}^{g}\boldsymbol{p}_{b1}$	−334.00 −370.00 −900.00	-	−334.05 −370.03 −900.04	-	0.07	-
${}^{g}\boldsymbol{p}_{b2}$	−334.00 385.00 −885.00	-	−334.03 385.01 −885.04	-	0.05	-
${}^{g}\boldsymbol{p}_{b3}$	−334.00 370.00 −280.00	-	−334.03 370.02 −280.09	-	0.10	-
${}^{g}\boldsymbol{p}_{b4}$	−334.00 −385.00 −295.00	-	−334.04 −385.02 −295.08	-	0.09	-
${}^{g}\boldsymbol{p}_{b5}$	316.00 −370.00 −900.00	-	316.04 −370.07 −900.02	-	0.08	-
${}^{g}\boldsymbol{p}_{b6}$	316.00 385.00 −885.00	-	316.03 385.04 −855.02	-	0.03	-
${}^{g}\boldsymbol{p}_{b7}$	316.00 370.00 280.00	-	316.07 370.01 −280.06	-	0.05	-
${}^{g}\boldsymbol{p}_{b8}$	316.00 −385.06 −295.10	-	316.02 −385.10 −295.08	-	0.05	-
${}^{b}\boldsymbol{X}_{p}$	0 0 0	−0.04 −0.04 −180.00	−0.03 −0.06 −0.07	−0.04 −0.06 −180.07	0.10	0.07
${}^{v}\boldsymbol{X}_{e}$	6.30 38.39 92.70	−89.22 −89.22 −90.79	6.29 38.40 92.61	−89.25 −89.25 −90.81	0.09	0.05

https://doi.org/10.3390/s26041147

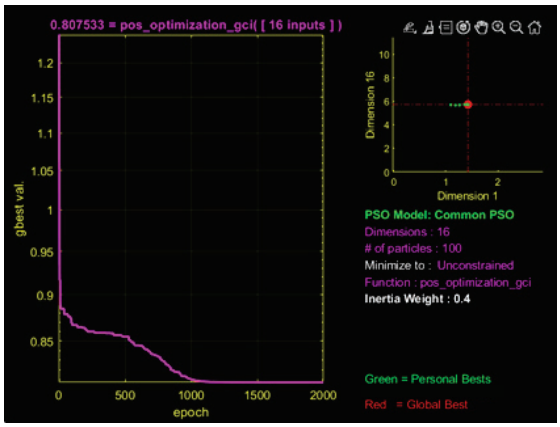
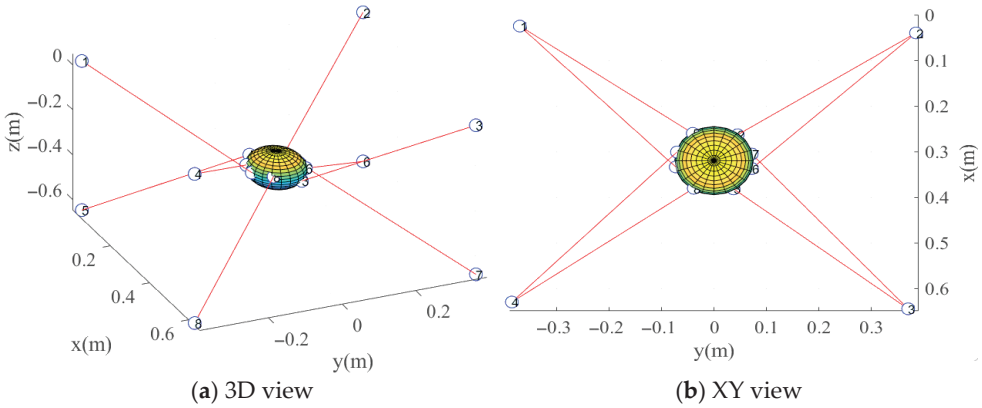


Figure 16. Optimization interface for dynamic anchor seat pose based on the PSO method.

Table 8. Parameters of several typical trajectories.

	Name	Trajectory Equations
Case A	Straight line trajectory	$\begin{cases} \boldsymbol{X}_{ee}^{ini} = [0.0074,\ -0.1995,\ -0.6164,\ 1.5732,\ -1.3077,\ -0.0025]^{T} \\ \boldsymbol{X}_{ee}^{fin} = [-0.2925785,\ 0.1005394,\ -0.3163925,\ 1.5732,\ -1.3077,\ -0.0025]^{T} \end{cases}$ where $\boldsymbol{X}_{ee}^{ini}$ and $\boldsymbol{X}_{ee}^{fin}$ are the starting pose and the ending pose of the end-effector, respectively.
Case B	Spiral trajectory	$\begin{cases} \boldsymbol{X}_{ee}^{ini} = [0.0074,\ -0.1995,\ -0.6164,\ 1.5732,\ -1.3077,\ -0.0025]^{T} \\ \boldsymbol{X}_{center} = [-1.9926,\ 0,\ -0.5000,\ 1.5732,\ -1.3077,\ -0.0025]^{T} \\ x_{ee} = 0.0074 - t_i \cdot h_{step} \\ y_{ee} = r_{ctraj} \cdot \cos(\theta_{ini} + t_i \cdot \theta_{step}) \\ z_{ee} = -0.6164 - r_{ctraj} \cdot \sin(\theta_{ini} + t_i \cdot \theta_{step}) \\ \alpha_{ee} = 1.5732 \\ \beta_{ee} = -1.3077 \\ \gamma_{ee} = -0.0025 \end{cases}$ where $\boldsymbol{X}_{ee}^{ini}$ and $\boldsymbol{X}_{ee}^{fin}$ are the starting pose and the ending pose of the end−effector, respectively, r_{ctraj} is the radius of the circle, $r_{ctraj} = 0.2309$, $h_{step} = 0.01$, θ_{ini} is the initial angle of the circle trajectory corresponding to $\boldsymbol{X}_{ee}^{ini}$, $\theta_{step} = \pi/10$.
Case C	"2"-type trajectory	$\begin{cases} \boldsymbol{X}_{ee}^{ini} = [0.0074,\ -0.1995,\ -0.6164,\ 1.5732,\ -1.3077,\ -0.0025]^{T} \\ \boldsymbol{X}_{ee}^{fin} = [0.0074,\ 0.1005,\ -0.3164,\ 1.5732,\ -1,\ 3077,\ -0.0025]^{T} \end{cases}$ where the circular arc segment of shape 2 is an 1800 semicircle, and the remaining part is a straight-line segment.



(**a**) 3D view (**b**) XY view

Figure 17. Visualization of the optimized distribution.

From the illustration of the optimization results, the distribution of the eight dynamic anchor seats resembles a regular octagon. When the anchor seats are arranged in this shape

https://doi.org/10.3390/s26041147

on a horizontal cross-section, the mobility of the moving platform is maximized, and the probability of collisions between the cables and the platform is minimized. The designed distribution of the anchor seats on the moving platform was symmetric with respect to the platform's center of mass, with square holes left in both the upper and lower layers to accommodate the wiring of the SM motors. Eye bolts were installed on the sides of the moving platform, with the center of each eyebolt ring corresponding to the location of the dynamic anchor seat. The eye bolts were secured with nuts, and the force direction of the eyebolts was aligned with the direction of the cable. The fixation method is non-tightening, allowing the eyebolts to rotate axially to some extent under the cable tension.

8.1.3. Simulation of Self-Calibration of Structural Parameters

In this simulation, $N_{\text{calib}} = 30$, ${}^{\text{b}}\widetilde{\boldsymbol{X}}_{\text{m}}^{(0)} = [0, 0, 0, 0, 0, -\pi]^{\text{T}}$, ${}^{\text{v}}\widetilde{\boldsymbol{X}}_{\text{e}}^{(0)} = [0, 0, 0, 0, 0, 0]^{\text{T}}$. The error between the measured and true values can be described as:

$$\Delta X = \left\| (\cdot) - \widetilde{(\cdot)} \right\|_2 \tag{50}$$

where $(\cdot)$ and $\widetilde{(\cdot)}$ represent the true and calibrated values of the unknowns, respectively.

As shown in Table 7, these errors are at the sub-millimeter and sub-tenth-degree level (max. 0.10 mm and 0.07°), which indicates that the proposed self-calibration can provide an accurate geometric mapping for subsequent planning and control. Considering the prototype scale (workspace on the order of hundreds of millimeters), the residual positioning error corresponds to a relative error well below 0.1%, and thus the calibration accuracy is sufficient for the intended manipulation and tracking tasks.

8.1.4. Simulation of Trajectory Tracking

To verify the established model, three typical trajectories were simulated and analyzed. The specific parameters of the simulations are presented in Table 8, and the results are shown in Table 9.

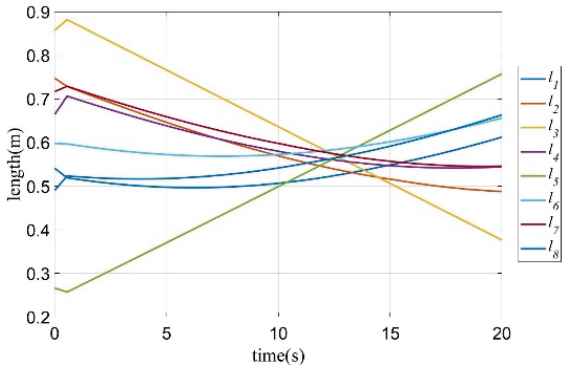
Table 9. Simulation results of the several typical trajectories.

Categories	Name	Maximum Dimension Errors						Maximum 2-Norm Error	
		x (mm)	y (mm)	z (mm)	α (°)	β (°)	γ (°)	Position (mm)	Attitude (°)
Case A	Straight line trajectory	0.0460	0.0360	0.0190	0.0099	0.0062	0.0116	0.0480	0.0147
Case B	Spiral trajectory	0.0390	0.0300	0.0180	0.0124	0.0050	0.0129	0.0440	0.0179
Case C	"2"-type trajectory	0.0400	0.0300	0.0140	0.0117	0.0049	0.0128	0.0440	0.0172

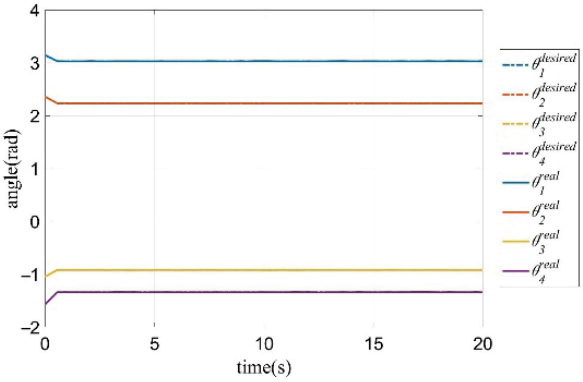
The initial pose is chosen as a representative central feasible pose (inside the FCW) to avoid boundary effects and ensure all constraints are satisfied at the start of tracking. Assume that the initial generalized pose is $\boldsymbol{Q}_{\text{gp}}^{(0)}$ = [0.1550, −0.2000, −0.7400, 0, −1.5701, 0, 3.14159274, 2.3561945, −1.04719758, −1.57079637]$^{\text{T}}$, $\lambda_1 = 1$, $\lambda_2 = 1$, the number of sampling points is N_i = 50,000. The pose variation trajectory of the mobile platform during the linear trajectory tracking process is shown in Figure 18a,b, the trajectory curve of the generalized joint is shown in Figure 18c, and the motion trajectory of the tips of the CDHR is shown in Figure 18d. From the simulation results, it can be observed that the mobile platform produces only a small deflection to avoid generating a large cable tension.

In the spiral trajectory tracking process, the pose variation curves of the mobile platform are shown in Figure 19a,b, the variation curves of the generalized joint are shown in Figure 19c, and the comparison results of the actual and desired trajectories of the CDHR are shown in Figure 19d. Figure 19 shows that the poses of the mobile platform and arm shape of the RSR are motional complementary.

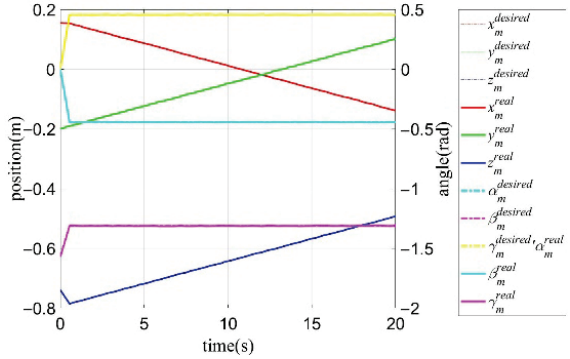
https://doi.org/10.3390/s26041147



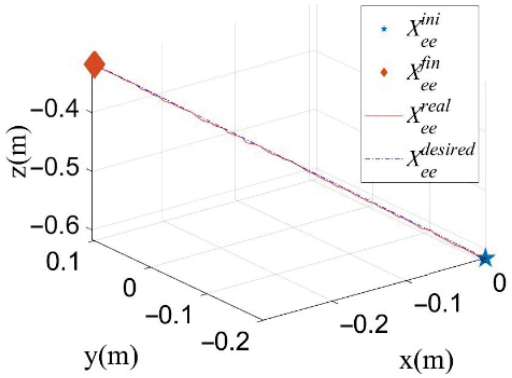
(**a**) The length variation curve of the cables



(**b**) The angle variation curve of the serial joints

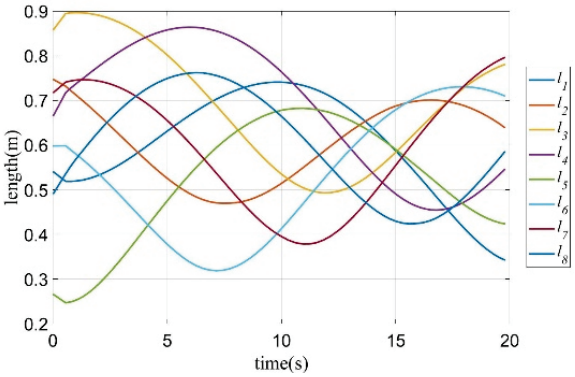


(**c**) The pose variation curve of the platform

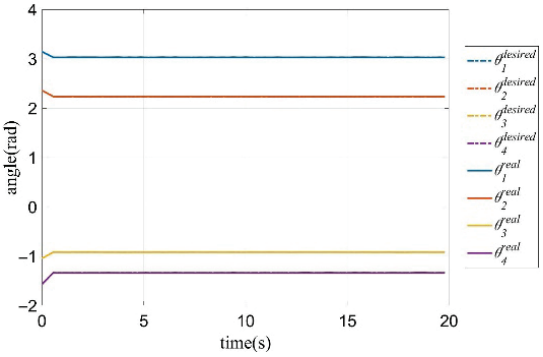


(**d**) The position variation curve of the end-effector

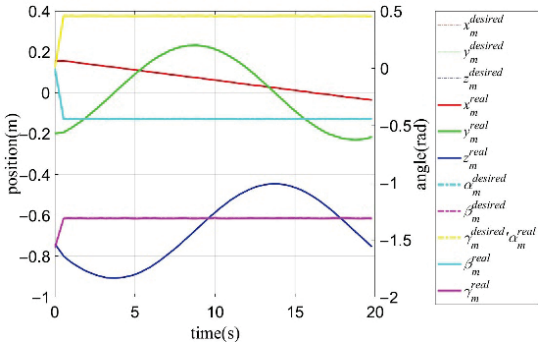
Figure 18. Trajectory variation curves of the mobile platform pose, generalized joints and the end-effector with Case A.



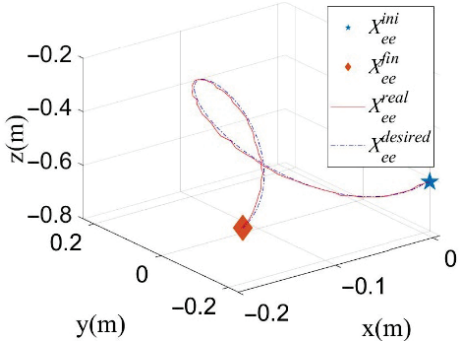
(**a**) The length variation curve of the cables



(**b**) The angle variation curve of the serial joints



(**c**) The pose variation curve of the platform

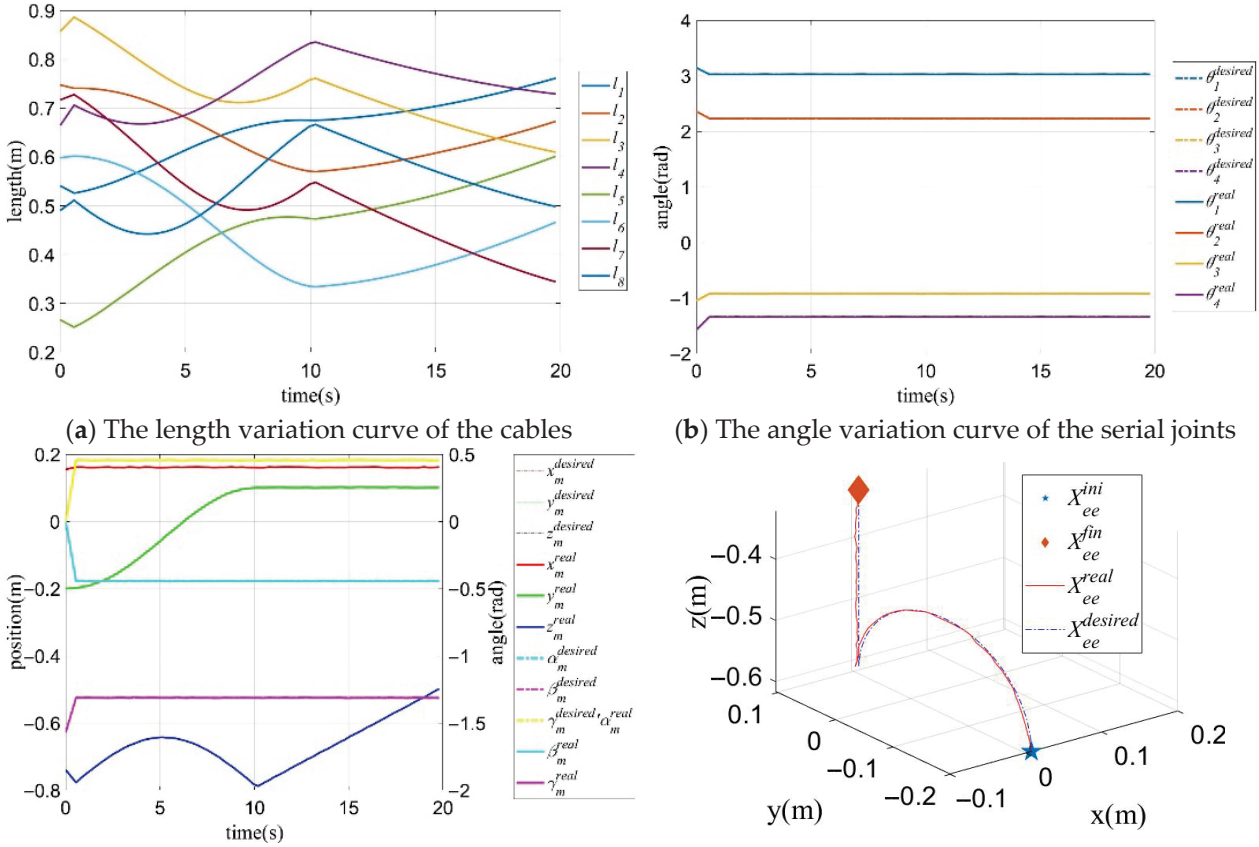


(**d**) The position variation curve of the end-effector

Figure 19. Trajectory variation curves of the mobile platform pose, generalized joints and the end-effector with Case B.

https://doi.org/10.3390/s26041147

In the "2"-type trajectory tracking process, the pose variation curves of the mobile platform are shown in Figure 20a,b, the variation curves of the generalized joint are shown in Figure 20c, and the comparison results of the actual trajectory and the desired trajectory of the CDHR are shown in Figure 20d. Figure 20 shows that the poses of the mobile platform and arm shape of the RSR are motional complementary.



(**a**) The length variation curve of the cables

(**b**) The angle variation curve of the serial joints

(**c**) The pose variation curve of the platform

(**d**)The position variation curve of the end-effector

Figure 20. Trajectory variation curves of the mobile platform pose, generalized joints and the end-effector with Case C.

As summarized in Table 9, the maximum position 2-norm error is 0.044–0.048 mm, and the maximum attitude 2-norm error is 0.0147–0.0179° across the three representative trajectories. These errors are significantly smaller than typical tolerances required for platform positioning and end-effector manipulation in our prototype-scale tasks, indicating that the proposed method achieves stable and accurate tracking performance in simulation.

8.2. Experiment Study

8.2.1. Experiment System Setup

As shown in Figure 21, the CDHR prototype consists of a CDPR with eight cables, a moving platform, a 4-DOF SM, and two vision systems. Fixed housing and bottom hardware support beams were assembled using aluminum alloy materials from the Euro-standard 3030 series, along with the corresponding angle brackets and L-shaped aluminum alloy connectors. The tension sensor support plate was machined from high-strength wear-resistant 304 stainless steels. The motors used to drive the winches were MAXON RE25 models equipped with 500-line encoders. The motor reduction ratio was approximately 128, whereas the remaining motors exhibited a reduction ratio of approximately 372. Incremental encoders were connected to the transmitters with RS485 transceiver chips, and the lower controller read the encoder counts stored in the transmitters.

https://doi.org/10.3390/s26041147

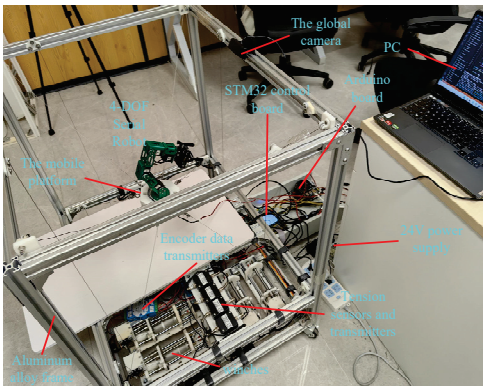


Figure 21. The CDHR prototype.

Owing to the differences in the cable winding angles and input impedance of the voltage acquisition channels, the linear ratio between the tension of each cable and the corresponding sensor's output analog voltage varies. To control the cable tension more accurately, it was necessary to calibrate the parameters of the tension sensors prior to starting the experiment. The specific calibration experiment is detailed in Supplementary Information.

8.2.2. Self-Calibration Experiment of System Parameters

Before starting the experiment, the QR code for calibration was mounted on an aluminum frame, as shown in Figure 22. The number of calibration datasets was chosen as $N_{\text{calib}} = 15$, and the initial pose of the moving platform was set as $\boldsymbol{X}_{\text{m0}} = [0.3350\text{m}, -0.0120\text{m}, -0.3550\text{m}, 0, 0, 0]^{\text{T}}$. Based on Equation (16), the initial cable lengths were calculated as $\boldsymbol{l}_0 = [0.5808, 0.5905, 0.5902, 0.5815, 0.5092, 0.5307, 0.5301, 0.5093]^{\text{T}}$. The initial joint angles of SM were $\boldsymbol{\Theta}_{\text{s0}} = [0, 2.2689, -1.0472, -1.0472](^\circ)$. The joint angle variation was within the range of $[-15^\circ, 15^\circ]$. Based on the modeling in Section 6, the self-calibration results of the system are shown in Table 10. The identified parameters in Table 10 provide the geometric mappings among the fixed anchor seats, the moving platform, and the hand–eye system, which are required for accurate trajectory planning and control. Although ground-truth values are not directly available in the physical experiment, the effectiveness of the self-calibration is reflected in the subsequent trajectory tracking results, where the end-effector paths closely follow the desired ones under the calibrated structural parameters (see Table 11).

Table 10. Self-calibration experimental results.

Value	Calibrated Value	
	Position (m)	**Orientation (°)**
${}^{\text{g}}\boldsymbol{p}_{b1}$	$[0.3513, -0.0093, -0.6181]^{\text{T}}$	-
${}^{\text{g}}\boldsymbol{p}_{\text{b2}}$	$[-0.3698, -0.0079, -0.5995]^{\text{T}}$	-
${}^{\text{g}}\boldsymbol{p}_{\text{b3}}$	$[-0.3633, -0.0169, -0.0153]^{\text{T}}$	-
${}^{\text{g}}\boldsymbol{p}_{\text{b4}}$	$[0.3817, -0.0198, -0.0288]^{\text{T}}$	-
${}^{\text{g}}\boldsymbol{p}_{b5}$	$[0.3495, 0.6103, -0.6187]^{\text{T}}$	-
${}^{\text{g}}\boldsymbol{p}_{b6}$	$[-0.3784, 0.6182, -0.6076]^{\text{T}}$	-
${}^{\text{g}}\boldsymbol{p}_{b7}$	$[-0.3442, 0.6081, -0.0285]^{\text{T}}$	-
${}^{\text{g}}\boldsymbol{p}_{b8}$	$[0.3575, 0.6032, -0.0533]^{\text{T}}$	-
${}^{\text{v}}\boldsymbol{X}_{\text{e}}$	$[0.0235, 0.0198, 0.1066]^{\text{T}}$	$[0.0344, -90.7450, 0.0344]^{\text{T}}$
${}^{\text{b}}\boldsymbol{X}_{\text{m}}$	$[-0.0190, -0.0061, -0.0093]^{\text{T}}$	$[0.0688, 0.1375, 180.5963]^{\text{T}}$

https://doi.org/10.3390/s26041147

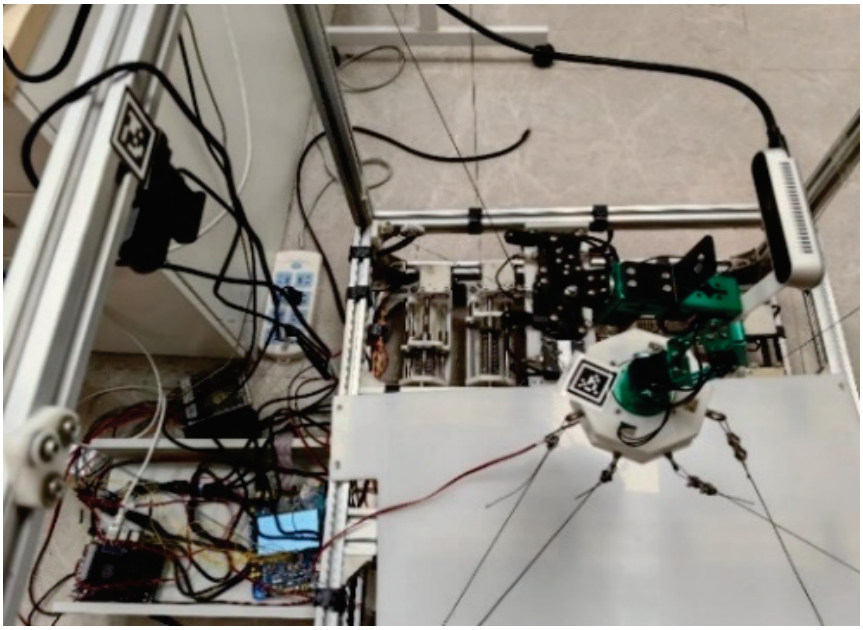


Figure 22. Calibration scene of the CDHR.

Table 11. Results of the several typical trajectories.

Categories	Name	Maximum Dimension Errors						Maximum 2-Norm Error	
		x (mm)	y (mm)	z (mm)	α (°)	β (°)	γ (°)	Position (mm)	Attitude (°)
Case A	Straight line trajectory	0.1167	0.0091	0.0023	0.0169	0.3076	0.4794	0.1171	0.5689
Case B	Arc trajectory	0.1182	0.1899	0.0014	0.0154	0.3314	0.4338	0.1907	0.5462
Case C	"2"-type trajectory	0.1003	0.1958	0.0020	0.0181	0.3257	0.4902	0.2224	0.5887

8.2.3. Trajectory Tracking Experiment

The CDHR was commanded to track three representative trajectories: a straight line (Case A), an arc (Case B), and a "2"-type trajectory (Case C). The "2"-type trajectory parameters are shown in Table 12.

Table 12. Starting point pose of the tracking trajectory.

Semicircular Arc Starting Point		Semicircular Arc Endpoint (Line Starting Point)		Line Segment Endpoint	
Position (m)	Attitude (°)	Position (m)	Attitude (°)	Position (m)	Attitude (°)
$\begin{bmatrix} 0.355 \\ -0.040 \\ -0.359 \end{bmatrix}^T$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}^T$	$\begin{bmatrix} 0.355 \\ 0.150 \\ -0.359 \end{bmatrix}^T$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}^T$	$\begin{bmatrix} 0.250 \\ 0.150 \\ -0.359 \end{bmatrix}^T$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}^T$

The results of the several typical trajectories are summarized in Table 11. As shown in Figure 23, during the motion process, the rate of change in the actual cable length closely matched the expected rate of change (i.e., the maximum cable length error was 2.1582 mm). The desired pose of the moving platform also closely matched the actual pose (i.e., the maximum position error was 3.2751 mm and the maximum orientation error was 1.2193°). The measured end-effector paths closely follow the desired ones, and the pose variation in the mobile platform remains small, indicating that the platform motion and the arm deformation are still kinematically complementary in real hardware. Based on the trajectory tracking method described in Section 7, the state diagram of the tracking process is shown in

https://doi.org/10.3390/s26041147

Figure 24, and the comparison between the desired and actual cable length curves is shown in Figure 25. The experimental results indicate that there is a lag between the desired and actual cable lengths, which is primarily caused by the communication delays between the upper and lower control systems. We found that the trajectory data on the upper computer is generated according to the upper-computer timeline, whereas the actual cable lengths on the lower computer are sampled according to the lower-computer timeline. A time offset exists between the two, and communication between the upper and lower computers incurs latency, which causes the measured cable lengths to lag behind the desired cable lengths. Based on the literature on distributed robotic control and embedded communication [36,37], the observed lag between desired and measured cable lengths is consistent with typical PC–microcontroller systems communicating over serial/fieldbus links without hardware-level clock synchronization. For example, experiments on Ethernet-based and CAN/serial robot controllers report end-to-end communication delays on the order of 5–20 ms, and show that a 10 ms delay can shift the measured joint/cable states by about 1–2 sampling periods when the control loop runs at 100–200 Hz, producing an apparent phase lag in position or length responses. These works further confirm that unsynchronized clocks between master and slave controllers lead to timestamp misalignment and additional effective delay in logged data, which directly manifests as the "delayed" actual trajectory relative to the reference one. It should be noted that the current implementation assumes reliable sensor feedback (e.g., encoder readings and continuous marker visibility for vision-based measurements) and uses the adopted modeling assumptions in Sections 4 and 7. For systems with higher DOFs (more cables/joints) or heavier payloads, the calibration and optimization variables will increase, potentially raising computational cost and sensitivity to unmodeled effects (e.g., cable elasticity and structural compliance). These aspects will be further addressed in future work by incorporating a full dynamic model and deformation compensation.

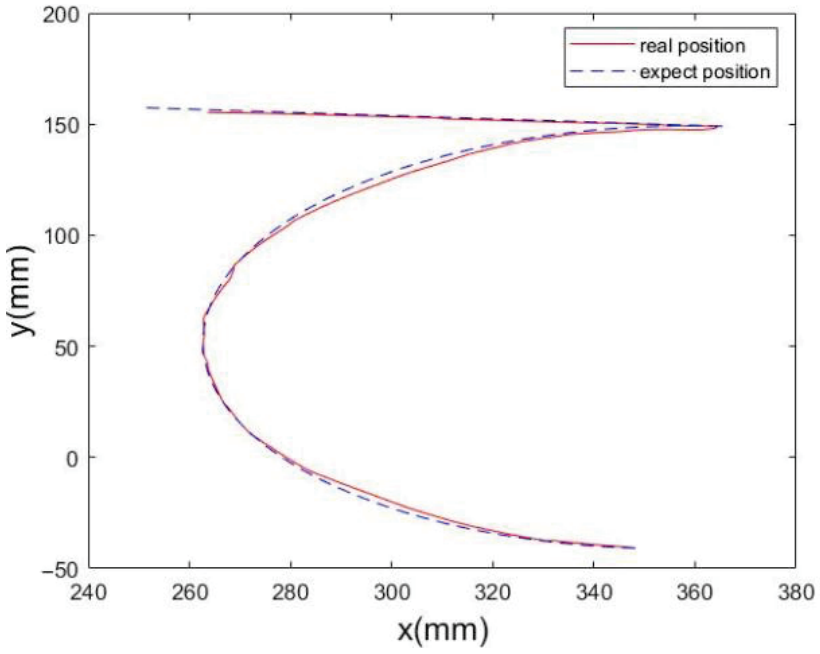


Figure 23. Comparison of actual end-effector pose vs. desired end-effector pose.

https://doi.org/10.3390/s26041147

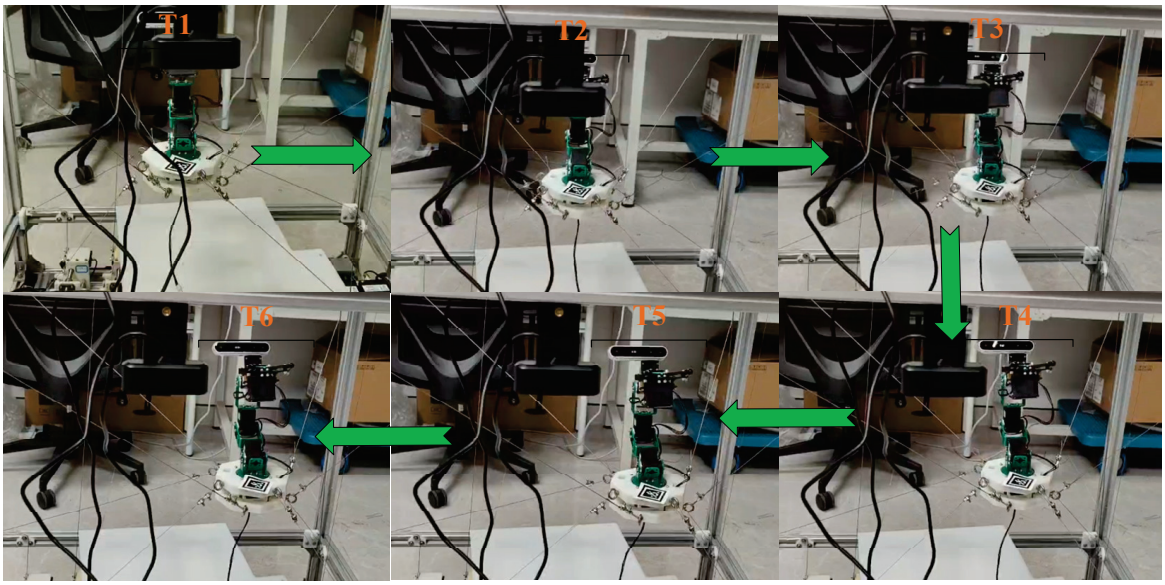


Figure 24. Process diagram of the trajectory tracking experiment.

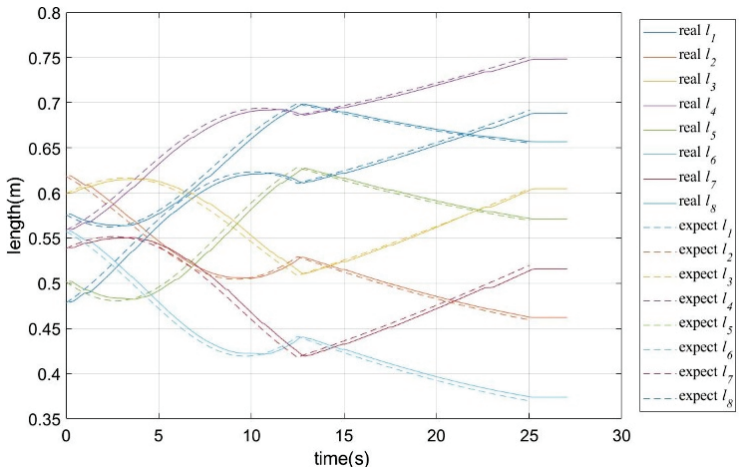


Figure 25. Comparison of desired cable length vs. actual cable length.

9. Discussion

The CDHR shows promising potential in fields such as rapid spatial transportation and autonomous control in complex environments. In this study, we propose a CDHR prototype with high integration and a compact structure. The moving platform and the on-platform SM are optimized for operation in confined and complex spaces. A kinematic modeling method for a hybrid robot was proposed, incorporating both serial and parallel elements, and a corresponding static model was developed. Additionally, the force-coupling effects between the SM and moving platform under collision constraints, along with the force-closure conditions of the hybrid robot, were analyzed. The workspace of the hybrid robot is examined under various conditions. Based on global vision and hand-eye coordination, a synchronous self-calibration method was proposed for the multi-mapping relationships between fixed anchor seats, the arm platform, and hand-eye systems, addressing the time-consuming issue of step-by-step structural parameter calibration. Using the calibrated structural parameters, a trajectory tracking control method for the CDHR is presented. Finally, the performances of the prototype and proposed methods were fully validated using both the simulation system and physical prototype. Future work will build upon a full dynamic model of the CDHR to enable compliant hybrid force–motion control, which jointly regulates end-effector motion and contact force during physical interaction. We

https://doi.org/10.3390/s26041147

will adopt task-space impedance/admittance control to shape end-effector compliance: impedance control defines a desired dynamic relation between motion error and interaction force (virtual mass–spring–damper behavior), while admittance control maps measured force into motion commands. Compliance will be allocated across the hybrid system, with the CDPR responsible for low-frequency large-scale motions and the on-platform serial mechanism providing high-bandwidth fine dexterity. In parallel, we will model material flexibility by representing cable stretch, creep, and hysteresis together with platform and idler compliance, and we will compensate deformation through real-time stretch estimation from tension and length residuals.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/s26041147/s1, Video S1: Simulation and Experimental Video of a Cable-Driven Hybrid Robot.

Author Contributions: Z.X.: Writing—original draft, Conceptualization. Z.Y.: Writing—review & Methodology. J.H.: Writing—review & editing. B.Z.: Writing—review & editing, Conceptualization. J.P.: Writing—review & editing, Conceptualization, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key R&D Program of China under Grant 2022YFB4703100, the Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515010540, the Shenzhen Science and Technology Program under Grant JCYJ20220530150006014 and ZDCY20250901104359008, and the University-Industry Collaborative Education Program under Grant 240506429202244.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Osumi, H.; Utsugi, Y.; Koshikawa, M. Development of a Manipulator Suspended by Parallel Wire Structure. In *Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000) (Cat. No.00CH37113)*; IEEE: New York, NY, USA, 2000; Volume 1, pp. 498–503.
2. Tang, X. An Overview of the Development for Cable-Driven Parallel Manipulator. *Adv. Mech. Eng.* **2014**. [CrossRef]
3. An, H.; Yu, D.; Xu, W.; Yuan, H. Modeling and Cooperation Method for Movable Anchor Winch Cable-Driven Parallel Robot. *IEEE/ASME Trans. Mechatron.* **2024**, *30*, 3584–3595. [CrossRef]
4. Xu, W.; Yan, P.; Wang, F.; Yuan, H.; Liang, B. Vision-Based Simultaneous Measurement of Manipulator Configuration and Target Pose for an Intelligent Cable-Driven Robot. *Mech. Syst. Signal Process.* **2022**, *165*, 108347. [CrossRef]
5. Youssef, K.; Otis, M.J.-D. Reconfigurable Fully Constrained Cable Driven Parallel Mechanism for Avoiding Interference between Cables. *Mech. Mach. Theory* **2020**, *148*, 103781. [CrossRef]
6. Bales, I.; Zhang, H. A Six Degrees-of-Freedom Cable-Driven Robotic Platform for Head–Neck Movement. *Sci. Rep.* **2024**, *14*, 8750. [CrossRef]
7. Tang, S.; Yao, J.; Yu, Y. Design and Mechanical Analysis of a Novel Modular Bionic Earthworm Robot with Upright Functionality. *Sci. Rep.* **2025**, *15*, 3829. [CrossRef] [PubMed]
8. Chen, Q.; Zi, B.; Sun, Z.; Li, Y.; Xu, Q. Design and Development of a New Cable-Driven Parallel Robot for Waist Rehabilitation. *IEEE/ASME Trans. Mechatron.* **2019**, *24*, 1497–1507. [CrossRef]
9. An, H.; Yuan, H.; Tang, K.; Xu, W.; Wang, X. A Novel Cable-Driven Parallel Robot with Movable Anchor Points Capable for Obstacle Environments. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 5472–5483. [CrossRef]
10. Rasheed, T.; Long, P.; Caro, S. Wrench-Feasible Workspace of Mobile Cable-Driven Parallel Robots. *J. Mech. Robot.* **2019**, *12*, 031009. [CrossRef]
11. Pinto, A.M.; Moreira, E.; Lima, J.; Sousa, J.P.; Costa, P. A Cable-Driven Robot for Architectural Constructions: A Visual-Guided Approach for Motion Control and Path-Planning. *Auton. Robot.* **2017**, *41*, 1487–1499. [CrossRef]

https://doi.org/10.3390/s26041147

12. Chaikalis, D.; Khorrami, F.; Tzes, A. Adaptive Control Approaches for an Unmanned Aerial Manipulation System. In *Proceedings of the 2020 International Conference on Unmanned Aircraft Systems (ICUAS)*; IEEE: New York, NY, USA, 2020; pp. 498–503.
13. El-Ghazaly, G.; Gouttefarde, M.; Creuze, V. Hybrid Cable-Thruster Actuated Underwater Vehicle-Manipulator Systems: A Study on Force Capabilities. In *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: New York, NY, USA, 2015; pp. 1672–1678.
14. Gao, B.; Song, H.; Zhao, J.; Guo, S.; Sun, L.; Tang, Y. Inverse Kinematics and Workspace Analysis of a Cable-Driven Parallel Robot with a Spring Spine. *Mech. Mach. Theory* **2014**, *76*, 56–69. [CrossRef]
15. Lau, D.; Eden, J.; Tan, Y.; Oetomo, D. CASPR: A Comprehensive Cable-Robot Analysis and Simulation Platform for the Research of Cable-Driven Parallel Robots. In *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: New York, NY, USA, 2016; pp. 3004–3011.
16. Pham, C.B.; Yeo, S.H.; Yang, G.; Kurbanhusen, M.S.; Chen, I.-M. Force-Closure Workspace Analysis of Cable-Driven Parallel Mechanisms. *Mech. Mach. Theory* **2006**, *41*, 53–69. [CrossRef]
17. Diao, X.; Ma, O. A Method of Verifying Force-Closure Condition for General Cable Manipulators with Seven Cables. *Mech. Mach. Theory* **2007**, *42*, 1563–1576. [CrossRef]
18. Yang, G.; Pham, C.B.; Yeo, S.H. Workspace Performance Optimization of Fully Restrained Cable-Driven Parallel Manipulators. In *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*; IEEE: New York, NY, USA, 2006; pp. 85–90.
19. Song, C.; Lau, D. Workspace-Based Model Predictive Control for Cable-Driven Robots. *IEEE Trans. Robot.* **2022**, *38*, 2577–2596. [CrossRef]
20. Nguyen, D.Q.; Gouttefarde, M. Study of Reconfigurable Suspended Cable-Driven Parallel Robots for Airplane Maintenance. In *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*; IEEE: New York, NY, USA, 2014; pp. 1682–1689.
21. Arai, T.; Matsumura, S.; Yoshimura, Y.; Osumi, H. A Proposal for a Wire Suspended Manipulator: A Kinematic Analysis. *Robotica* **1999**, *17*, 3–9. [CrossRef]
22. Khalilpour, S.A.; Khorrambakht, R.; Taghirad, H.D.; Cardou, P. Robust Cascade Control of a Deployable Cable-Driven Robot. *Mech. Syst. Signal Process.* **2019**, *127*, 513–530. [CrossRef]
23. Tjahjowidodo, T.; Zhu, K.; Dailey, W.; Burdet, E.; Campolo, D. Multi-Source Micro-Friction Identification for a Class of Cable-Driven Robots with Passive Backbone. *Mech. Syst. Signal Process.* **2016**, *80*, 152–165. [CrossRef]
24. Yuan, H.; You, X.; Zhang, Y.; Zhang, W.; Xu, W. A Novel Calibration Algorithm for Cable-Driven Parallel Robots with Application to Rehabilitation. *Appl. Sci.* **2019**, *9*, 2182. [CrossRef]
25. Borgstrom, P.H.; Borgstrom, N.P.; Stealey, M.J.; Jordan, B.; Sukhatme, G.S.; Batalin, M.A.; Kaiser, W.J. Design and Implementation of NIMS3D, a 3-D Cabled Robot for Actuated Sensing Applications. *IEEE Trans. Robot.* **2009**, *25*, 325–339. [CrossRef]
26. Zhang, Z.; Xie, G.; Shao, Z.; Gosselin, C. Kinematic Calibration of Cable-Driven Parallel Robots Considering the Pulley Kinematics. *Mech. Mach. Theory* **2022**, *169*, 104648. [CrossRef]
27. Wang, H.; Gao, T.; Kinugawa, J.; Kosuge, K. Finding Measurement Configurations for Accurate Robot Calibration: Validation with a Cable-Driven Robot. *IEEE Trans. Robot.* **2017**, *33*, 1156–1169. [CrossRef]
28. Ma, Q.; Goh, Z.; Ruan, S.; Chirikjian, G.S. Probabilistic Approaches to the AXB = YCZ Calibration Problem in Multi-Robot Systems. *Auton. Robot.* **2018**, *42*, 1497–1520. [CrossRef]
29. Peng, J.; Xu, W.; Wang, F.; Han, Y.; Liang, B. A Hybrid Hand–Eye Calibration Method for Multilink Cable-Driven Hyper-Redundant Manipulators. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 5010413. [CrossRef]
30. Harms, S.; Bizcocho, C.G.; Wakizono, H.; Murasaki, K.; Kawagoe, H.; Nagaoka, K. Tetherbot: Experimental Demonstration and Path Planning of Cable-Driven Climbing in Microgravity. *Robotics* **2024**, *13*, 130. [CrossRef]
31. Michelin, M.; Hervé, P.E.; Tempier, O.; Izard, J.B.; Gouttefarde, M. Path Following Demonstration of a Hybrid Cable-Driven Parallel Robot. In *Proceedings of the Cable-Driven Parallel Robots*; Gouttefarde, M., Bruckmann, T., Pott, A., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 323–335.
32. Song, J.; Wang, M.; Wu, J.; Liu, Q.; Yang, S. Workspace Analysis and Dynamic Modeling of 6-DoF Multi-Pattern Cable-Driven Hybrid Mobile Robot. *Machines* **2025**, *13*, 659. [CrossRef]
33. Peng, J.; Guo, Y.H.; Meng, D.; Han, Y. Kinematics, Statics Modeling and Workspace Analysis of a Cable-Driven Hybrid Robot. *Multibody Syst. Dyn.* **2023**, *61*, 163–193. [CrossRef]
34. Sun, C.; Gao, H.; Liu, Z.; Xiang, S.; Yu, H.; Li, N.; Deng, Z. Design and Optimization of Three-Degree-of-Freedom Planar Adaptive Cable-Driven Parallel Robots Using the Cable Wrapping Phenomenon. *Mech. Mach. Theory* **2021**, *166*, 104475. [CrossRef]
35. Olson, E. AprilTag: A Robust and Flexible Visual Fiducial System. In *Proceedings of the 2011 IEEE International Conference on Robotics and Automation*; IEEE: New York, NY, USA, 2011; pp. 3400–3407.

https://doi.org/10.3390/s26041147

36. Niemeyer, G.; Slotine, J.-J.E. Telemanipulation with Time Delays. *Int. J. Robot. Res.* **2004**, *23*, 873–890. [CrossRef]
37. Middleton, R.H.; Goodwin, G.C. *Digital Control and Estimation: A Unified Approach*; Prentice Hall Professional Technical Reference: Upper Saddle River, NJ, USA, 1990.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

https://doi.org/10.3390/s26041147

sensors

MDPI

Article

Real-Time Deep-Sea Mooring System with Inductive Telemetry and Multi-Sensor Integration: Deployment and Performance in the South China Sea

Tongmu Liu [1,2,3], Baocheng Zhou [1,2,*], Xinwen Zhang [1,2], Jianqing Peng [4,5], Hua Huang [1,2] and Tianhao Jian [1,2,3]

[1] South China Sea Marine Survey Center, Ministry of Natural Resources, Guangzhou 510300, China; liutongmu@126.com (T.L.); xinwenzh@foxmail.com (X.Z.); hhaostos@126.com (H.H.); 15737682122@163.com (T.J.)
[2] Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Natural Resources, Guangzhou 510300, China
[3] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 511458, China
[4] School of Intelligent Systems Engineering, Shenzhen Campus of Sun Yat-Sen University, Shenzhen 518107, China; pengjq7@mail.sysu.edu.cn
[5] Guangdong Provincial Key Laboratory of Fire Science and Technology, Guangzhou 510641, China
* Correspondence: 18312022312@163.com

Abstract

Obtaining real-time data from the deep ocean remains a major challenge in marine observation. This study presents a deep-sea mooring system that integrates inductive telemetry (EUM6000 modules) with Tiantong satellite communication to achieve real-time, long-term hydrological monitoring in the South China Sea. The system incorporated 25 sensors, including CTDs and ADCPs, and was deployed at a depth of 1247 m. Over one year of continuous operation, it maintained a data reception rate > 90%, with a latency of <15 min from seabed to shore. Compared to acoustic-based systems, the inductive telemetry design significantly improved energy efficiency and reliability. The high-resolution multi-sensor data provide valuable insights into ocean dynamics and support applications in climate research and disaster early warning. This system offers a robust solution for real-time deep-sea observation and serves as a reference for future ocean network development.

Keywords: real-time mooring; inductive telemetry; South China Sea; ADCP; CTD; Tiantong satellite

1. Introduction

The ocean is a vital component of the Earth system, harboring abundant resources and playing a crucial role in regulating the global climate. The deep ocean, constituting the bulk of the ocean, has long been at the forefront of research in physical oceanography, marine geology, and global climate change studies, with its internal hydrodynamic processes, material and energy exchanges, and ecological evolution serving as key focal points [1]. Acquiring long-term, continuous, high-resolution in situ observation data from the deep ocean holds irreplaceable scientific value for profoundly understanding ocean circulation structures, mesoscale eddy dynamics, internal ocean mixing processes, and accurately predicting the evolution trends of climate models [2–4].

Ocean observation plays a crucial role in understanding ocean processes, climate change prediction, environmental monitoring, and resource management. Long-term, continuous ocean data collection is vital for studying physical oceanography, biogeochemical

cycles, and ecosystem dynamics. Traditionally, ocean observation relies on mooring systems, including surface moorings [5] and subsurface mooring systems [6]. Surface mooring systems enable real-time data transmission via buoys and satellite communications but are susceptible to surface weather disturbances, waves, and human interference. Subsurface mooring systems, typically deployed in deep waters, avoid surface disruptions but can only retrieve data upon system recovery, lacking real-time capability. This limitation restricts their application in operational oceanography and real-time early warning systems [7,8].

To achieve real-time acquisition of deep-sea data, research institutions worldwide have developed multiple technical approaches, with the core challenge being the transmission of underwater data to shore-based centers [9,10]. Morrison et al. [11] developed the McLane Moored Profiler (MMP), which operates by utilizing a towing drive system to autonomously ascend and descend along an anchor mooring while simultaneously conducting measurements and observations. Send et al. [12] developed the SeaCycler autonomous ocean profiling system. This system employs an underwater winch to deploy and retrieve cables, controlling the ascent and descent of communication buoys. The buoys are raised to the surface only during data transmission. Similar approaches were proposed by Xu [13], Lin [14], Yu [15], and others [16,17]. Tian et al. [18] designed a low-cost, reusable underwater mooring system capable of periodic data uploads, primarily employing inductive coupling and timed release of communication buoys.

Zhang [19] developed a timed communication mooring that releases data-carrying communication buoys at predetermined intervals. After surfacing, the buoys transmit data via satellite. Wang et al. [20] developed a real-time communication mooring integrating acoustic and satellite communication. Its operating principle involves using acoustic communication technology to upload deep-sea data to surface buoys, which then relay the data back via satellite systems such as Iridium or BeiDou. Despite significant progress in real-time communication mooring technology, challenges remain in areas such as deep-water deployment, multi-sensor integration, and satellite communication capabilities.

With the rapid advancement of satellite communications, underwater sensing, and data transmission technologies, establishing real-time deep-sea observation systems has become a key development direction in international ocean observation [21]. Among these, inductive coupling transmission technology utilizes seawater as a medium and mooring cables as conductors to achieve non-contact penetration for electrical power and data transmission [22]. This provides a critical technical pathway for enabling long-term, reliable, high-speed data transfer between underwater observation platforms and surface floats [23]. Concurrently, the establishment of China's satellite mobile communication systems [24], such as Tiantong-1, provides stable and reliable long-distance communication support for transmitting deep-sea observation data back to shore-based receiving centers from remote locations [5]. Integrating inductive coupling underwater transmission with Tiantong satellite remote communication to establish an end-to-end real-time data link spanning "seabed-surface-shore stations" represents a highly promising innovative solution for overcoming the bottleneck of real-time deep-sea observation data transmission.

This paper presents the design, deployment, and validation of a deep-sea real-time transmission mooring system based on inductive coupling transmission technology and Tiantong satellite communication. This paper will detail the overall design, core technology integration plan, and implementation approaches for key technologies of this real-time communication hydrographic mooring system. It will focus on analyzing field test data from the South China Sea to validate the system's long-term stability, data reliability, and overall performance in real ocean environments. This aims to provide critical technical support and engineering practice references for constructing next-generation deep-sea real-time observation networks.

https://doi.org/10.3390/s26030937

2. System Design and Methodology

2.1. Overall System Design

The deep-sea hydrographic mooring system primarily consists of a communication buoy, main buoy body, inductively coupled cable, acoustic release mechanism, and composite anchor. Components are interconnected via inductively coupled steel cables, stainless steel anchor chains, Kevlar mooring ropes, inductively coupled anti-twist modules, shackles, and other connectors. The communication buoy is equipped with a Tiantong satellite real-time communication system.

Two 75K ADCPs are mounted on the main buoy, transmitting data via watertight cables to the inductive telemetry module compartment installed on the main buoy. Inductively coupled steel cables are attached above and below the main float, carrying inductively coupled sensors such as temperature-depth, temperature-salinity-depth, and current meters. Data collected by these sensors at various water layers is transmitted via inductively coupled steel cables and the inductively coupled data transmission system to the inductive telemetry module compartment on the main float. The inductive telemetry module compartment then transmits the integrated and packaged datasets in batches via inductively coupled cables to the data acquisition and control module of the satellite communication buoy.

The mooring system anchorage is designed as a three-tier structure with a communication buoy, sub-float, and main float. The system structure diagram is shown in Figure 1.

2.2. Sensor Integration and Configuration

The mooring system was designed to capture high-resolution, multi-parameter hydrological data throughout the entire water column. To achieve this, a suite of 25 sensors was strategically integrated onto the inductive coupling telemetry cable, forming a distributed observation array. The selection of sensors prioritized long-term stability, high accuracy, and seamless compatibility with the inductive data transmission system. The specific types, models, quantities, and key specifications of the deployed sensors are summarized in Table 1.

Table 1. Sensor inventory and key specifications deployed on the deep-sea mooring system.

Sensor Type	Model	Quantity	Key Specifications	Depth Rating
Inductive CTD	SBE37-SM	12	Accuracy: Temp.: ±0.002 °C; Cond.: ±0.003 mS/cm; Sal.: ±0.005 PSU	7000 m
Inductive T-D	OceanPlot TD-IM	8	Accuracy: Temp.: ±0.002 °C; Depth: ±0.05% FS	2000 m
Inductive Current Meter	ZPulseIM	3	Accuracy: ±1% ± 0.5 cm/s; Range: 0–5 m/s	6000 m
Acoustic Doppler Current Profiler (ADCP)	Teledyne Workhorse Long Ranger 75 kHz	2	Max Range: 600 m; Accuracy: ±1% ± 0.5 cm/s	1500 m

2.2.1. Stratified Hydrological Parameters

The core temperature and salinity data were provided by nineteen instruments: seven SBE37-SM MicroCAT CTD sensors (Sea-Bird Scientific, Bellevue, WA, USA) and twelve OceanPlot TD-IM inductive temperature-depth sensors (OceanPlot, Hangzhou, China). The SBE37-SM sensors were chosen for their proven high accuracy and long-term stability in measuring the fundamental parameters of conductivity, temperature, and pressure, from which salinity is derived. The TD-IM sensors provided complementary high-precision

https://doi.org/10.3390/s26030937

temperature and depth measurements, significantly increasing the spatial density of the thermal structure observation.

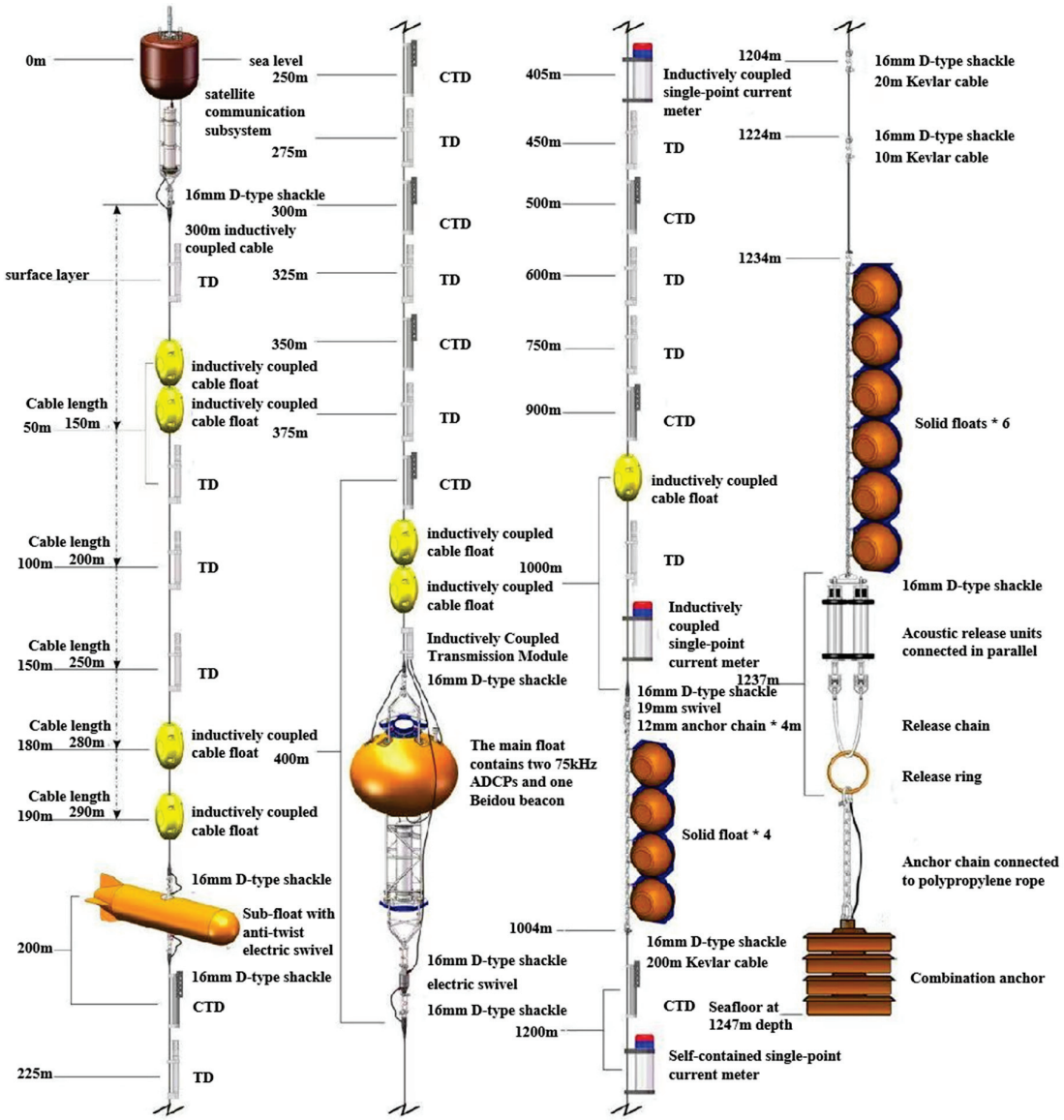


Figure 1. Overall design of coupled transmission mooring system.

These sensors were distributed at critical depths along the mooring cable, with a higher concentration within the upper ocean (0–400 m) to resolve the strong vertical gradients of the thermocline and halocline. This configuration allows for the detailed reconstruction of the water column's thermodynamic structure and the identification of fine-scale processes such as internal waves and intrusions.

2.2.2. Current Measurement Strategy

A dual-method approach was employed to measure ocean currents:

Point Measurements: Two ZPulseIM inductive single-point current meters (Aanderaa, Bergen, Norway) were installed at predetermined depths (405 m and 1000 m) to provide long-term, stable records of current velocity and direction at specific, strategic locations.

Profile Measurements: Two 75 kHz Teledyne Workhorse Long Ranger ADCPs (Teledyne RD Instruments, Poway, CA, USA) were mounted on the main subsurface buoy at approximately 400 m depth. This configuration was designed to maximize the coverage of the water column: one ADCP was oriented upwards to profile currents from the buoy

https://doi.org/10.3390/s26030937

towards the surface, and the other was oriented downwards to measure currents from the buoy towards the seabed. This setup provides a comprehensive view of the current shear and structure over most of the water column, which is essential for understanding dynamics like eddies and deep currents.

2.2.3. Data Integration via Inductive Telemetry

A key innovation of this system is the seamless integration of these diverse sensors via the EUM6000 inductive telemetry modules. Each sensor equipped with an IM interface communicates through the single, specially designed inductive coupling cable. This design eliminates the need for individual data cables for each sensor, drastically reducing the system's complexity, weight, and potential points of failure. The EUM6000 modules aggregate data from all sensors, packetize them, and relay the integrated datasets upwards to the surface communication buoy for satellite transmission. This integrated architecture ensures the synchronous acquisition and real-time transmission of all hydrological parameters, making the entire water column remotely accessible.

2.3. Inductive Telemetry Module

2.3.1. Working Principle of Inductive Coupling

The inductive telemetry technology serves as the critical data link between submerged sensors and the data acquisition unit in this mooring system. Its operation is based on the fundamental principle of electromagnetic induction, enabling robust and efficient data transmission through a standard, plastic-coated steel mooring cable without the need for physical electrical contacts [25].

As illustrated in Figure 2, the system primarily consists of two key components: an Inductively Coupled Master (Host) unit, typically integrated within the data logger, and multiple Inductively Coupled Slave units connected to individual sensors (e.g., pressure sensors) [6]. Each unit contains a ferrite core around which a coil is wound, forming a magnetic coupling circuit.

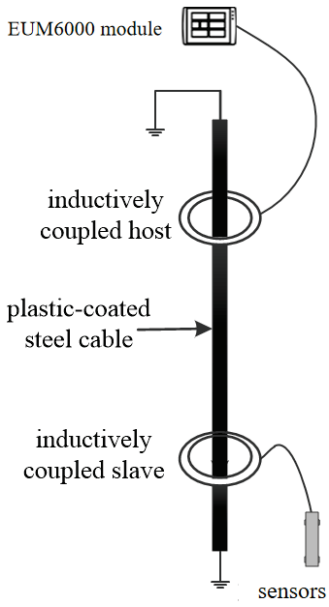


Figure 2. Schematic diagram of inductive coupling transmission.

The working principle can be described in three steps:

(1) Signal Modulation and Magnetic Field Generation: At the transmitting end (e.g., the slave unit connected to a pressure sensor), the digital data signal is first modulated.

https://doi.org/10.3390/s26030937

This modulated electrical signal is then fed into the coupling coil, generating an alternating magnetic field around the coil according to Ampère's circuital law. This magnetic field is confined and guided by the high magnetic permeability ferrite core.

(2) Magnetic Flux Coupling via the Cable: The key innovation lies in using the steel mooring cable itself as a conduit for the magnetic flux. The alternating magnetic field produced by the transmitting coil magnetically couples with the steel cable. The cable, acting as a low-reluctance path, channels this magnetic flux over long distances with minimal loss. This process effectively transforms the entire cable into a "virtual wire" for magnetic signals.

(3) Signal Reception and Demodulation: At the receiving end (e.g., the master unit), the alternating magnetic flux flowing through the cable induces an electrical current in the receiving coil, following Faraday's law of induction. This induced electrical signal is then demodulated to recover the original digital data transmitted by the sensor.

This non-contact, magnetic induction-based method offers significant advantages over other underwater communication techniques:

High Reliability: Unlike acoustic telemetry, which is susceptible to noise, multipath interference, and absorption losses in water, inductive coupling provides a stable and predictable channel because magnetic fields are less affected by the aquatic environment.

Low Power Consumption: The power required for short-range inductive coupling is substantially lower than that for acoustic modems, which is crucial for long-term deployments.

Simplified Integration: A single cable serves both the mechanical mooring function and the data/power transmission bus, drastically reducing system complexity compared to using multiple dedicated electrical cables for each sensor.

In this system, the EUM6000 modules implement this principle, ensuring reliable, real-time transmission of data from all distributed sensors to the central data logger for subsequent processing and satellite transmission.

2.3.2. EUM6000 Module Specifications and Performance

The EUM6000 inductive telemetry module is the core component that implements the working principle described above, serving as the data gateway for the entire sensor array. Its specifications are meticulously engineered to meet the demands of long-term, deep-sea observation, and its performance is a critical determinant of the system's overall reliability. The key specifications of the EUM6000 module are summarized in Table 2.

Table 2. Key specifications of the EUM6000 inductive telemetry module.

Parameter	Specification	Note/Implication
Operating Depth	6000 m	Compatible with abyssal seafloor deployments.
Data Transmission Rate	4800 bps	Adequate for high-frequency data from CTD, ADCP, and pressure sensors.
Maximum Range	6000 m	Supports data transmission over the full length of deep-sea moorings.
Power Consumption	<1 W (active)	Essential for long-term ($\geq$1 year) battery-powered operation.
Communication Interface	RS-232/Inductive Coupling	Enables seamless connection to both standard sensors and the inductive cable.
Number of Slave Units Supported	Up to 32	Allows for dense multi-sensor integration on a single mooring line.

The performance of the EUM6000 module was rigorously validated during the system's development and sea trials, demonstrating several key advantages:

https://doi.org/10.3390/s26030937

(1) High Data Fidelity: The module employs robust error-checking protocols to ensure data integrity. The use of electromagnetic induction avoids the high bit error rates typical of underwater acoustic channels, which was a primary factor in achieving the overall system data reception rate of >90% during the South China Sea deployment.
(2) Low-Latency Communication: Data transmission through the EUM6000 system is nearly instantaneous, with latency determined primarily by the signal propagation speed (approaching the speed of light in the cable medium). This is in stark contrast to acoustic telemetry, which suffers from slow propagation speeds (~1500 m/s), resulting in significant delays over long distances. This low latency is crucial for creating a synchronized, real-time profile of the water column.
(3) System Simplicity and Reliability: A significant performance metric is the mean time between failures (MTBF). The solid-state design of the EUM6000, with no moving parts and minimal underwater connectors, contributes to a high MTBF. This design eliminates the primary points of failure associated with traditional multi-core watertight cables, which are prone to leakage and damage under deep-sea pressure and dynamic loading.

In conclusion, the EUM6000 module is not merely a communication component but an enabler of the system's core innovation. Its specifications for depth, data rate, and low power consumption, combined with its demonstrated performance in reliability and latency, directly support the successful integration of multiple sensors and the establishment of a robust real-time data link from the deep sea to the shore station.

2.4. Tiantong Satellite Communication System

The Tiantong-1 satellite system is a cornerstone of China's independent space-based infrastructure, operating in geosynchronous Earth orbit (GEO) to provide secure mobile communication services over its coverage area, which encompasses the South China Sea. For deep-sea mooring systems requiring real-time data telemetry, the selection of a communication system must balance data rate, reliability, power consumption, and national security considerations. While systems like Iridium and Inmarsat offer global coverage, the Tiantong system was selected for this study due to its superior data rate compared to the BeiDou Short Message Service (SMS), lower operational cost, and critically, its status as a sovereign, secure communication asset.

2.4.1. System Architecture and Link Budget

The communication chain follows a standardized architecture: Mooring System $\rightarrow$ Tiantong User Terminal (on surface buoy) $\rightarrow$ Tiantong GEO Satellite $\rightarrow$ Ground Gateway Station $\rightarrow$ Shore-based Data Center. The user link operates in the S-band, which offers a favorable balance between antenna size, atmospheric attenuation, and beam width, making it suitable for small marine platforms. The terminal employed in this system features a directional antenna with a servo-stabilized gimbal mechanism. This mechanism is imperative for maintaining a high-quality link under dynamic sea states. An integrated Inertial Measurement Unit (IMU) provides real-time attitude data (pitch, roll) to an Application Processor (AP), which calculates compensation angles to ensure precise antenna pointing towards the satellite, achieving a tracking accuracy of better than 2° even with buoy tilts of $\pm 30^\circ$.

2.4.2. Communication Protocol Stack and Data Handling

The system implements a structured protocol stack to manage the complexity of satellite communication. Interaction with the Tiantong modem is governed by a set of AT commands sent via a UART interface. Key commands used in our implementation include

https://doi.org/10.3390/s26030937

AT + CSQ (to query signal strength),AT + CGREG = 1 (to register with the packet-switched network), and ATD*98*4# (to initiate a PPP dial-up connection) [5].

To efficiently utilize the available bandwidth and ensure data integrity, a custom application-layer protocol was designed, as illustrated in Figure 3.

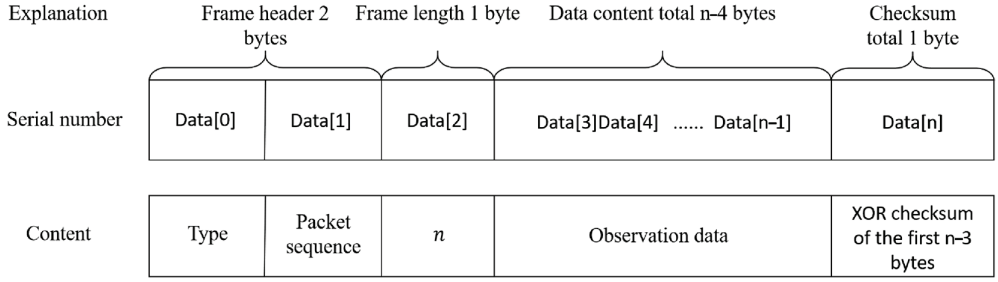


Figure 3. Encapsulated data frame structure.

Framing and Encapsulation: Large datasets, particularly from the ADCP, exceed the typical maximum transmission unit (MTU) of a single satellite packet. Therefore, data is fragmented into frames. Each frame consists of a Header, Payload, and Trailer.

Header: Contains a Start Delimiter, a Data Type identifier and a Frame Sequence Number.

Payload: Contains the actual sensor data.

Trailer: Contains a Cyclic Redundancy Check (CRC) or an XOR checksum computed over the entire frame to detect transmission errors.

Quality of Service (QoS) and Retransmission Mechanism: A critical aspect of the protocol is its QoS strategy to handle packet loss. The shore station monitors the sequence numbers of incoming frames. Upon detecting a gap, it issues a NACK (Negative Acknowledgement) command via the Tiantong return link, specifying the missing frame number. The buoy-side controller stores recent data packets in a buffer and retransmits the requested frame. This process is repeated for a predefined maximum number of attempts to ensure robustness without depleting power resources.

2.4.3. Performance Metrics

In this deployment, the Tiantong terminal operated with an effective data rate of 9.6 kbps. The average latency for a data packet traveling from the buoy to the shore station was measured to be less than 15 min, which includes processing, satellite hop, and terrestrial network routing delays. The system's link budget, designed with a sufficient margin, coupled with the automatic beam-tracking, resulted in a measured data delivery success rate of >90% during the sea trial, validating the effectiveness of the chosen communication architecture and protocols for real-time ocean observation.

2.5. Power System and Energy Consumption Analysis

To ensure the system can achieve long-term, continuous observation exceeding one year, we conducted a detailed analysis and design of the power requirements for the entire observation chain. The system adopts a distributed power supply strategy: the inductively coupled transmission link and satellite communication module are powered by independent battery packs, while the sampling power consumption of certain sensors (such as CTDs) is handled by their built-in batteries. This approach optimizes system redundancy and reliability.

Power consumption analysis for major components is as follows: The 12 TD-IM temperature-salinity-depth loggers and 7 CTD-IM inductively coupled modules, due to their low power characteristics, are projected to operate for over 28,000 h each. The two

https://doi.org/10.3390/s26030937

inductively coupled single-point current meters and the inductively coupled transmission module for the ADCP, operating on a 30 min cycle, are projected to last 60,000 h and 35,000 h, respectively.

The satellite communication module represents the core power consumption component of the entire system. During its one-hour operational cycle, the total power consumption for completing one inductively coupled data acquisition and satellite data transmission cycle is approximately 18 mAh. Based on this calculation, the equipped 238 Ah high-capacity battery can support continuous operation of this module for over 10,000 h (approximately 416 days), fully meeting the design life requirements. Considering the combined power consumption of all equipment and the distributed power supply solution, the entire observation system possesses the energy assurance to operate continuously and stably for over one year while executing its designated observation tasks.

3. Field Deployment and Validation

3.1. Pre-Deployment Preparation and Laboratory Testing

Prior to the sea deployment, all sensors, including the CTDs and TD sensors, were calibrated against a shipboard SBE 911 + CTD system. The results showed a root mean square deviation (RMSD) of 0.003 PSU for salinity and 0.002 °C for temperature, ensuring data accuracy. The entire mooring system was assembled and tested dockside to verify the integrity of mechanical connections, the functionality of the acoustic release mechanism, and the end-to-end data flow from seabed sensors to the satellite transmission. A 72 h continuous test confirmed the stability of the inductive telemetry link and the power management system, with the measured average power consumption projecting a battery life exceeding 400 days.

3.2. Sea Trial Deployment: Procedure and Site Description

The mooring system was deployed on 22 June 2024. The deployment site was selected at a depth of 1247 m in the northern South China Sea, a region known for its strong internal solitary wave activities, providing an ideal environment for validating the system's capability to capture dynamic processes. The deployment procedure followed the standard sequence: first, the surface communication buoy was lowered, followed by the controlled deployment of the mooring line with sensors, the main float, release and finally, the anchor weight. The field operation diagram for deploying the moored buoy is shown in Figure 4.



(**a**) (**b**)

Figure 4. Deployment Site Diagram for Submerged Buoys. (**a**) Field Operation Diagram; (**b**) Tiantong Communications Buoy.

3.3. Data Reception and System Operational Status

The system commenced data transmission immediately after deployment. Over the initial 42-day observation period, the shore station achieved a data reception rate of 90.45%.

https://doi.org/10.3390/s26030937

As shown in Figure 5, the depth data from the communication buoy in July is analyzed. The buoy remained above sea level most of the time, but occasionally dipped into the water, which compromised its communication effectiveness. Analysis of buoy communication data reveals a significant correlation between the frequency of communication interruptions and the astronomical spring tide cycle, with failures primarily occurring around the periods of new moon and full moon spring tides. The fundamental cause of this phenomenon likely lies in the active internal solitary waves in the South China Sea. During the passage of internal solitary waves, extremely strong sudden currents are generated at the sea surface, which will pull the buoy into the sea. In the future, we will optimize the anchor system design and consider adopting longer mooring line or bigger buoyancy to enhance its stability under strong current impacts. Simultaneously, we will strengthen communication system redundancy to enable automatic data re-transmission upon recovery, fundamentally improving system reliability and data acquisition rates.

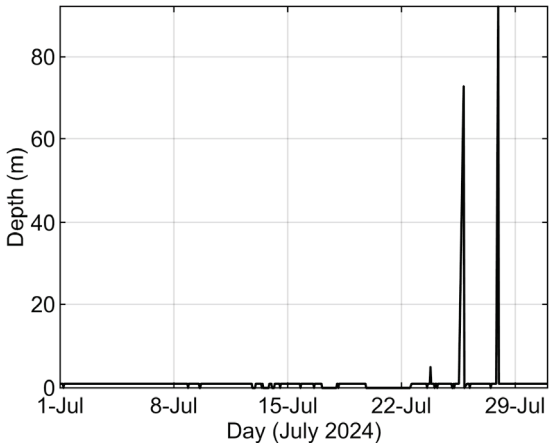


Figure 5. The Depth of the Communication Buoy.

3.4. Analysis of Hydrological Data Products

The collected data successfully captured the detailed hydrological structure of the water column. The temperature-salinity profiles from the 19 sensors revealed a pronounced thermocline and the characteristic salinity minimum layer, consistent with historical observations in the region. The simultaneous measurements from the two ADCPs provided a continuous, high-resolution view of the current velocity profile, clearly showing the strengthening of the subsurface current maximum and its vertical shear during the study period.

3.4.1. Temperature-Depth Profile

As shown in Figure 6, the real-time moored buoy system constructed in this study successfully acquired high-resolution vertical temperature profiles in the northern South China Sea. The observational data clearly reveal the typical marine thermal stratification structure in this sea area. The upper mixed layer extends from the sea surface to a depth of approximately 50–60 m, where water temperature uniformly decreases from about 22 °C to 20 °C with a relatively small vertical gradient. This indicates that the water in this layer is thoroughly mixed under the influence of dynamic processes such as wind and wave mixing, forming relatively homogeneous surface water. The main thermocline extends from approximately 60 to 800 m depth, where temperature decreases sharply with depth—plunging abruptly from 20 °C to about 5 °C. This layer exhibits the most dramatic vertical temperature variation. Its strong density gradient forms a barrier to vertical exchange, exerting a decisive influence on the transport of energy and matter within the ocean.

https://doi.org/10.3390/s26030937

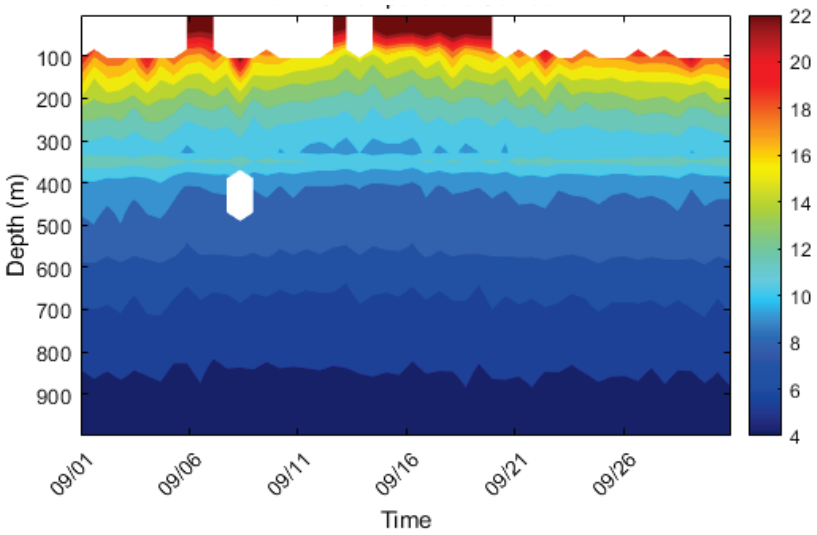


Figure 6. Typical Temperature-Depth Profile (the white section denotes data of poor quality).

3.4.2. Salinity Profile

As shown in Figure 7, the system's salinity observations reveal the complex vertical structure and temporal variability within the study area. The surface layer, characterized by low salinity water, exhibits relatively low salinity (approximately 33.6–34.2) at depths shallower than 50 m and demonstrates high temporal variability. This low-salinity feature likely originates from summer precipitation or freshwater input from the surrounding continental shelf. A pronounced hypersaline core layer exists between approximately 100 and 300 m depth, characterized by salinities exceeding 34.8. Maximum salinities within this core layer can surpass 35.0. This represents a signature feature of North Pacific Tropical Water (NPTW) intrusion into the South China Sea. This hypersaline mass enters through conduits such as the Luzon Strait, subsequently sinking and spreading throughout the South China Sea interior. Deep-water salinity distribution: Below the hypersaline core layer, salinity gradually decreases with depth. At depths greater than approximately 800 m, salinity stabilizes around 34.4–34.5, consistent with the characteristics of deep-water in the South China Sea.

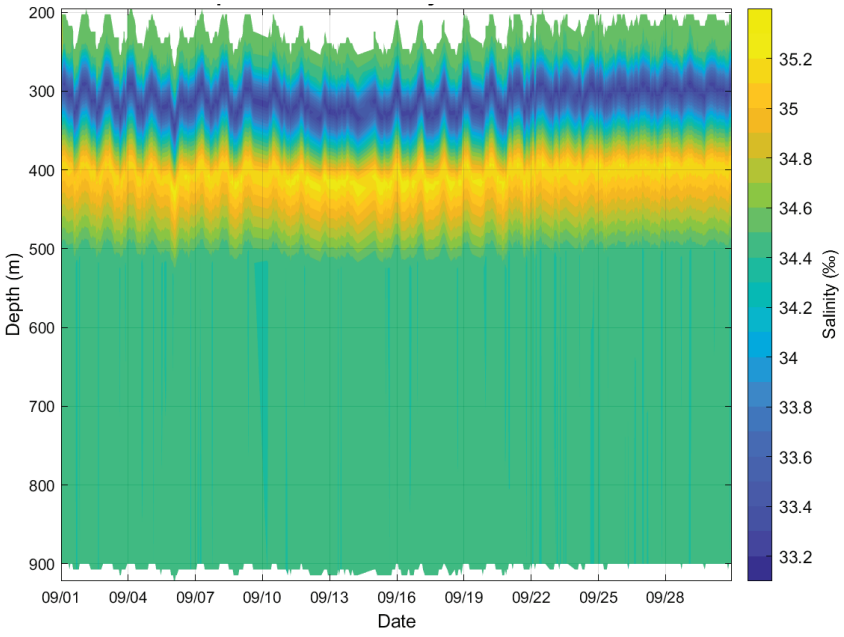


Figure 7. Typical Salinity-Depth Profile.

https://doi.org/10.3390/s26030937

3.4.3. Temperature-Salinity Diagram

The T-S diagram (Figure 8) provides a integrated view of the water mass structure in the northern South China Sea. The data points, colored by depth, form a tight, coherent curve that aligns well with the known T-S characteristics of this region [26]. The temperature and salinity distribution ranges of this water mass are 4.95–13.39 °C and 34.39–34.55‰, respectively. This water mass is the least altered oceanic-type water mass in the South China Sea region, retaining the fundamental characteristics of mid-depth water masses in the western Pacific. The stability and clarity of this T-S relationship, derived from our multi-sensor array, validates the high quality and internal consistency of the data delivered by the inductive telemetry system.

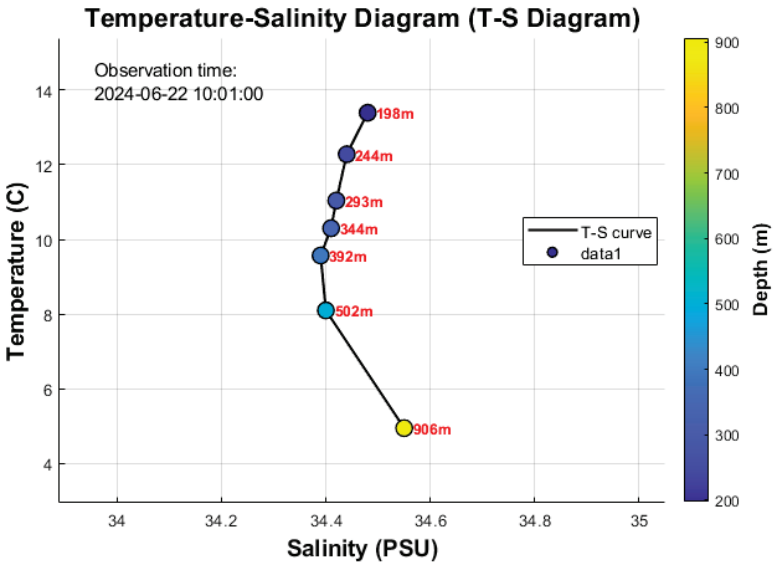


Figure 8. Temperature-Salinity (T-S) diagram.

3.4.4. Current Profile

Figure 9 presents high-resolution ocean current profile data acquired during the passage of Typhoon Yagi in 1–10 September 2024, analyzing the dynamic response of the upper ocean to the intense typhoon. Observations reveal a strong surface current pulse of up to 0.6 m/s in the northward component during the typhoon's core passage, rapidly decaying with depth. Figure 10 shows the distribution of near-inertial energy during the typhoon. It can be observed that energy is primarily confined within the upper 200 m, exhibiting distinct near-inertial oscillatory characteristics. The signal period is approximately 5.6 h, nearly identical to the local inertial period, indicating that near-inertial motion dominates the flow field. This exhibits typical wind-driven flow structure and near-inertial oscillation characteristics. Concurrently, the eastward component velocity shows frequent directional shifts correlated with the typhoon's rotating wind field. The overall flow field was dominated by a northward component, indicating the observation point was located on the typhoon's right-side path—the region receiving the strongest energy input. This study visually reveals how Typhoon Yagi injected immense energy into the upper ocean and drove intense currents, providing a key case study for understanding typhoon-ocean interactions.

https://doi.org/10.3390/s26030937

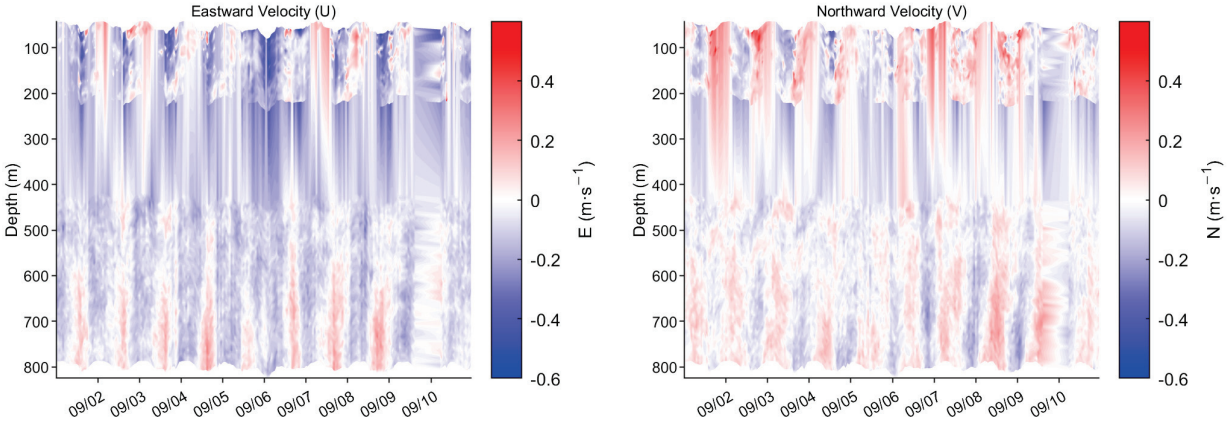


Figure 9. Current profile observed during the passage of Typhoon Yagi.

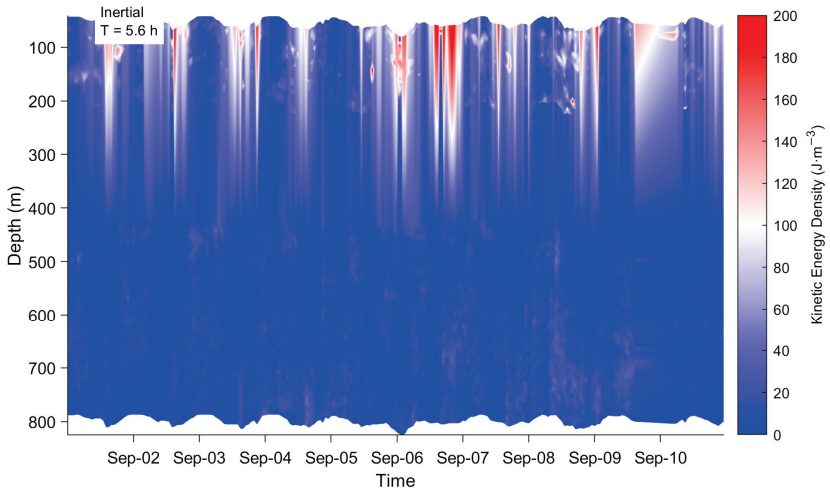


Figure 10. The kinetic energy during the passage of Typhoon Yagi.

4. Discussion

The successful deployment and operation of the real-time mooring system in the South China Sea demonstrate the efficacy of the inductive telemetry + Tiantong satellite solution for deep-sea observation. The core achievement of this study lies in establishing a highly reliable, continuous, and near-real-time data link from the seabed to the shore station, a feat that has been a significant challenge for conventional approaches.

The performance of our system stands in stark contrast to existing real-time telemetry schemes. Firstly, compared to systems relying on acoustic modems, e.g., ref. [20], our inductive coupling method eliminates the inherent vulnerabilities of the underwater acoustic channel, such as signal attenuation, multi-path interference, and bio-fouling noise. This is evidenced by the sustained data reception rate of >90% over the deployment period. Acoustic systems, while capable of greater depths, often suffer from unpredictable data packet loss and require high-power consumption for transmission, compromising both data integrity and operational longevity. Our approach provides a stable, low-power, and high-fidelity connection akin to a "virtual cable," ensuring the complete transmission of high-resolution datasets from multiple sensors.

Secondly, when compared to mobile profiling platforms (e.g., MMPs or SeaCyclers [11]), our fixed-depth, multi-sensor array provides a true synoptic view of the entire water column at a single moment. Profilers offer high vertical resolution but at the cost of poor temporal resolution for any given depth, potentially aliasing rapid dynamic processes like internal solitary waves or frontal movements. Furthermore, the mechanical complexity of profilers, with their moving parts and winching systems, presents a non-trivial risk of

https://doi.org/10.3390/s26030937

failure during long-term deployments. Our stationary design boasts greater mechanical simplicity and reliability, which is paramount for year-long unattended operations.

Lastly, against pop-up buoy systems, our solution offers genuine real-time capability. While pop-up buoys excel in covert, long-term data storage and are low-risk during their submerged phase, they provide only retrospective data delivery after a significant delay (weeks or months). This makes them unsuitable for time-critical applications such as ocean forecasting, typhoon monitoring, or rapid response to oceanic events. Our system's <15 min latency is a transformative feature for operational oceanography and dynamic process studies.

The capability of our system to deliver full-water-column, real-time thermohaline and current profiles with high vertical resolution opens up new possibilities for studying submesoscale processes and validating high-resolution numerical models. For instance, the detailed structure of the thermocline and halocline captured by our 19 sensors provides a ground truth for model simulations of stratification. Furthermore, the ability to resolve near-inertial oscillations and their vertical propagation, as hinted at in the current response to Typhoon Yagi, is crucial for understanding the pathway of wind energy into the ocean interior and its role in driving mixing.

Nevertheless, the system's surface expression remains a vulnerability, exposing it to potential damage from fishing activities, vessel traffic, and extreme sea states. To enhance resilience against strong internal waves and extreme surface conditions, we plan to increase buoyancy and extend mooring length in future deployments, which will help maintain antenna elevation and alignment. Furthermore, to mitigate data loss during communication interruptions, the system will be enhanced with robust data buffering and automatic retransmission capabilities, ensuring data integrity even during prolonged link outages. Future iterations could also explore hybrid communication strategies, such as combining inductive telemetry with a submerged, satellite-ready buoy that only surfaces briefly for data transmission, further mitigating surface-related risks.

5. Conclusions

This paper develops a deep-sea real-time communication moored buoy system, providing a reliable platform for long-term real-time observation in the deep ocean. Based on inductive telemetry technology (EUM6000 module) and Tiantong satellite communication, this deep-sea real-time transmission moored buoy system integrates 25 sensors including CTD and ADCP. It achieves continuous observation at a water depth of 1247 m for $\geq$1 year with a data reception rate >90%. Multi-sensor data fusion enhances profiling resolution for layered observations at 50–1000 m. This mooring system holds promising applications in ocean dynamics research, climate model validation, and disaster early warning. This system demonstrates a viable and robust pathway for real-time data transmission from the deep sea, serving not only as a standalone observatory but also as a potential prototype for future nodes in a comprehensive, integrated ocean observation network. Future studies may explore the impact of biofouling on buoyancy, long-term reliability testing of sealing components under deep-sea high-pressure conditions, and system safety.

Author Contributions: Conceptualization, B.Z. and T.L.; methodology, T.L.; software, X.Z.; validation, T.L., J.P. and T.J.; formal analysis, H.H.; investigation, T.L.; resources, B.Z.; data curation, B.Z.; writing—original draft preparation, T.L.; writing—review and editing, X.Z.; visualization, J.P.; supervision, H.H.; project administration, B.Z.; funding acquisition, B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key Research and Development Program of China (Grant Nos. 2022YFC3104204 and 2023YFC3107502) and the Hainan Province Science and

https://doi.org/10.3390/s26030937

Technology Special Fund (Grant No. ZDYF2023GXJS151). The article processing charge (APC) was funded by Grant No. 2022YFC3104204.

Data Availability Statement: Data is available upon reasonable request to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADCP	Acoustic Doppler Current Profiler
TD	Temperature–Depth sensor
TD-IM	Temperature–Depth sensor with an Inductive Modem
CTD	Conductivity–Temperature–Depth sensor
CTD-IM	Conductivity–Temperature–Depth sensor with an Inductive Modem

References

1. Zhao, W.; Zhou, C.; Zhang, Z.; Huang, X.; Guan, S.; Yang, Q.; Sun, Z.; Qin, C.; Guan, Y.; Tian, J. The South China Sea Mooring Array and Its Applications in Exploring Oceanic Multiscale Dynamics. *Sci. Bull.* **2025**, *70*, 609–623. [CrossRef] [PubMed]
2. Li, R.; Chen, C.; Dong, W.; Beardsley, R.C.; Wu, Z.; Gong, W.; Liu, Y.; Liu, T.; Xu, D. Slope-Intensified Storm-Induced Near-Inertial Oscillations in the South China Sea. *J. Geophys. Res. Oceans* **2021**, *126*, e2020JC016713. [CrossRef]
3. Yang, Q.; Zhao, W.; Li, M.; Tian, J. Spatial Structure of Turbulent Mixing in the Northwestern Pacific Ocean. *J. Phys. Oceanogr.* **2014**, *44*, 2235–2247. [CrossRef]
4. Liu, T.; Yu, J.; Yu, Y.; Zhang, X.; Zhou, B.; Yin, L. Characterization of Wave Power Resources off the Coast of Guangdong. *Processes* **2023**, *11*, 2221. [CrossRef]
5. Zhang, X.; Lin, G.; Li, R.; Yang, W.; Liu, T.; Zhou, B.; Yin, L.; Ding, Y. Design and Application of Internal Solitary Wave Monitoring System Based on the Tiantong Communication. *J. Trop. Oceanogr.* **2024**, *43*, 180–189. [CrossRef]
6. Zhou, B.; Zhang, X.; Wan, X.; Liu, T.; Liu, Y.; Huang, H.; Chen, J. Development and Application of a Novel Tsunami Monitoring System Based on Submerged Mooring. *Sensors* **2024**, *24*, 6048. [CrossRef]
7. Trask, R.P.; Farrar, J.T. Near Real-Time Data Recovery from Oceanographic Moorings. In *Observing the Oceans in Real Time*; Venkatesan, R., Ed.; Springer: Cham, Switzerland, 2018; pp. 263–279. [CrossRef]
8. Zhang, S.; Tian, C.; Zhou, F. Ocean Observation System Design of Mooring Buoy and Benthic Node with Electro-Optical-Mechanical Cable. *Front. Mar. Sci.* **2022**, *9*, 1018751. [CrossRef]
9. Xiao, D.; Zhu, Q.; Zhang, J.; Xie, T.; Ji, Q. The Design and Data Analysis of an Underwater Seismic Wave System. *Sensors* **2025**, *25*, 4155. [CrossRef]
10. Bian, J.W.; Huang, C.J. Real-Time and Long-Term Monitoring of Coastal Water Turbidity Using an Ocean Buoy Equipped with an ADCP. *Sensors* **2024**, *24*, 6979. [CrossRef] [PubMed]
11. Morrison, A.T.; Billings, J.D.; Doherty, K.W. The McLane Moored Profiler: An Autonomous Platform for Oceanographic Measurements. In Proceedings of the OCEANS 2000 MTS/IEEE Conference and Exhibition, Providence, RI, USA, 11–14 September 2000; pp. 353–358. [CrossRef]
12. Send, U.; Fowler, G.; Siddall, G.; Beanlands, B.; Pittman, M.; Waldmann, C.; Karstensen, J.; Lampitt, R. SeaCycler: A Moored Open-Ocean Profiling System for the Upper Ocean in Extended Self-Contained Deployments. *J. Atmos. Ocean. Technol.* **2013**, *30*, 1555–1565. [CrossRef]
13. Xu, M.; Li, Y.; Xu, Q. Sea Trial of a Real-Time Submerged Buoy Communication System Based on Wireless Networking. In Proceedings of the AIPCC '19, Sanya, China, 19–21 December 2019; Association for Computing Machinery: New York, NY, USA, 2019.
14. Lin, B.; Wang, D.; Dong, Y.; Ren, C. Design and Sea Trial of Real-Time Communication Subsurface Mooring System Based on Underwater Winch. *J. Ocean. Technol.* **2024**, *43*, 27–35.
15. Yu, F.; Chen, Y.; Zhou, C.; Zhang, X.; Deng, K.; Han, Y.; Liu, Y.; Liu, Q.; Wang, B.; Hu, H.; et al. Design of Deep-Sea Quasi-Real-Time-Communication Submerged Buoy Systems. *Mar. Sci.* **2020**, *44*, 194–200.
16. Li, W.; Zhang, X.; Li, J.; Li, M.; Zhang, W. Design of the Surface Communication Buoy of the Real-Time Sub-Surface Mooring System. *J. Ocean. Technol.* **2014**, *33*, 47–51.
17. Yu, R.; Li, H.; Feng, S.; Yin, H.; Ren, S. Development, Application and Prospect of Submarine Buoy Technology. *Meteorol. Hydrol. Mar. Instrum.* **2022**, *39*, 112–116.

https://doi.org/10.3390/s26030937

18. Tian, C.; Deng, Z.D.; Xu, X.; Yang, Q.; Zhao, W. Timed Communication Buoy System: A Subsurface Mooring System for Efficient Sensor Data Recovery. *J. Mar. Sci. Eng.* **2015**, *49*, 117–126. [CrossRef]
19. Zhang, Y. Overview on Development of Marine Environment Monitoring Equipment and Technologies. *Mine Warf. Ship Self-Def.* **2018**, *1*, 7–14.
20. Wang, F.; Wang, J.; Xu, L.; Zhang, X.; Yan, S.; Chen, Y. The Development of a New Real-Time Subsurface Mooring. *J. Oceanol. Limnol.* **2020**, *38*, 1080–1091. [CrossRef]
21. Howe, B.M.; Chao, Y.; Arabshahi, P.; Roy, S.; McGinnis, T.; Gray, A. A Smart Sensor Web for Ocean Observation: Fixed and Mobile Platforms, Integrated Acoustics, Satellites and Predictive Modeling. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 507–521. [CrossRef]
22. Bhagat, S.N.; Malarkodi, A.; Raguraman, G.; Latha, G. Inductive Telemetry for Passive Acoustic Monitoring in the Ocean. In Proceedings of the 2023 International Symposium on Ocean Technology (SYMPOL), Kochi, India, 6–9 December 2023; pp. 1–5. [CrossRef]
23. Luo, P.; Song, Y.; Xu, X.; Wang, C.; Zhang, S.; Shu, Y.; Ma, Y.; Shen, C.; Tian, C. Efficient Underwater Sensor Data Recovery Method for Real-Time Communication Subsurface Mooring System. *J. Mar. Sci. Eng.* **2022**, *10*, 1491. [CrossRef]
24. Liu, Y.; Xu, Y.; Zhang, K.; Feng, Y. Hardware Design of Marine Buoy Data Communication Terminal Based on the 1st Generation Beidou Navigation Positioning System. *J. Coast. Res.* **2019**, *93*, 639–645. [CrossRef]
25. Xu, J.; Li, X.; Xie, Z.; Zhang, H.; Wu, T.; Fang, C. Research on a Multiple-Receiver Inductively Coupled Power Transfer System for Mooring Buoy Applications. *Energies* **2017**, *10*, 519. [CrossRef]
26. Cheng, G.; Sun, J.; Zu, T.; Chen, J.; Wang, D. Analysis of Water Masses in the Northern South China Sea in Summer 2011. *J. Trop. Oceanogr.* **2014**, *33*, 10–16. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

https://doi.org/10.3390/s26030937

sensors

MDPI

Article

Development of a Robot-Assisted TMS Localization System Using Dual Capacitive Sensors for Coil Tilt Detection

Czaryn Diane Salazar Ompico [1], Julius Noel Banayo [2], Yamato Mashio [1], Masato Odagaki [1,*], Yutaka Kikuchi [3], Armyn Chang Sy [2,4] and Hirofumi Kurosaki [5]

[1] Systems and Bioengineering Department, Faculty of Engineering, Maebashi Institute of Technology, Maebashi 371-0816, Gunma, Japan; czarynompico@gmail.com (C.D.S.O.); yamax.gmax@gmail.com (Y.M.)
[2] Evelyn D. Ang-Institute of Biomedical Engineering and Health Technologies, De La Salle University, Manila 1004, Philippines; julius.banayo@dlsu.edu.ph (J.N.B.); armyn.sy@dlsu.edu.ph (A.C.S.)
[3] Institute of Brain and Blood Vessels, Mihara Memorial Hospital, Isesaki 372-0006, Gunma, Japan; viconmx@gmail.com
[4] Manufacturing Engineering and Management Department, De La Salle University, Manila 1004, Philippines
[5] Gunma Industrial Technology Center, Maebashi 379-2147, Gunma, Japan; hirofumi.kurosaki@gmail.com
* Correspondence: odagaki@maebashi-it.ac.jp

Highlights

What are the main findings?

- A robotic-assisted TMS system using a 3D camera to detect facial landmarks can successfully locate the C3 motor hotspot with minimal calibration.
- Sensor ratio balance and moderate contact pressure result in higher MEP amplitudes, indicating effective coil–scalp alignment and tilt detection.

What are the implications of the main findings?

- Low-cost textile capacitive sensors are able to provide real-time feedback on coil positioning and tilt, enabling safer and more reproducible TMS procedures.
- Robotic assistance has the potential to reduce setup time and operator dependency compared to traditional manual localization, improving efficiency and consistency in clinical and research settings.

Abstract

Transcranial Magnetic Stimulation (TMS) is a non-invasive technique for neurological research and therapy, but its effectiveness depends on accurate and stable coil placement. Manual localization based on anatomical landmarks is time-consuming and operator-dependent, while state-of-the-art robotic and neuronavigation systems achieve high accuracy using optical tracking with head-mounted markers and infrared cameras, at the cost of increased system complexity and setup burden. This study presents a cost-effective, markerless robotic-assisted TMS system that combines a 3D depth camera and textile capacitive sensors to assist coil localization and contact control. Facial landmarks detected by the depth camera are used to estimate the motor cortex (C3) location without external tracking markers, while a dual textile-sensor suspension provides compliant "soft-landing" behavior, contact confirmation, and coil-tilt estimation. Experimental evaluation with five participants showed reliable C3 targeting with valid motor evoked potentials (MEPs) obtained in most trials after initial calibration, and tilt-verification experiments revealed that peak MEP amplitudes occurred near balanced sensor readings in 12 of 15 trials (80%). The system employs a collaborative robot designed in accordance with international human–robot interaction safety standards, including force-limited actuation and monitored stopping. These results suggest that the proposed approach can improve the

accessibility, safety, and consistency of TMS procedures while avoiding the complexity of conventional optical tracking systems.

Keywords: Transcranial Magnetic Stimulation; robotic-assisted TMS; coil localization; 3D camera; capacitive sensors; tilt detection

1. Introduction

Transcranial Magnetic Stimulation (TMS) is a widely used non-invasive brain stimulation technique for both research and clinical applications. TMS induces localized electric currents in cortical regions by generating rapidly changing magnetic fields, thereby modulating neuronal activity. It is commonly used for treating major depressive disorder (MDD), particularly for patients resistant to pharmacological therapy, and also clinically accepted in treating other neurological and psychiatric disorders, including postpartum depression (PPD) and substance use disorders (SUDs) [1]. High-frequency stimulation generally enhances cortical excitability, while low-frequency stimulation suppresses overactive areas. Several studies have reported a decrease in depressive symptoms following stimulation of regions such as the dorsolateral and medial prefrontal cortex [2,3].

Despite these advances, TMS efficiency and reproducibility have certain limitations. Even subtle deviations can reduce stimulation efficacy. Hence, accurate and stable coil placement is critical. Manual positioning can lead to operator fatigue, small drifts during long sessions, and inter-session variability, all of which compromise treatment consistency [4–6].

To address these challenges, neuronavigation and robotic systems have been developed. Advanced neuronavigation solutions often integrate MRI or fMRI data to guide coil placement with millimeter-level precision [7]. Recent robotic systems developed for TMS incorporate real-time force feedback and imaging. Examples include robots with multi-axis force sensors that regulate contact pressure [8], and spherical robots with neuronavigation that use visual tracking to maintain precise alignment [9]. However, these systems are often costly, requiring equipment such as laser trackers and head markers that can compromise patient comfort. This study does not aim to achieve MRI-level anatomical targeting of specific cortical sulci or gyri. It focuses on improving the repeatability, accessibility, and functional consistency of scalp-based TMS coil positioning using a markerless robotic approach.

The present study proposes a low-cost, markerless robotic-assisted TMS system that combines a 3D depth camera with textile capacitive sensors. The system automatically detects facial landmarks to localize the motor cortex (C3) and employs a soft, compliant suspension structure with dual textile sensors to ensure even coil contact and detect tilt. Unlike recent similar systems that depend on force sensors or external trackers, this method emphasizes affordability, accessibility, patient comfort, and ease of use. This work describes the design, implementation, and preliminary evaluation of the system, with the aim of enhancing coil placement accuracy and reducing operator workload in future TMS applications.

2. Materials and Methods

2.1. System Overview

The proposed system integrates an igus ReBeL 5-DOF collaborative robot, Intel RealSense D435 depth camera, and two custom textile capacitive sensors placed on a suspension platform, which can be seen in Figure 1. The robot holds the TMS coil and adjusts orientation based on camera-derived anatomical landmarks and sensor feedback. The igus

https://doi.org/10.3390/s26020693

ReBeL collaborative robot provides repeatable and stable positioning as a commercially available platform compliant with international safety standards. Evaluation of intrinsic robot positioning accuracy and joint-level stability was beyond the scope of this work.

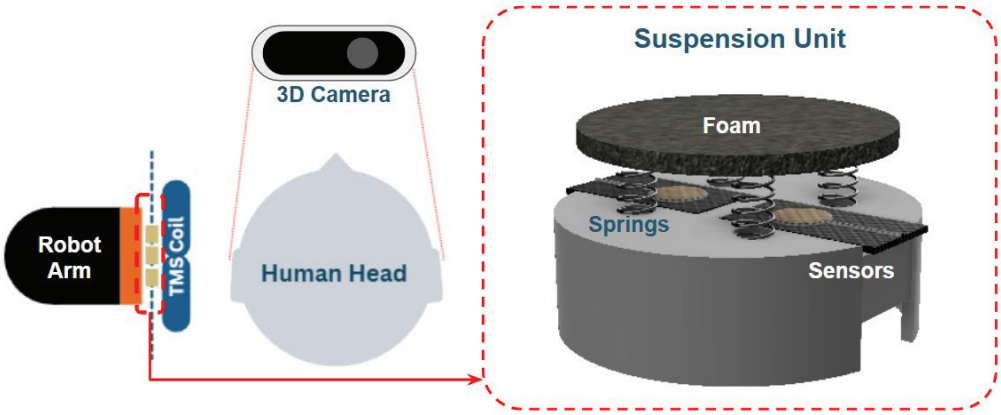


Figure 1. System overview showing camera and robot placement. Suspension platform housing the two capacitive sensors.

All system logic, data acquisition, and robot control were implemented in C# using Visual Studio 2022, with communication to the igus Robot Control (iRC) Release 13 software via the CRI (Container Runtime Interface) protocol. The RealSense SDK and OpenCV-Sharp handled image and depth processing, while DlibDotNet provided 68-point facial landmark detection.

2.2. Hardware Configuration

2.2.1. igus ReBeL 5-DOF Cobot

An igus ReBeL (igus GmbH, Cologne, Germany) 5-DOF cobot serves as the primary actuator for holding and positioning the TMS coil. While 6-DOF robots offer full 3D orientation control, the ReBeL's 5-DOF architecture limits independent pitch and yaw adjustments relative to the patient's head. Accordingly, the control and sensing framework in this study was designed to maintain proper coil alignment and manage tilt within this constrained workspace.

As a collaborative robot, the ReBeL adheres to safety requirements outlined in ISO 10218-1:2011, ISO 10218-2:2011, and ISO/TS 15066:2016, which define functions such as monitored standstill, hand-guiding, speed-and-separation monitoring, and power-and-force limiting [10–12]. In compliance with these standards—particularly the $\leq$250 mm/s limit for safety-rated reduced speed—the robot was operated at only 10–20% of its maximum speed during all experiments. This ensured low-risk operation and stable, controlled coil handling.

For tilt manipulation, the end-effector orientation was controlled in Cartesian space using the robot's tool coordinate system rather than by directly actuating a single joint such as A5, which is the fifth joint of the robot that changes the pitch. Specifically, the tool-frame orientation parameter B was adjusted to vary the coil tilt. This parameter B is displayed in Figure 2, corresponding to pitch rotation of the end effector in the robot's Cartesian coordinate representation (X,Y,Z,A,B,C). Although geometric offsets accounting for the TMS coil thickness and the 3D-printed holder have not yet been incorporated into the tool model, preliminary calibration confirmed that tilt-dependent changes in motor-evoked potential (MEP) amplitude could still be reliably observed.

https://doi.org/10.3390/s26020693

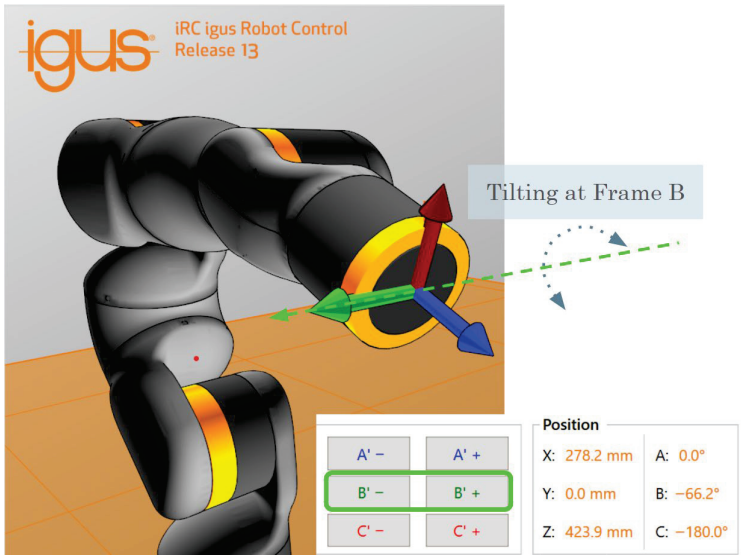


Figure 2. Visualization of the robot tool Cartesian, highlighting Frame B, taken from the iRC software.

2.2.2. Textile Capacitive Sensors

Textile capacitive sensors measure variations in capacitance resulting from deformation and changes in proximity between conductive layers. When integrated behind the TMS coil within a compliant suspension structure, these capacitance variations reflect changes in contact pressure and mechanical compression at the coil–scalp interface. Unlike rigid force or torque sensors, textile-based capacitive sensors provide a soft, distributed sensing surface that conforms to scalp curvature, thereby improving subject comfort and tolerance to minor positioning errors.

Compared to ultrasonic or infrared distance sensors, textile capacitive sensors do not require line-of-sight, are insensitive to hair coverage and surface reflectivity, and can be embedded directly into a compliant mechanical interface. These properties make them particularly suitable for close-contact human–robot interaction scenarios, such as TMS coil positioning, where continuous and stable scalp contact is required.

The textile capacitive sensors used in this study were custom-fabricated in-house by a former doctoral researcher in the same laboratory, in collaboration with the Gunma Industrial Technology Center. Each sensor consists of flexible conductive fabric layers separated by a compliant dielectric material, allowing measurable changes in capacitance in response to pressure or deformation. While the sensors employed in this system were custom-built, the design is readily replicable, and alternative fabric-based or capacitive sensors may be substituted, provided they offer stable analog outputs compatible with the data acquisition hardware. This flexibility supports the broader goal of maintaining affordability and adaptability across different research environments.

Two textile capacitive sensors were integrated between the robot flange and the TMS coil holder, with one sensor positioned at the upper and the other at the lower portion of the end effector. This dual-sensor configuration enables real-time estimation of coil tilt by comparing differential capacitance responses arising from asymmetric compression of the suspension structure.

Data processing for sensor reading:

- A 50-sample baseline for each sensor is collected for calibration
- Baseline is subtracted from real-time reading (capacitance count):

$$\mathrm{Adj_n} = \mathrm{Raw_n} - \mathrm{BaseLine_n}, \quad (1)$$

- Normalized ratio between the two sensors is computed:

https://doi.org/10.3390/s26020693

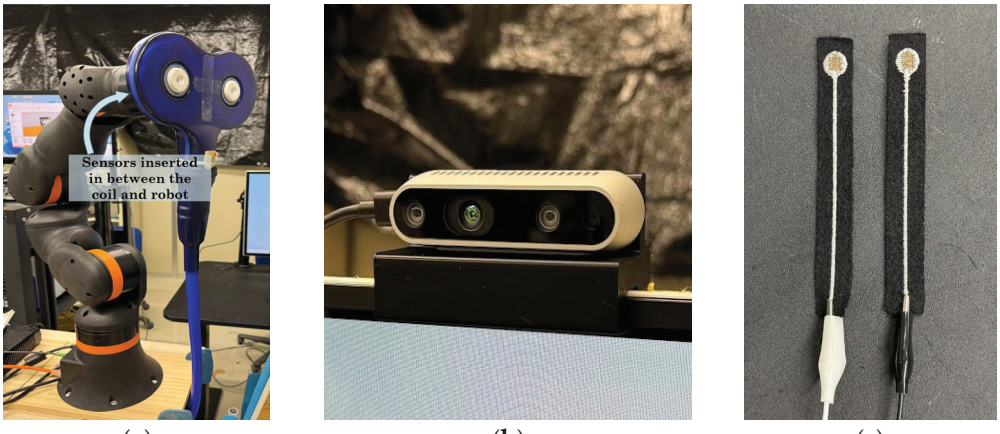
$$\text{Norm}_n = \frac{\text{Adj}_n}{\text{Adj}_1 + \text{Adj}_2} \quad (2)$$

Tilt detection is based on asymmetric compression of the suspension structure. When the coil is approximately perpendicular (tangential) to the scalp surface, both textile sensors undergo similar deformation, resulting in comparable adjusted capacitance values. When the coil is tilted, one side of the suspension compresses more than the other, producing an imbalance between the upper and lower sensor readings. The normalized ratio therefore provides an indirect estimate of coil tilt relative to the local scalp surface.

With the total load (Adj_1 + Adj_2) normalized to unity, normalized values approaching 0.5 indicate balanced contact between the two sensors, corresponding to a coil orientation that is approximately tangential to the scalp. Deviations from this balance indicate increasing tilt.

An automatic tilt-correction algorithm utilizing this sensor ratio was implemented in the control software. However, this closed-loop correction was not activated during the experimental procedures reported in this study. The algorithm was functionally tested during system development to verify correct execution and safe behavior, but it was excluded from formal experiments to isolate and evaluate the sensing and estimation performance of the textile capacitive sensors themselves. Consequently, all tilt adjustments during experimentation were performed manually, and experimental validation of the automatic correction mechanism remains future work.

Overall, capacitive textile sensing was selected because it enables sensitive, continuous measurement of coil–scalp interactions without requiring rigid mechanical interfaces or point-force sensing. The combination of flexibility, durability under repeated compression, and spatially distributed sensing provides a richer representation of load distribution and coil orientation than single-point force sensors, supporting its suitability for robotic-assisted TMS applications. The overview of the hardware components are shown in Figure 3.



(**a**) (**b**) (**c**)

Figure 3. Hardware Components: (**a**) TMS coil attached to igus ReBeL 5-DOF cobot; (**b**) Intel RealSense D435 with 3D printed mount; (**c**) two custom textile capacitive sensors.

2.2.3. Suspension System

To ensure accurate load detection and consistent contact between the TMS coil and the scalp surface, a custom suspension system was designed and 3D-printed to support the textile capacitive sensors. Since the sensors were positioned behind the TMS coil, the direct change in capacitance upon coil–head contact was initially minimal, with the coil itself acting as a source of interference. Instead of filtering out this signal, the system design

https://doi.org/10.3390/s26020693

leveraged it as a form of load detection. Figure 4 displays the effect of the suspension system on the displacement of the coil when pushed.

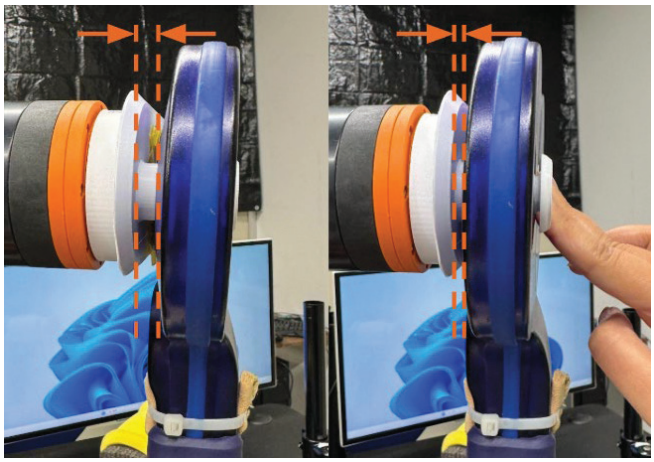


Figure 4. Subtle displacement of the TMS coil, visualized by the dotted lines, when pushed on the suspension system.

The suspension structure consists of a lightweight 3D printed platform supported by soft polyurethane foam and compression springs. When the coil contacts the scalp, the mechanism compresses backward, pushing against the sensors and generating a capacitance change proportional to load. This "soft-landing" design also provides a cushioning effect, which improves safety and comfort during approach. This suspension system aids tilt estimation through differential readings across the two sensors.

2.3. Software Architecture and C3 Localization

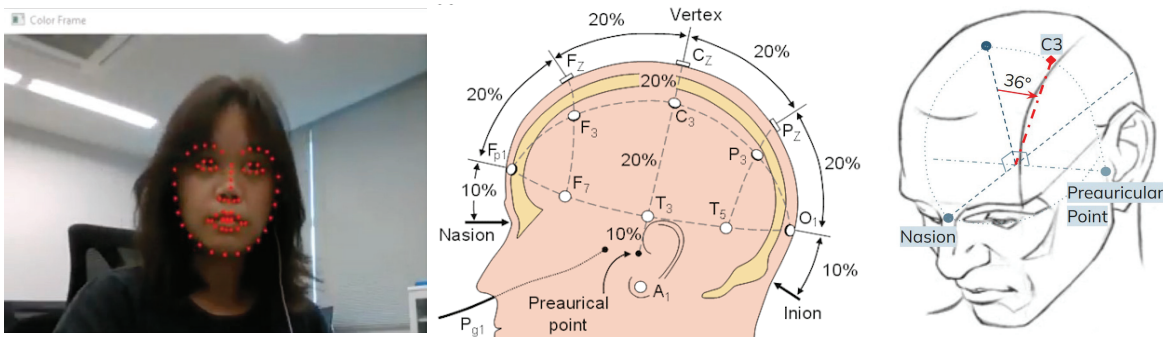
The system was programmed in C# using Microsoft Visual Studio 2022. An Intel RealSense D435 (Intel Corporation, Santa Clara, CA, USA) [13] provided synchronized RGB–D data through the RealSense SDK 2.0 for real-time facial landmark detection. Although the D435 has no major technical advantage over other depth cameras, it is inexpensive, widely supported, and easily integrated with OpenCV and DlibDotNet. The system does not depend on this specific model and any depth camera capable of landmark detection would be suitable.

Facial landmarks were detected using the DlibDotNet library, a C# wrapper for the Dlib C++ machine learning library. This library provides pre-trained models for detecting key facial points [14], including the nasion and preauricular landmarks, which were then used to estimate the C3 position according to the international 10–20 EEG system. In this study, C3 was used as an anatomically defined reference location rather than a subject-specific functional motor hotspot. C3 was selected because it lies over the primary motor cortex in many individuals and is commonly used as an initial targeting point in both clinical and research TMS settings.

Functional motor mapping was not performed to redefine the optimal hotspot location. Instead, MEPs elicited during stimulation were used as a physiological confirmation that the coil was positioned over a motor-responsive cortical region, thereby validating the effectiveness and repeatability of the proposed localization approach. This strategy allowed evaluation of the system's ability to reproducibly position the coil at a consistent anatomical target while avoiding confounds introduced by individualized functional mapping.

Extracted facial landmarks for this project include the nasion, and right and left preauricular points along the jaw. Facial landmark detection and simplified 10–20 method are visualized in Figure 5.

https://doi.org/10.3390/s26020693



(**a**) (**b**) (**c**)

Figure 5. C3 Localization (**a**) Using DlibDotNet shape_predictor_68_face_landmarks.dat [14] to detect facial landmarks; (**b**) International 10–20 Method [15]; (**c**) Visualization of C3 estimation from 10–20 method.

2.3.1. C3 Estimation Procedure

Because the inion cannot be captured in the camera view, the full 10–20 system was adapted using only visible landmarks. The procedure is:

1. Define the transverse plane.
   The distance between the left and right preauricular points is treated as 100% of the transverse arc (180° in the simplified angular model).
2. Approximate the vertex.
   Normally at 50% of the nasion–inion distance; here, it is estimated from the geometric midpoint of the preauricular points and the nasion, which also provides pitch/roll estimation.
3. Locate C3.
   In the 10–20 system, C3 lies 20% of the transverse distance from the vertex. In the simplified model, 20% of 180° is 36°. The C3 direction is therefore the 36° angle from the vertex toward the left ear.
4. Estimate head orientation.
   Using the nasion–ear triangle, pitch and roll are computed, enabling the C3 point to be projected onto a tilt-corrected anatomical plane.

Thus, C3 is estimated as the point lying 36° from the approximated vertex on the tilt-corrected transverse plane.

2.3.2. Coordinate Computation

Because the 5-DOF cobot has limited translation along the depth (Z) axis, Z is treated as constant for the localization task. The target X–Y location is computed as:

$$\mathrm{Ear_x} = \mathrm{RightEar_x} - \mathrm{LeftEar_x};\ \mathrm{Ear_y} = \mathrm{RightEar_y} - \mathrm{LeftEar_y} \tag{3}$$

$$\theta = \mathrm{Arctan}\left(\frac{\mathrm{Ear_y}}{\mathrm{Ear_x}}\right) \tag{4}$$

$$\alpha = 100\cos\theta \tag{5}$$

$$\mathrm{Target_x} = \mathrm{MP_x} - \alpha\tan(36 + \theta) \tag{6}$$

$$\mathrm{Target_y} = \mathrm{MP_y} + \alpha \tag{7}$$

https://doi.org/10.3390/s26020693

These equations can be visualized in Figure 6. The α is an allotted offset as the exact height of the head cannot be accurately detected. 100 mm from the midpoint of the preauricular points is the added offset so that the robot can simply move forward, toward the midpoint at the computed angle.

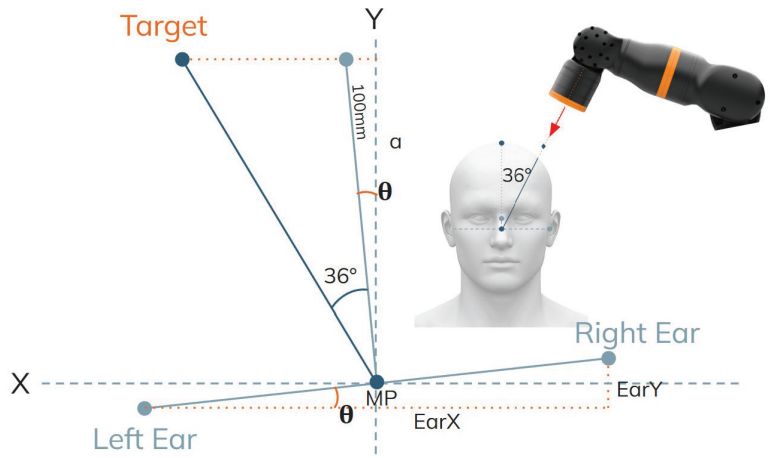


Figure 6. Visualization of C3 coordinate computation.

The resulting XYZ coordinates were sent via the iRC interface using the CRI protocol. Using a Renesas microcontroller board, textile sensors continuously collect capacitance counts to monitor contact and coil tilt. The control logic automatically stopped or retracted the robot whenever the sensor readings indicated excessive capacitance, ensuring safe interaction with the scalp.

2.4. Experimental Procedures

MEPs were recorded from the right first dorsal interosseous (FDI) muscle using surface EMG in a standard three-electrode bipolar configuration, with two electrodes placed over the muscle belly and a reference electrode on a nearby electrically neutral site. EMG signals were amplified (500× gain) and digitized at 5000 Hz using a laboratory-developed MATLAB R2024a-based acquisition system. Recordings were synchronized with TMS pulses via an external TTL trigger, with a 160 ms recording epoch per trial. Signals were preprocessed by voltage conversion (accounting for gain and 2.5 V DC offset), band-pass filtered (20–1000 Hz, 4th-order IIR), and notch filtered at 50 Hz (2nd-order IIR). Zero-phase digital filtering (filtfilt) was applied to avoid phase distortion.

MEP amplitude was defined as the peak-to-peak value (Vmax–Vmin) within a 20–40 ms post-stimulation window. Responses exceeding 50 μV were classified as valid MEPs. A MagStim 200 figure-eight TMS coil was mounted on the igus ReBeL cobot and positioned over the estimated C3 target based on camera-detected facial landmarks. Meanwhile, the textile sensors provided real-time feedback on coil–scalp contact.

2.4.1. MEP Threshold Determination

To ensure comparable data across tilt verification testing, the resting motor threshold (RMT) for each subject was determined using standard clinical procedures. The coil was positioned at the C3 point, and ten single pulses, with resting periods, were delivered at a given intensity. If at least 5 of 10 pulses produced an MEP greater than 50 μV peak-to-peak, that intensity was defined as the RMT. If fewer or more responses were elicited, intensity was increased or decreased accordingly until a 50% response rate was achieved.

All subsequent stimulation was delivered at 120% of the participant's RMT to ensure reliable suprathreshold activation while minimizing discomfort. For example, if the RMT

https://doi.org/10.3390/s26020693

determined is of 50% maximum stimulator output (MSO), 120% of the RMT was used in experimental setups, which in this case results in 60% MSO. It is important to note that different manufacturers and device models cause variations in stimulation output scale. The values reported in this study correspond specifically to the MagStim 200 system (Magstim Co., Ltd., Whitland, UK) [16], as it was the TMS model used for all experiments.

2.4.2. C3 Localization Experiment

This experiment evaluated the accuracy and usability of the C3 localization program and was performed alongside the Tilt Verification Experiment. With the participant stabilized using a headrest, the program computed the XYZ coordinates of the C3 target, and the robot moved the coil to the calculated position. For this test, the final advancement to the scalp was performed manually until sensor-confirmed contact was reached.

After stimulation at a suprathreshold TMS intensity (120% RMT), MEP responses were recorded, whether present or absent. If no MEP was detected, the system was recalibrated and the procedure repeated. For each trial, the number of attempts, the number of camera recalibrations, and any relevant observations were documented.

2.4.3. Tilt Verification Experiment

This experiment was conducted to verify the optimal coil tilt for effective stimulation, determine any other relationships between MEP responses and coil orientation, and assess the performance of the sensors in detecting tilt. Five healthy volunteers completed three trials each. For every trial, the robot positioned the coil at the C3 target coordinate.

Based on prior literature indicating that a ~45° orientation relative to the midsagittal plane optimizes corticospinal activation by aligning the induced field perpendicular to the central sulcus [17,18], the coil was first positioned at 45° and advanced until scalp contact was confirmed by the sensors. From this reference position, the coil was tilted superiorly and inferiorly in 1° increments, spanning ten total conditions (−4° to +5°).

Instead of simply changing the joint angles, tilt adjustments were applied by modifying the pitch orientation of the end-effector (Frame B), ensuring tilt changes occurred without altering the XYZ position of the end-effector. While coil thickness offset has not yet been fully integrated in this system, all tilt conditions represent consistent relative angular deviations from a uniform contact reference. Thus, while absolute orientations may include a small offset, comparisons across tilt conditions remain valid. Rest periods between trials reduced fatigue-related variability.

For every tilt angle variation:

- One suprathreshold TMS pulse was delivered;
- MEP amplitude and graph were recorded;
- Upper and lower textile sensor total load and normalized readings were logged.

3. Results

3.1. Motor Thresholds of Participants

Resting motor thresholds (RMTs) were successfully obtained for all five participants before the tilt verification experiment. All volunteers were male; four were in their early twenties and one in his late forties. RMT values ranged from 50% to 64% MSO, consistent with typical inter-individual differences in corticospinal excitability. These RMT values, which are listed in Table 1, were subsequently used to standardize the stimulation intensity for the tilt test at 120% RMT for each participant.

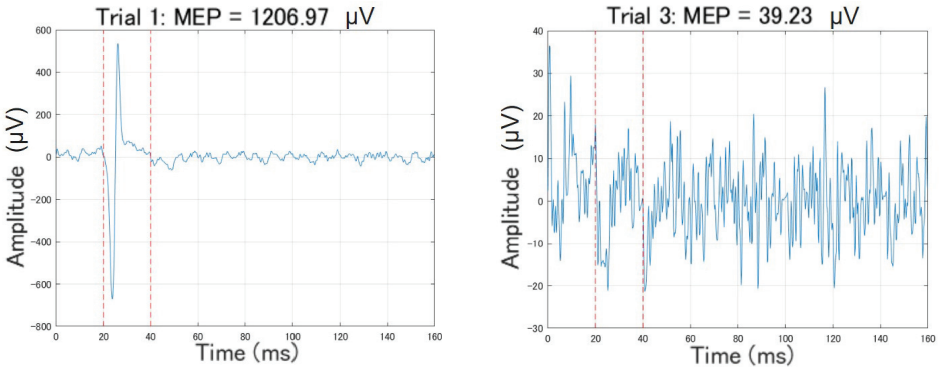
https://doi.org/10.3390/s26020693

Table 1. Participant information and corresponding motor thresholds.

Subject	Dominant Hand	RMT (%)	120% RMT (%)
A	Right	62	74
B	Right	50	60
C	Right	55	66
D	Right	45	54
E	Right	64	77

3.2. C3 Localization

All attempts to position the TMS coil using the developed C3 localization program were recorded, including the total number of trials and any notable deviations in coil placement or subject response. A stimulation attempt was considered successful when the MEP amplitude exceeded 50 μV, which is a commonly used physiological threshold for confirming activation of the affected muscles. Subjects reporting index-finger twitching were also taken into consideration when recording the MEP response as successful or not. MEP responses can be visually differentiated as shown in Figure 7.



(**a**) (**b**)

Figure 7. Contrasting MEP responses: (**a**) Good MEP response with clear peak and >50 μV amplitude; (**b**) Poor MEP response with no clear peaks, very noisy, and <50 μV amplitude. Dashed lines indicate the 20–40 ms post-stimulation window used for MEP evaluation.

Attempts and remarks for all subjects are summarized in Table 2.

Table 2. Summary of localization attempts and relevant notes.

Subject	Trial	Attempts	Remarks
A	1	5	Attempts 1–4: coil not positioned at 45 degrees
	2	1	Successful
	3	1	Successful
B	1	3	Recalibrated for height difference
	2	1	Successful
	3	1	Successful
C	1	4	Physically moved the cobot slightly backward
	2	2	Recalibrated target a bit closer to vertex
	3	1	Successful
D	1	2	Reverted to original target calibration
	2	1	Successful
	3	1	Successful
E	1	1	Successful
	2	1	Successful
	3	1	Successful

https://doi.org/10.3390/s26020693

3.3. Tilt Ratio Verification

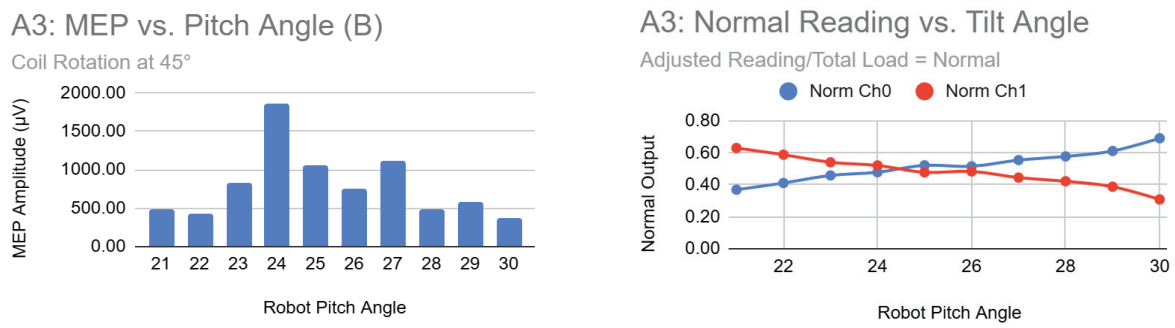
During the tilt verification experiment, the program allowed pausing of sensor updates to record static values per trial. For each recorded angle, the following were documented: robot pitch angle, total load, normalized readings from each of the two sensors, and the resulting MEP amplitude.

Table 3 shows sample data from Subject A, Trial 3. The reference angle, which is the computed 36° from the subject's vertex, is represented by a different value in the robot pitch angle. In this example, the reference angle in the robot pitch is 25°, which varies across trials due to head tilting.

Table 3. Subject A Trial 3—Sample measurements.

Trial	Robot Pitch Angle	Total Load	Norm Ch0	Norm Ch1	MEP (μV)
Trial 3	21	1867	0.37	0.63	492.47
	22	1777	0.41	0.59	434.43
	23	1863	0.46	0.54	834.70
	24	1769	0.48	0.52	1861.40
	25	1669	0.52	0.48	1067.20
	26	1731	0.52	0.48	748.72
	27	1795	0.55	0.45	1108.10
	28	1752	0.58	0.42	483.87
	29	1741	0.61	0.39	572.38
	30	1614	0.69	0.31	367.76

Using the collected data, graphs were generated to illustrate the relationship between tilt angle, sensor values, and MEP amplitude, as shown in Figure 8. Complete plots for all subjects and trials are provided in Appendix A.



(**a**) (**b**)

Figure 8. Subject A Trial 3: Graphs (**a**) MEP Amplitude vs. Robot Pitch Angle; (**b**) Normal Sensor Reading vs. Robot Pitch Angle.

4. Discussion

This study evaluated the feasibility of a robotic-assisted TMS positioning system using textile capacitive sensors and a 3D camera to support coil alignment, contact monitoring, and tilt detection. The findings provide preliminary evidence that the system can (1) localize the C3 motor hotspot with minimal correction after initial calibration, (2) detect coil tilt through differential sensor ratios, and (3) identify a contact-pressure range ("total load") associated with improved MEP responses. The implications of these results are discussed in detail below.

https://doi.org/10.3390/s26020693

4.1. C3 Localization Reliability

Across all participants, the C3 localization program successfully positioned the TMS coil at the target location, requiring only minor calibration adjustments after initial setup. In most trials, a single positioning attempt was sufficient.

Subject A required additional attempts during the first trial due to an incorrect initial coil rotation instead of the recommended 45° for motor cortex stimulation, which prevented reliable MEP generation. Once corrected, localization accuracy was comparable to that of the other participants. Subjects B–E consistently achieved acceptable positioning with one attempt per trial after initial calibration, with Subject E representing the optimal user experience, requiring only one attempt in all trials.

This suggests that the computational method for determining C3 is generally reliable, and when combined with operator familiarity, its reproducibility improves across sessions. Minor calibration discrepancies are likely attributable to camera-related factors, such as depth noise or inaccuracies in point-to-pixel and pixel-to-point transformations. Future work should focus on improving data acquisition stability and refining these coordinate transformations.

4.2. Tilt Verification Implications

4.2.1. Relationship Between Coil Tilt Angle and MEP Amplitude

Across the 15 trials (five participants, three trials each), a clear trend emerged in 9 trials (60%) shown in Figure 9, where MEP amplitudes decreased progressively as the coil was tilted away from the top of the head (increasing robot pitch). In several trials, amplitudes remained high when the coil was tilted toward the top. Compared to the theoretical reference angle of 36° from the top of the head, these observations suggest that tilting away from the top reduces the coil's alignment with the scalp surface, whereas tilting toward the top improves tangent orientation.

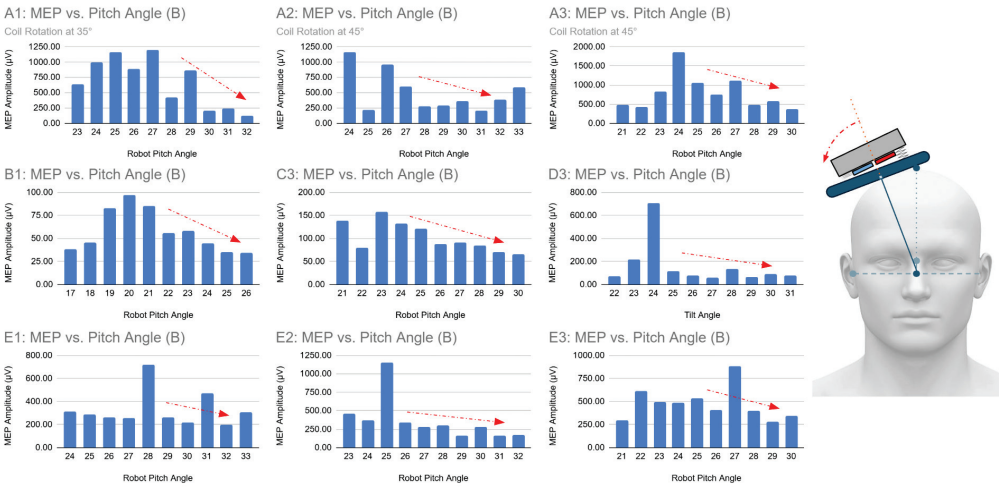


Figure 9. MEP vs. Pitch Angle Graphs showing similar patterns of MEP amplitude decreasing, visualized with dashed arrows, as coil is tilted downwards (higher robot pitch). The tilt direction that causes the pattern is also visualized with an arrow.

In the remaining six trials, the amplitude-to-tilt-angle relationship was less monotonic. However, peak responses still generally occurred near tilt angles that imply a tangent-to-scalp position. These deviations likely reflect inter-subject variability, such as head shape or hair thickness, or sensor noise.

https://doi.org/10.3390/s26020693

4.2.2. Sensor Ratio Correlation with MEP Amplitude

The dual textile sensors mounted vertically on the end-effector provided a reliable proxy for coil tilt. In 12 of 15 trials (80%) shown in Figure 10, the highest MEP amplitudes coincided with sensor readings approaching a 1:1 ratio, indicating symmetric displacement of the suspension system and, by extension, tangent-to-scalp orientation. This supports the principle that maintaining the coil tangent to the scalp optimizes stimulation.

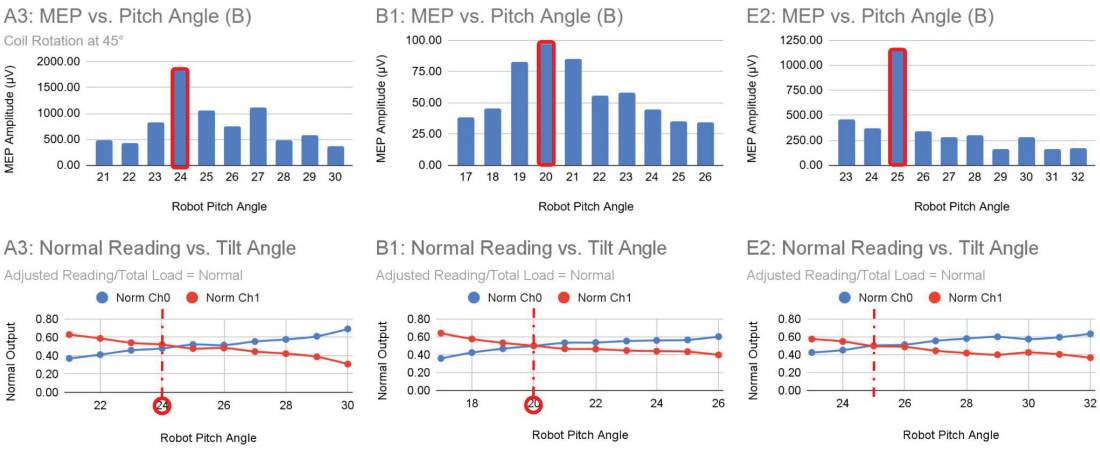


Figure 10. Peak MEP amplitudes, highlighted in red, in three representative trials align with dual sensor readings near a 1:1 ratio, indicating near-tangent coil–scalp orientation. Complete data for all trials are provided in Appendix A.

Importantly, even in trials with noisy or non-intersecting sensor readings, the peaks in MEP amplitude still corresponded to conditions where the sensor ratio approached 1:1. This demonstrates that low-cost textile sensors can provide actionable real-time feedback on coil–scalp alignment, even without precise force sensors or complex impedance control, highlighting their practicality for robotic-assisted TMS.

4.2.3. Inter-Subject Variability

MEP amplitude magnitudes varied substantially across participants, as expected due to differences in head shape and curvature, scalp thickness, individual MT levels, and coil-scalp contact geometry. However, despite these differences, the overall trends wherein the peak MEP amplitude occurs when near tangent-to-scalp alignment and reduced amplitude with increased tilt, remained consistently observable. This indicates that the textile sensor approach performs robustly across anatomical variability and does not require individual anatomical models.

4.3. Limitations in Sensor Consistency

While the textile sensors effectively detected tilt in most trials, some readings were ambiguous. Possible causes include slight sensor misalignment, inconsistent baseline calibration, or variability in head geometry. Additionally, only a single sensor reading was recorded per angle, increasing susceptibility to transient noise.

To improve reliability, future experiments will incorporate averaging over multiple samples (e.g., $\geq$10 readings per angle). This approach should reduce noise sensitivity and yield more stable indicators of tilt. Refinements to sensor mounting and calibration procedures are also expected to improve accuracy and repeatability.

4.4. Additional Findings

While not included in the main objectives of the experiment, a relationship between the MEP responses and total load readings was observed. Stimulations performed under

https://doi.org/10.3390/s26020693

moderate load ranges tended to produce stronger MEPs, while loads that were too low or too high resulted in weakened responses. Specifically, trials with a total load of 1500–4000 were associated with the highest and most consistent MEP amplitudes, whereas loads below 800 and above 6000 produced poor responses. These readings correspond to capacitance counts from the textile sensors, which are arbitrary units reflecting coil–scalp contact pressure rather than absolute force.

This suggests the presence of an optimal contact-pressure window for stimulation, consistent with studies [7] showing that both insufficient and excessive coil pressure can reduce scalp–coil coupling quality. This finding could be integrated into future closed-loop control strategies, allowing the robot to automatically approach the scalp and maintain optimal pressure during stimulation.

4.5. Future System Improvements

Short-term refinements include averaging multiple capacitance readings per angle, standardizing textile sensor calibration to reduce baseline drift, and upgrading camera calibration using an OpenCV checkerboard method to improve depth accuracy and facial landmark reliability. Additional improvements include refining pixel-to-point and point-to-robot transformations and incorporating coil thickness offsets into the kinematic model. Integration of the identified optimal load range into control logic could enable automatic maintenance or correction of coil pressure, advancing the system toward closed-loop operation. Furthermore, incorporating automated tilt correction to consistently achieve tangent-to-scalp orientation can further enhance positioning accuracy, although this feature has not yet been experimentally validated. Future studies could also compare setup time for C3 localization between the robotic-assisted system and traditional manual methods to demonstrate improved speed and usability. Conducting a comparison study between the accuracy and repeatability of this system and other similar systems would also help further verify the feasibility of the study.

Longer-term improvements, requiring additional resources, could significantly enhance system performance. A 6-DOF robot would allow full tilt compensation, while additional textile sensors or higher-resolution 3D/multi-view imaging could further improve scalp–coil coupling. Integration of neuroimaging, such as MRI-based targeting, would increase localization precision and allow individualized stimulation. Together, these upgrades outline a pathway toward a fully autonomous, high-precision robotic TMS positioning system.

5. Conclusions

This study demonstrates the feasibility of a robotic-assisted TMS positioning system integrating textile capacitive sensors and a 3D camera to support coil alignment and tilt detection. The C3 localization program reliably estimated and stimulated the C3 location with minimal calibration, while the dual textile sensors provided actionable, real-time feedback on coil tilting. Experimental results indicate that coil tilt and contact pressure significantly influence MEP amplitude, with optimal responses observed when the coil is in tangent to the scalp and within a moderate load range.

The proposed system shows promise for reducing operator dependency, improving reproducibility, and potentially reducing setup time compared to traditional manual localization methods. A tilt-correction algorithm utilizing differential sensor readings was implemented in the control software. However, its experimental validation was outside the scope of this study and is planned for future work. Short-term refinements, including improved sensor calibration, signal averaging, and camera calibration, are expected to further enhance system stability and estimation accuracy. Longer-term developments, such

https://doi.org/10.3390/s26020693

as the use of higher-DOF robotic platforms, additional sensing modalities, and integration of neuroimaging-based targeting, could advance the system toward more autonomous and higher-precision TMS delivery.

Because the proposed approach relies on scalp landmarks and does not incorporate individual neuroimaging data, anatomical targeting error relative to cortical anatomy or standardized spaces (e.g., MNI) was not quantified in this study. While the overall system includes a robotic arm and a commercial TMS device, the primary contribution of this work lies in reducing the cost and complexity of navigation and contact-sensing components. Compared to MRI-based neuronavigation or optical tracking systems, the proposed approach avoids external markers, dedicated tracking cameras, and force–torque sensors, thereby reducing system complexity and setup burden. Overall, these findings support the practical utility of combining low-cost sensing with robotic assistance for TMS applications and provide a clear roadmap for future enhancements aimed at improving precision, efficiency, and safety in both research and clinical settings.

Author Contributions: Conceptualization, C.D.S.O., J.N.B. and M.O.; methodology, C.D.S.O. and J.N.B.; software, C.D.S.O.; validation, C.D.S.O., Y.M. and M.O.; formal analysis, C.D.S.O.; investigation, C.D.S.O. and Y.M.; resources, M.O. and H.K.; data curation, C.D.S.O.; writing—original draft preparation, C.D.S.O. and J.N.B.; writing—review and editing, C.D.S.O., M.O. and A.C.S.; visualization, C.D.S.O., J.N.B. and Y.M.; supervision, M.O., Y.K. and A.C.S.; project administration, M.O. and Y.K.; funding acquisition, M.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethical Review Committee for Research Involving Human Subjects, Maebashi Institute of Technology (25-003).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Acknowledgments: During the preparation of this manuscript, the author used ChatGPT (GPT-5, OpenAI) to check grammar and provide feedback. The authors have reviewed and edited all outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

TMS	Transcranial Magnetic Stimulation
MDD	Major Depressive Disorder
PDD	Postpartum Depression
SUD	Substance Use Disorder
MRI	Magnetic Resonance Imaging
fMRI	Functional Magnetic Resonance Imaging
DOF	Degrees of Freedom
iRC	Igus Robot Control
CRI	Container Runtime Interface
MEP	Motor Evoked Potential
FDI	First Dorsal Interosseous
RMT	Resting Motor Threshold
MSO	Maximum Stimulator Output

https://doi.org/10.3390/s26020693

Appendix A

This appendix shows all recorded trials for the tilt verification experiment across all participants. Each figure includes the graphs for MEP amplitude vs. robot pitch angle, normalized textile sensor readings vs. robot pitch angle of all trials per subject. These graphs provide a complete overview of coil orientation effects on MEP amplitude and sensor symmetry, complementing the representative data presented in the main text.

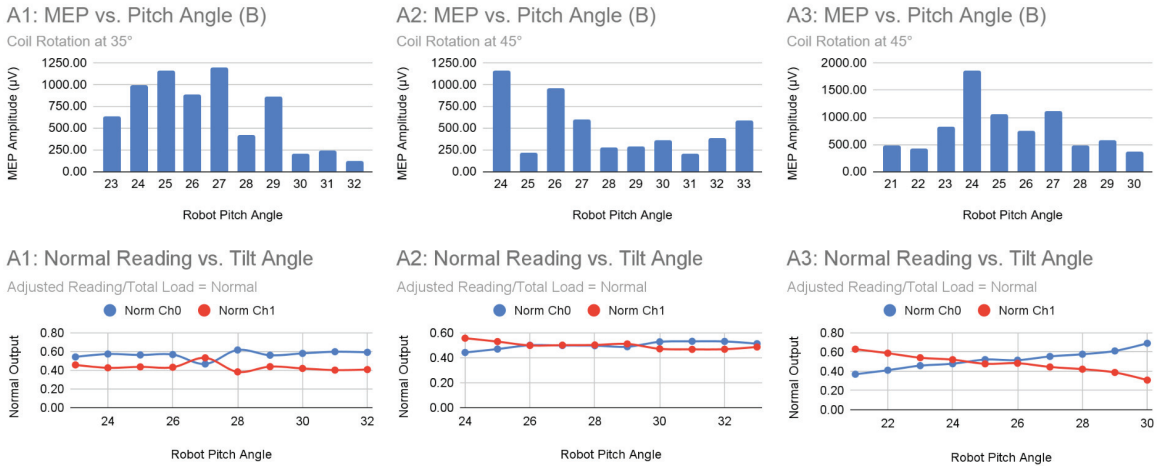


Figure A1. All trials from Subject A.

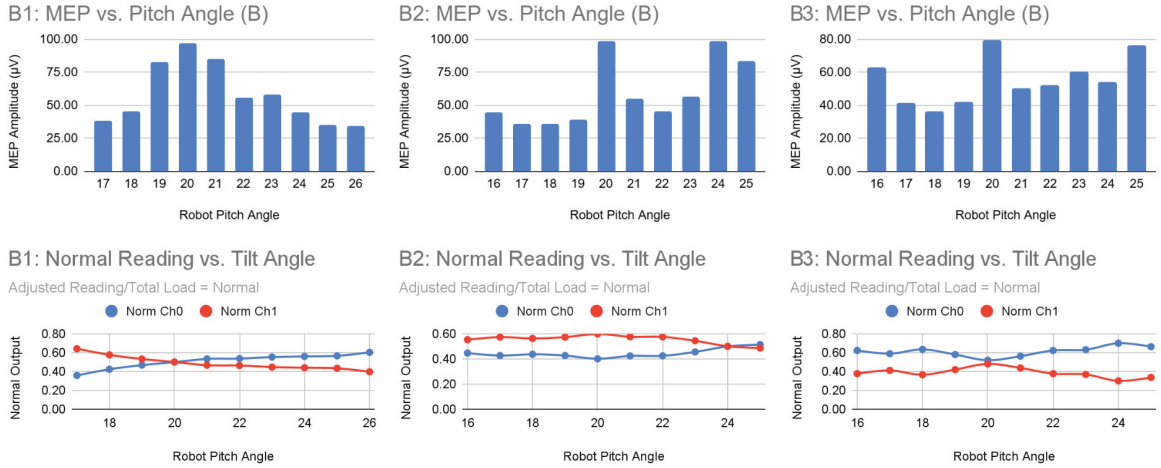


Figure A2. All trials from Subject B.

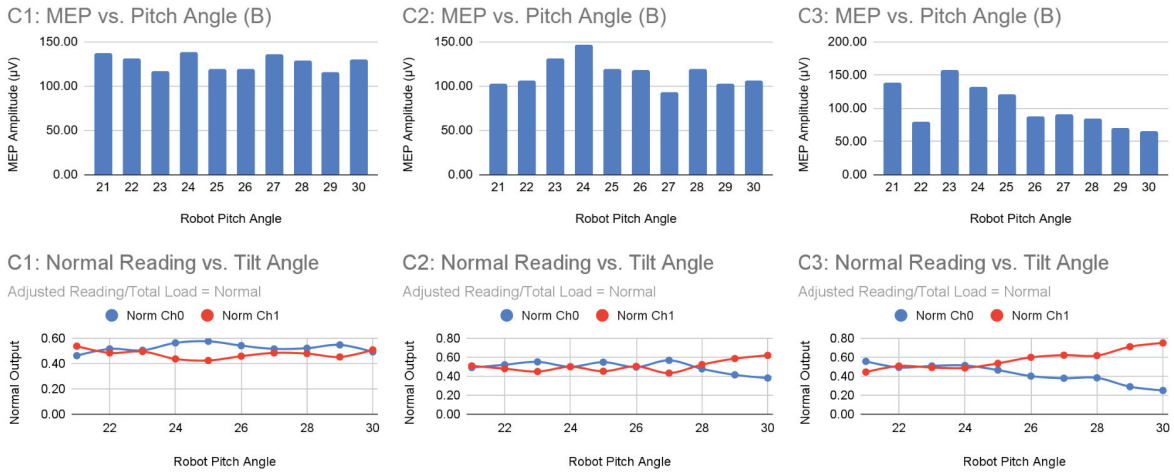


Figure A3. All trials from Subject C.

https://doi.org/10.3390/s26020693

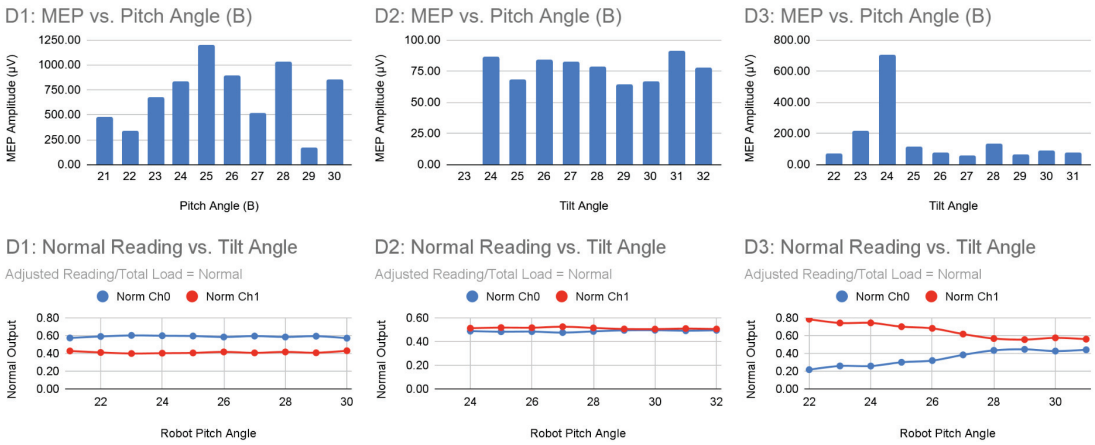


Figure A4. All trials from Subject D.

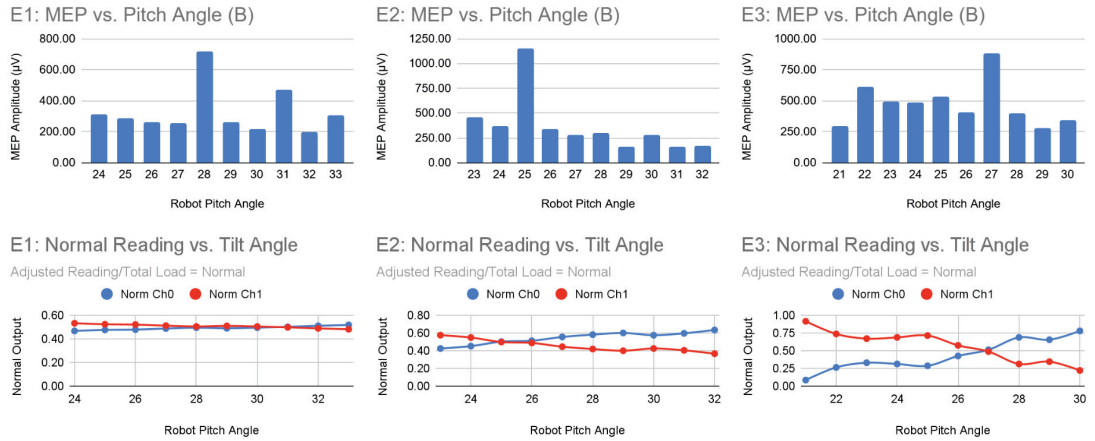


Figure A5. All trials from Subject E.

References

1. Varone, G.; Biabani, M.; Tremblay, S.; Brown, J.C.; Kallioniemi, E.; Rogasch, N.C. The golden age of online readout: EEG-informed TMS from manual probing to closed-loop neuromodulation. *Neuroimage* **2025**, *322*, 121543. [CrossRef] [PubMed]
2. Huddle, M.M.; Costello, S.C.; Barton, D.A. A Systematic Review of the Efficacy of Repetitive Transcranial Magnetic Stimulation Treatment for Women with Postpartum Depression. *Psychiatry Int.* **2021**, *2*, 265–276. [CrossRef]
3. Steele, V.R. Transcranial Magnetic Stimulation as an Interventional Tool for Addiction. *Front. Neurosci.* **2020**, *14*, 592343. [CrossRef] [PubMed]
4. Gellersen, H.M.; Kedzior, K.K. Antidepressant outcomes of high-frequency repetitive transcranial magnetic stimulation (rTMS) with F8-coil and deep transcranial magnetic stimulation (DTMS) with H1-coil in major depression: A systematic review and meta-analysis. *BMC Psychiatry* **2019**, *19*, 139. [CrossRef] [PubMed]
5. MacMaster, F.P.; Croarkin, P.E.; Wilkes, T.C.; McLellan, Q.; Langevin, L.M.; Jaworska, N.; Swansburg, R.M.; Jasaui, Y.; Zewdie, E.; Ciechanski, P.; et al. Repetitive Transcranial Magnetic Stimulation in Youth with Treatment Resistant Major Depression. *Front. Psychiatry* **2019**, *10*, 170. [CrossRef] [PubMed]
6. Sondergaard, R.E.; Martino, D.; Kiss, Z.H.T.; Condliffe, E.G. TMS Motor Mapping Methodology and Reliability: A Structured Review. *Front. Neurosci.* **2021**, *15*, 709368. [CrossRef] [PubMed]
7. Kingsly, E.N.; Bozkurt, I.; Chaurasia, B. Neuro-navigation: Equipment, Tips, and Tricks on Brain Navigated Surgery. *Indian J. Neurosurg.* **2023**, *12*, 193–202. [CrossRef]
8. Jaroonsorn, P.; Neranon, P.; Smithmaitrie, P.; Dechwayukul, C. Robot-Assisted Transcranial Magnetic Stimulation Using Hybrid Position/Force Control. *Adv. Robot.* **2020**, *34*, 1559–1570. [CrossRef]
9. Shin, H.; Ryu, W.; Cho, S.; Yang, W.; Lee, S. Development of a Spherical Positioning Robot and Neuro-Navigation System for Precise and Repetitive Non-Invasive Brain Stimulation. *Appl. Sci.* **2019**, *9*, 4561. [CrossRef]
10. igus GmbH. ReBeL Cobot 5 DOF; igus: Cologne, Germany, 2025. Technical Data. Available online: https://www.igus.eu/product/22534?artNr=REBEL-5DOF-03 (accessed on 23 November 2025).

https://doi.org/10.3390/s26020693

11. igus GmbH. Safety Instructions for ReBeL Cobots; igus: Cologne, Germany, 2024. EC Machinery Directive, Safe Spatial Borders. Available online: https://www.igus.eu (accessed on 23 November 2025).
12. *ISO/TS 15066:2016*; Robots and Robotic Devices—Collaborative Robots, 1st ed. International Organization for Standardization (ISO): Geneva, Switzerland, 2016; Last Reviewed and Confirmed 2022. Available online: https://www.iso.org/standard/62996.html (accessed on 23 November 2025).
13. RealSense, Inc. *Stereo Depth Camera D435*; Product Page; RealSense: Santa Clara, CA, USA, 2025. Available online: https://www.realsenseai.com/products/stereo-depth-camera-d435/ (accessed on 12 January 2026).
14. Takeuchi, T. DlibDotNet: Dlib .NET Wrapper Written in C++ and C#; GitHub Repository, 2025. Available online: https://github.com/takuya-takeuchi/DlibDotNet (accessed on 23 November 2025).
15. Artinis Medical Systems. What Is the 10–20 System for EEG? TMSi—An Artinis Company Blog. 24 January 2024. Available online: https://www.tmsi.artinis.com/blog/the-10-20-system-for-eeg (accessed on 23 November 2025).
16. Magstim Co., Ltd. *MagStim 200*; Magstim: Whitland, UK, 2025. Available online: https://www.magstim.com/product/magstim-m200%C2%B2/ (accessed on 26 November 2025).
17. Janssen, A.M.; Oostendorp, T.F.; Stegeman, D.F. The coil orientation dependency of the electric field induced by TMS for M1 and other brain areas. *J. Neuroeng. Rehabil.* **2015**, *12*, 47. [CrossRef] [PubMed]
18. Adank, P.; Kennedy-Higgins, D.; Maegherman, G.; Hannah, R.; Nuttall, H.E. Effects of coil orientation on motor evoked potentials from Orbicularis Oris. *Front. Neurosci.* **2018**, *12*, 683. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

https://doi.org/10.3390/s26020693

sensors

MDPI

Article

Automatic Grasping System and Hybrid Controller Towards Multi-Drone Parcel Delivery

Bruno J. Guerreiro [1,2,*], Francisco Azevedo [2], Paulo Oliveira [2,3] and Rita Cunha [2]

[1] DEEC/CTS/LASI, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
[2] ISR/LARSYS, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal; francisco.azevedo@tecnico.ulisboa.pt (F.A.); paulo.j.oliveira@tecnico.ulisboa.pt (P.O.); rita.cunha@tecnico.ulisboa.pt (R.C.)
[3] LAETA and DEM, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal
* Correspondence: bj.guerreiro@fct.unl.pt

Abstract

This paper presents the development of an autonomous grasping mechanism for drone-based parcel delivery systems towards developing capabilities for in-flight package transfer. The approach integrates a mechanical gripper fitted with sensors and a pose estimation method for parcels, all coordinated through a hybrid Model Predictive Control (MPC) architecture. The gripper's mechanical structure and prototype are developed using 3D printing technology for both the main framework and gear components. A hybrid dynamical model is formulated that integrates the gripper mechanics with simplified drone dynamics, capturing distinct operational phases including package acquisition, transport, and release. The hybrid MPC framework computes reference trajectories for both the gripper arm configuration and the drone's spatial path toward designated target positions. Experimental validation is conducted using the operational gripper prototype and pose estimation system, while drone behavior is represented through simulation.

Keywords: hybrid MPC; pose estimation; drone delivery; gripper; 3D printing

1. Introduction

In the early stages of unmanned aerial vehicle (UAV) adoption, commonly known as drones, the dominant applications centered on observation and surveillance tasks. However, contemporary technological advances and growing demands for automation and remote operation have expanded the scope of UAV capabilities. Equipping these platforms with manipulation functionality enables not merely environmental observation but active physical interaction with surroundings as well. The inherent mobility of UAVs can increase the potential for ubiquitous object grasping and transportation, eliminating requirements for elaborate ground-based infrastructure. Addressing these opportunities, the REPLACE project [1] pursues the development of a rapid parcel delivery system for urban settings using drone platforms, acknowledging that relay operations between multiple drones become essential when individual vehicle range and endurance prove insufficient for complete delivery missions.

This paper focuses on developing a lightweight, fast, and autonomous grasping system for drone-based parcel exchange operations. The primary objectives are as follows: (i) a gripper design capable of handling parcels with known characteristics under constraints of mass, velocity, and energy consumption; (ii) establishing a perception and control

architecture for the grasping device that enables package pose detection and responsive actuation. This foundation work can afterwords support the creation of basic grasping agents suitable for executing parcel transfers in operational contexts.

The task of aerial package manipulation and transport presents challenges that span a spectrum of scenarios with varying complexity levels. At the simplest level, a vehicle retrieves a package from one stationary ground location and deposits it at another fixed position. Upon approaching the pickup zone, the gripper system engages and computes an optimal approach maneuver and path based on the detected package position. Once the payload is secured, the vehicle navigates to the delivery site for package release. A more demanding scenario involves acquiring objects from mobile platforms (such as another aerial vehicle or ground vehicle), where the cargo remains visible and accessible from above. Advanced scenarios might integrate the motion of both pickup and delivery platforms within a supervisory control strategy to optimize transfer speed and efficiency, such as the work published in [2,3].

Concerning gripper system design, ref. [4] provides an extensive kinematic analysis of 64 linkage-based gripper configurations, establishing a foundation for subsequent investigations. The work in [5] documents established industrial practices and design methodologies for gripping mechanisms, providing foundational principles for the present study. Additionally, ref. [6] addresses the integration of grasping mechanisms with aerial platforms, emphasizing vehicle dynamics and control architecture, whereas [7,8] explore the deployment of both impactive and ingressive gripper types on UAV systems. Comprehensive reviews of aerial manipulation technologies and methodologies are presented in [9–11], which discuss various gripper designs, control strategies, and application scenarios.

For grasping operations in uncertain or incompletely characterized environments, such as aerial load transport, perception and sensing capabilities become critically important. The research presented in [12] addresses this topic by examining UAV modeling and control with explicit consideration of environmental interactions. The work in [13] develops a vision-based control approach for quadrotor perching maneuvers on cables, though the challenge intensifies when accurate pose estimation of target objects becomes necessary, for which the methodology presented in [14] provides an effective and widely adopted solution. Recent developments include sophisticated rigid designs such as the dual-arm aerial manipulator with anthropomorphic grippers [15] or lightweight adaptive gripper for parcel delivery [16]. While rigid grippers offer high load capacity and predictable behavior, soft and soft–rigid hybrid designs provide compliant grasping with improved adaptability to geometric uncertainty, as demonstrated in [17] for food handling and in [18] with variable stiffness grippers. Specifically for the aerial parcel delivery application, ref. [19] provides an thorough review of the state-of-the-art technologies and challenges, highlighting the trade-offs between different gripper designs. The approach presented here emphasizes a streamlined simple grasping device combined with coordinated autonomous control of both vehicle and manipulation subsystems for package exchange operations.

Problems featuring multiple operational phases with both continuous and discrete state dynamics may require hybrid dynamical modeling approaches. The foundational concept of hybrid continuous-time dynamical systems was established in [20], which examined how discrete state variables can capture transitions among continuous dynamics modes and their stability properties. The hybrid automaton framework, introduced in [21] and elaborated in [22], provides a formal representation for hybrid systems, though numerous alternative formulations appear throughout the literature, such as [23]. The work in [24] advances toward more tractable frameworks for describing, analyzing, implementing, and controlling such systems through a Mixed Logical Dynamical (MLD) formulation.

https://doi.org/10.3390/s26020653

The main contributions of this paper are the design, implementation and experimental validation of a new automatic gripping system towards multi-drone parcel delivery comprising the mechanical and electronics systems, a pose estimation method, and a hybrid MPC strategy to achieve automatic planning and control for grasping parcels. The proposed hybrid MPC strategy is capable of coordinating both the drone motion and the gripper actuation through different discrete operation phases, including approach, grasping, and release maneuvers, alowing for the operation design to focus on goals and respective combination of cost functionals and contrainst, rather than using predefined trajectories that might not be able to cope with a moving parcel. The control algorithm is validated in simulation and an experimental validation trials for the gripper system is also presented.

The paper is organized as follows. Section 2 outlines the gripper mechanism design procedure, while Section 3 presents a package pose estimation method based on ArUco fiducial markers. Section 4 develops the dynamical models for both drone and gripper subsystems, subsequently integrating them into a unified hybrid model with an associated hybrid MPC controller. Section 5 describes the implemented prototype along with validation experiments that integrate the various system components, and Section 6 provides concluding observations and directions for future investigation.

2. Gripper Design

Gripping mechanisms represent well-established devices that exhibit diverse configurations depending on their intended application, operational constraints, and available fabrication materials. The objective here is not to present a universal gripper design methodology, but rather to document the development process for a mechanism tailored to the specific requirements of this application.

2.1. Motion Constraints and Prehension

The transported item (or its enclosure) is assumed to possess a rectangular prismatic geometry with uniform mass distribution, which given the vehicle's payload capacity limitations, is constrained to remain below 500 g. For the fundamental scenario where the vehicle acquires and releases cargo at fixed ground locations, the gripper's motion envelope is bounded only by the vehicle frame and ground surface. The gripper arm dimensions must respect both the vertical clearance when in the closed configuration and the lateral clearance required for the fully opened state. A slow prehension sequence would necessitate reducing vehicle speed such that arrival at the target position coincides with jaw closure completion. Following package acquisition, the gripper must retain secure hold throughout the transport trajectory to the destination.

To handle packages of varying dimensions and aspect ratios, the gripper jaw surfaces should maintain parallel orientation throughout their motion.Unlike angular jaw configurations, parallel jaw motion enables grasping from any accessible package surface, providing versatility across multiple acquisition scenarios while ensuring greater tolerance for positioning errors. Gripper mechanisms can be categorized into two primary types: ingressive and impactive. The present work focuses exclusively on the latter category, which operates through a straightforward principle where grasping is accomplished and sustained by normal forces applied by the jaw surfaces against opposing faces of the target object. Package retention results from the friction force F_f generated at these contact interfaces. Selecting appropriate jaw surface materials that yield suitable friction coefficients μ_s when paired with the package surface is essential. For cardboard–rubber material pairs, μ_s typically ranges from 0.5 to 0.8.

https://doi.org/10.3390/s26020653

2.2. Typical Forces During Operation

Developing an appropriate gripper mechanism requires identifying and quantifying the operational forces involved. Given the assumed cuboid package geometry, the contact region between the package and jaw surfaces forms a rectangular area. Larger contact areas enhance retention stability while simultaneously reducing required gripping forces. The adopted configuration employs symmetric bilateral grasping as illustrated in Figure 1.

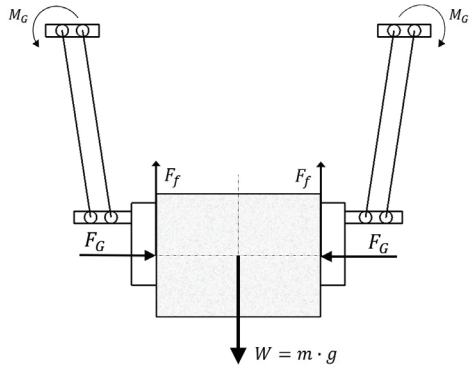


Figure 1. Forces acting on the object.

Each jaw applies a normal force F_G normal to the grasped package surface, whereas the friction force F_f required to prevent the package descent is expressed as $F_f = \frac{m\,g}{n}$, where n denotes the number of contact points (fingers and jaws, in this case, 2), m represents the package mass, and g is the gravitational acceleration. Given the relationship $F_f = \mu_s N$, where N represents the normal contact force, the friction force F_f and gripping force F_G are related through $F_G = \frac{F_f}{\mu_s}$.

During package transport, the vehicle may undergo accelerations beyond gravitational effects in the vertical direction. The most demanding condition arises during rapid ascent maneuvers, when the vehicle experiences maximum acceleration a_{max}, which compounds the gravitational acceleration. Under these circumstances, considering a gripper arm length l, the forces and moments required to maintain secure package retention are, respectively,

$$F_G^* = \frac{m \cdot (g + a_{max})}{\mu_s}, \tag{1}$$

$$M_G^* = F_G^*\, l. \tag{2}$$

2.3. Power Drive Chain

Selecting appropriate power transmission components is critical to meeting the system's minimum velocity, force, and torque specifications. As illustrated in Figure 2, an actuator, specifically an electric motor, supplies torque M_{motor} to the system.

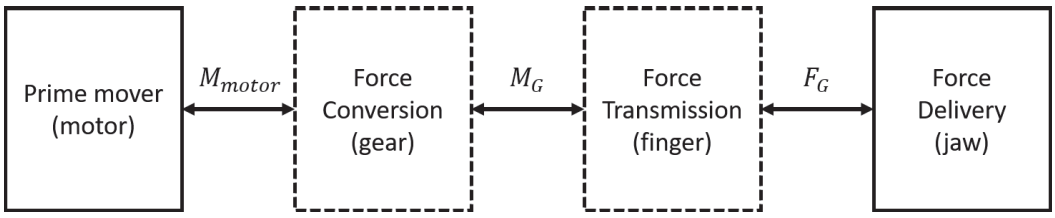


Figure 2. Structure of the gripper forces drive chain.

https://doi.org/10.3390/s26020653

Power transmission from the motor to the gripper finger rotation axis occurs through gear mechanisms, for which spur gears represent the most prevalent and elementary gear type. For a spur gear featuring N teeth, several fundamental geometric parameters can be established. The pitch circle defines a theoretical reference circle forming the basis for geometric calculations. The circular pitch, p, represents the arc length between corresponding points on adjacent teeth, measured along the pitch circle. The pressure angle α characterizes the inclination of the gear tooth profile. The module m serves as the standard ISO parameter for gear tooth sizing in gear nomenclature, defined as $m = p/\pi$, whereas the pitch diameter d_p relates to the module through $d_p = Nm$.

For successful meshing between two spur gears, three principal requirements must be satisfied:

1. The gears must be mounted on parallel shafts;
2. Both gears must share identical module values m;
3. The shaft separation, the center distance, must equal half the sum of the two pitch diameters.

For two properly meshed gears X and Y, their gear ratio is expressed as

$$S_{X-Y} = \frac{N_Y}{N_X} = \frac{d_Y}{d_X}. \tag{3}$$

During meshing operation, the pitch circles of both gears roll without slip, and the velocity at contact point c remains constant for gears with pitch radii r_X, r_Y, as well as angular velocities ω_X, ω_Y. Consequently, the angular velocities satisfy the relationship

$$\frac{\omega_X}{\omega_Y} = \frac{r_Y}{r_X} = \frac{d_Y}{d_X} = S_{X-Y}. \tag{4}$$

Given that properly designed gear meshes exhibit high efficiency with approximately 2% losses, power transmission through the mesh is treated as constant.

The torques on each gear can be determined by equating the mechanical power transmitted, resulting in

$$\frac{T_X}{T_Y} = \frac{\omega_Y}{\omega_X} = \frac{1}{S_{X-Y}}. \tag{5}$$

As such, the relationship between motor torque and the torque at the finger rotation axis is expressed as

$$M_G = \frac{M_{motor}\, S_{total}}{n}, \tag{6}$$

where S_{total} represents the overall gear ratio of the complete transmission train (accounting for multiple gear stages if present).

The grasping system comprises two pairs of gripper arms, with all elements designed to operate symmetrically and in synchronization, as depicted in Figure 3 for a single arm pair. Gears B and C maintain a 1:1 ratio, S_{B-C}, to preserve symmetry within each arm pair, while the ratios S_{A-B} and S_{D-C} are identical, and gear B functions as an idler, reversing the rotational direction of gear A.

Power transfer from the motor to the gear drive shaft employs a worm gear mechanism, which incorporates two components: a worm screw and a worm wheel (or worm gear). The transmission ratio for a worm drive is given by

$$S_{worm} = \frac{N_G}{N_W}, \tag{7}$$

where N_G denotes the tooth count on the worm wheel and N_W represents the number of thread starts on the worm screw. A key advantage of worm drives is their potential for self-

https://doi.org/10.3390/s26020653

locking behavior in certain configurations, where the worm wheel cannot back-drive the worm screw. Finally, combining Equations (1), (2) and (6) yields the minimum permissible total gear train ratio threshold, expressed as

$$S^* = \frac{M_g^*}{M_{motor}}. \tag{8}$$

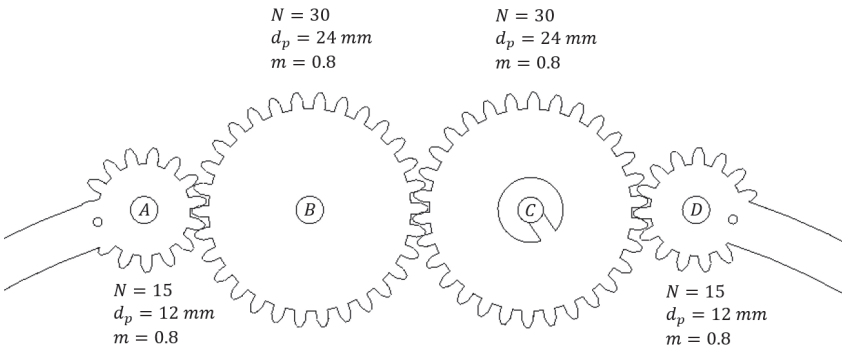


Figure 3. Gear train in one pair of gripper arms.

2.4. Final Prototype and Experimental Assessment

The final gripper design was subject to several tests to validate its performance against the specified requirements. Regarding the use of the worm gear, from (4) and (5), it is possible to calculate the expected values of the gripper jaw's maximum angular velocity and torque provided by the chosen motor and gear set combination, which are 5.39 Nm and 0.94 rad/s, respectively. However, empirical tests that account for losses and force transmission inefficiencies were performed for the angular velocity, as depicted in Figure 4.

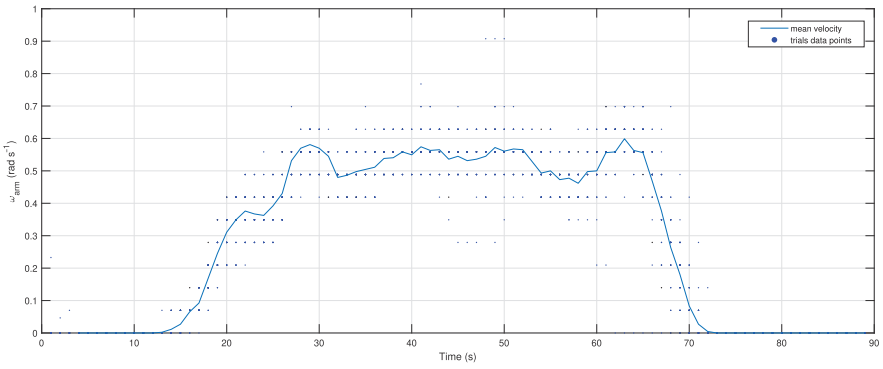


Figure 4. Mean gripper arm angular velocity measured from 30 different trials with the same step input reference.

The arm angular velocity was measured in 30 trials with the same step input reference, and the mean value was computed for each time step. From this data, it can be inferred that the maximum measured angular speed is $\bar{\omega}_{arm} \approx 0.6$ rad/s, which is about 64% of the expected values. Assuming the same losses for the available torque, $M_{motor} = 3.45$ Nm. From (1) and (2), the minimum torque necessary to hold a parcel with $m = 200$ g, considering $\mu_s = 0.8$ and a combined maximum allowable acceleration of 12 ms^{-2}, is $M_G^* = 3$ Nm, which is within the estimated gripper capabilities, as $M_{motor} \geq M_G^*$.

To determine the gripper's current angular position, a potentiometer is attached to one of the gripper arms. Experimental trials were performed to test the gripper going from fully open to fully closed, as shown in Figure 5, where the gripper arm angle θ_{arm} is plotted against time.

https://doi.org/10.3390/s26020653

In addition to position tracking, the system must also verify whether the package has been successfully secured. This is accomplished by identifying instances when the servo motor stalls or experiences exceptional loading, indicating insufficient power to overcome mechanical resistance from an obstacle. A current detection circuit interfaced with the microcontroller was developed to acquire this data, as illustrated in Figure 6.

When stalling is detected during testing, the motor temporarily halts before resuming motion. Current exceeding the calibrated threshold (red dashed line) triggers a motor stop and the system proceeds with the following steps defined for regular operation. Considering the trial depicted in Figure 5, corresponding the current measurement results are shown in Figure 7.

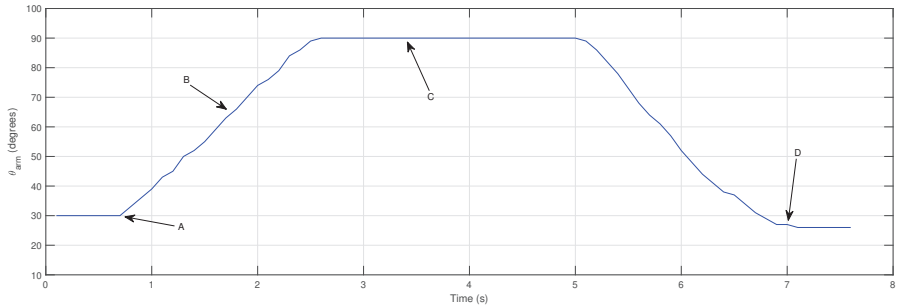


Figure 5. Evolution of the gripper arm angle θ_{arm}, from fully open (A), closing (B), fully closed (C), and back to fully open (D).

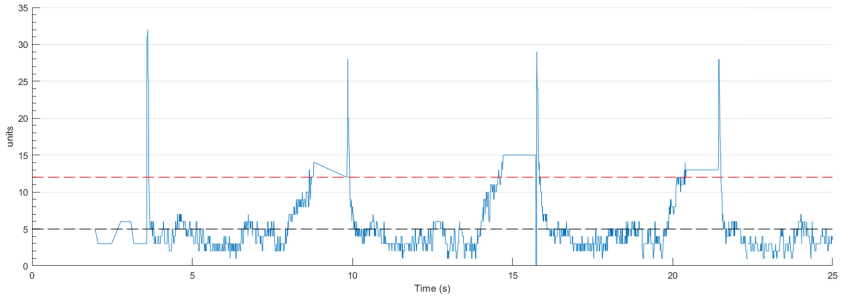


Figure 6. Current sensor circuit diagram. The black dashed line indicates the median current value, whereas the red dashed line represents the threshold established through a calibration procedure.

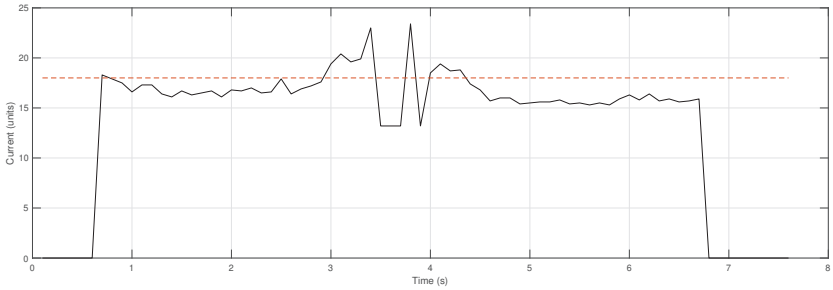


Figure 7. Current measurements and threshold (red dashed line) for the test showed in Figure 5.

Current measurements demonstrate proper system operation, with the microcontroller accurately detecting full gripper closure through characteristic current peaks. When measured current surpasses the predefined threshold, the gripper arms are confirmed to be in contact with either the target parcel or the opposing jaw.

The final gripper prototype depicted in Figure 8, both standalone and integrated in the drone holding a parcel, was also evaluated for its load-bearing capacity during static conditions.

https://doi.org/10.3390/s26020653

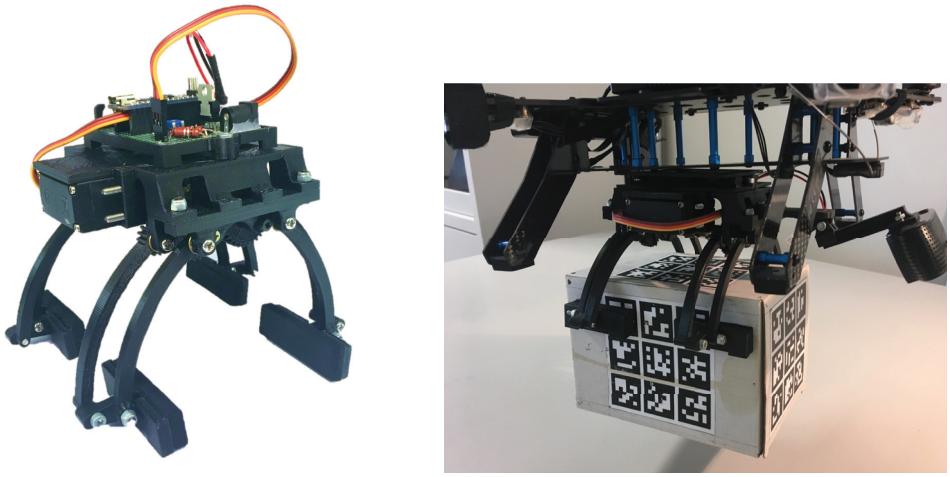


Figure 8. Final gripper prototype as well as the integration with the drone and grasping a parcel.

The bulk of its structure and all of the spur gears were 3D printed using Fused Deposition Modeling with a PLA material, weighing 250 g without the camera and 290 g with the used camera. Experimental tests were performed using a spring-scale dynamometer attached to the bottom surface of a test parcel box to assess the maximum weight that the gripper could hold without electric current being supplied to the motor. It was experimentally observed that the gripper mechanism equipped with the worm gear and end parts fitted with rubber mats could hold cargo of up to 1 kg.

In Table 1, a comparison between the proposed gripper and some recent works found in the literature is presented.

Table 1. Comparison of the proposed gripper with recent works found in the literature.

Gripper	Weight (kg)	Max. Payload (kg)
T Model [25]	0.4	1.5
Dual Arm [15]	1.8	0.75
Adaptive [26]	0.3	0.06
Compliant [27]	0.009	0.06
Proposed	0.25	1.0

Although its purpose is more specific to the application and usage detailed above, it can be seen that it is lighter than most of the rigid grippers presented in [10,11], which usually weigh more than 300 g, while still being able to handle parcels up to 1 kg, which is above the required payload for the intended application and for the payload capacity for which most small drones are designed to carry. Another design option related to the autonomy of the drone+gripper system is the use of the worm gear, which provides self-locking capabilities, avoiding the need for continuous power supply to the motor to keep the parcel grasped during transportation. This type of discussion is seldomly found in the literature, but it is an important aspect to consider when designing aerial manipulation systems, as it can significantly impact the overall energy consumption and flight endurance of the drone.

3. Parcel Pose Estimation

Accurate determination of the target package position can be achieved through various approaches. External positioning systems, such as GPS tracking, represent one possibility, but their positional uncertainties and limited update frequencies render them unsuitable for close-range operations and rapid maneuvers. Conversely, onboard sensing methods,

https://doi.org/10.3390/s26020653

including computer vision or proximity sensors, typically deliver improved accuracy at shorter ranges with fewer environmental obstructions. The adopted approach combines both methodologies: initial acquisition at larger relative distances with relaxed accuracy requirements, transitioning to precise measurement as separation decreases, corresponding to scenarios (a) and (b) depicted in Figure 9, respectively.

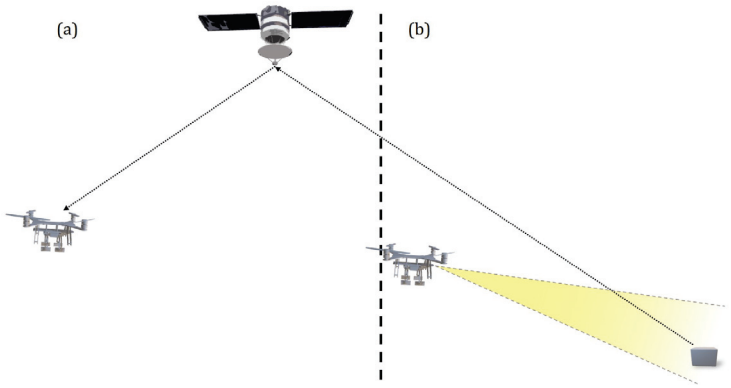


Figure 9. Combination of offboard (**a**) and onboard (**b**) position acquisition.

The present work concentrates exclusively on the onboard sensing phase (b), where a camera establishes correspondences between environmental features and their image plane projections. Incorporating passive markers on the package surface significantly enhances both image capture and processing performance by supplying the detection algorithm with predefined reference points. This approach, termed a fiducial marker system, enables parcel pose estimation relative to a monocular camera with low computational cost, substantial robustness, and rapid processing.

3.1. ArUco Marker System

Fiducial marker systems operate using predefined marker patterns and algorithms that execute detection, error correction, and pose estimation. Among the various implementations available, refs. [14,28] present a computationally efficient and robust square fiducial marker approach employing binary encoding, with capabilities for detecting and estimating poses of individual markers or marker arrays. The *ArUco* library provides an open-source implementation of this methodology, where the marker generation occurs offline through an optimization algorithm that maximizes inter-marker distance and bit transition count, with each marker assigned a unique identifier and stored in a dictionary. Marker detection within images is executed through the *ArUco* library function `detectMarkers()`.

ArUco tags can be deployed either individually or in collective arrangements, where the latter may be distributed across planar surfaces (termed boards) or three-dimensional structures. Three-dimensional marker arrangements prove particularly suitable for package applications, as they provide redundancy to compensate for marker occlusion or partial visibility, enabling detection from arbitrary viewing angles. Constructing a 3D marker structure requires specifying the spatial coordinates of each marker corner, assigning individual marker identifiers, and selecting the appropriate dictionary. Once these parameters are defined, the *ArUco* library function `Board_create()` generates the corresponding 3D structure object.

The package reference frame $\mathcal{P}$ is established based on the marker configuration geometry, as illustrated in Figure 10, with its origin located at the package centroid and axes x_P, y_P and z_P aligned with the box length, width, and height, respectively.

https://doi.org/10.3390/s26020653

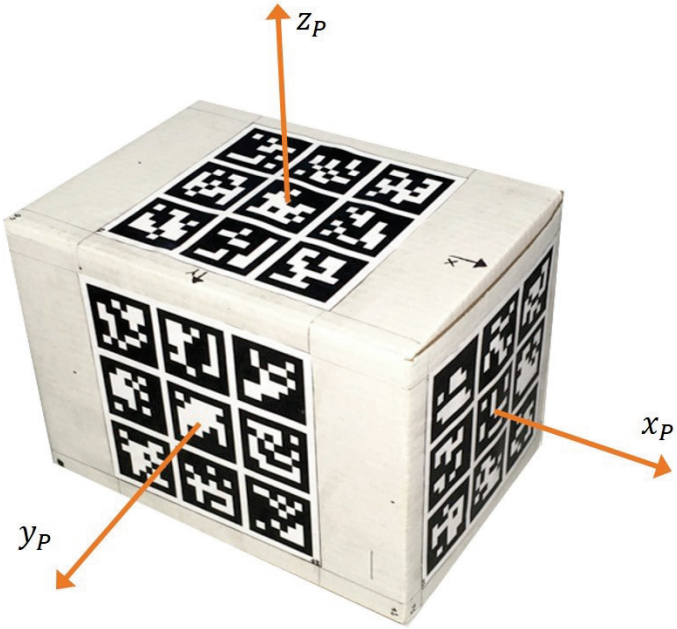


Figure 10. *ArUco* 3D structure with $\mathcal{P}$ coordinate frame.

Determining the pose of an *ArUco* marker structure requires the camera's intrinsic matrix C_I and distortion coefficient vector $\mathbf{D}_{cf}$, which are camera-specific and obtained through calibration procedures. Providing these camera parameters together with the corner coordinates of each of the N_d detected markers, their corresponding *ids*, and the predefined 3D marker geometry to the *ArUco* library function `estimatePoseBoard()` yields the package pose relative to the camera parameterized by a vector $\mathbf{r_v} \in \mathbb{R}^3$ representing the rotation and a vector $\mathbf{t_v} \in \mathbb{R}^3$ the translation, where the latter corresponds to the position of $\mathcal{P}$ expressed in $\mathcal{C}$, denoted by ${}^{C}\mathbf{p}_P$.

The vector $\mathbf{r_v}$ representing the rotation can be converted to the rotation matrix representing the relative attitude of $\mathcal{P}$ as seen by $\mathcal{C}$, via the *Rodrigues'* rotation formula

$$ {}^{C}_{P}R = I + S(\bar{\mathbf{r}}_v)\sin(\alpha) + S^2(\bar{\mathbf{r}}_v)(1 - \cos(\alpha)) \tag{9} $$

where the rotation angle is $\alpha = \|\mathbf{r_v}\|$, the rotation axis is $\bar{\mathbf{r}}_v = \mathbf{r}_v/\|\mathbf{r}_v\|$, I represents the 3×3 identity matrix, and $S(\mathbf{a})$ denotes the skew-symmetric matrix that defines the cross product as $S(\mathbf{a})\mathbf{b} = \mathbf{a} \times \mathbf{b}$, where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$.

3.2. Drone-with-Gripper Perception of Package Pose

The camera frame $\mathcal{C}$ is mounted beneath the vehicle frame $\mathcal{B}$ to maximize parcel visibility while minimizing obstructions during grasping operations. Specifically, $\mathcal{C}$ is offset from the $\mathcal{B}$ origin by a fixed displacement $\boldsymbol{l_o} = [l_o^x, l_o^y, 0]$ and rotated about the y_B axis by angle θ_o. The transformation relating the camera frame $\mathcal{C}$ to the vehicle body frame $\mathcal{B}$ is then expressed as

$$ {}^{B}_{C}R = \begin{bmatrix} \cos(\theta_o) & 0 & -\sin(\theta_o) \\ 0 & 1 & 0 \\ \sin(\theta_o) & 0 & -\cos(\theta_o) \end{bmatrix}, \tag{10} $$

with ${}^{B}\mathbf{p}_C = -l_o$. As such, the complete transformation from $\mathcal{P}$ to $\mathcal{W}$ is given by

$$ {}^{W}\mathbf{p}_P = {}^{W}\mathbf{p}_B - {}^{W}_{B}R\,({}^{B}\mathbf{p}_C + {}^{B}_{C}R\,{}^{C}\mathbf{p}_P), \tag{11} $$

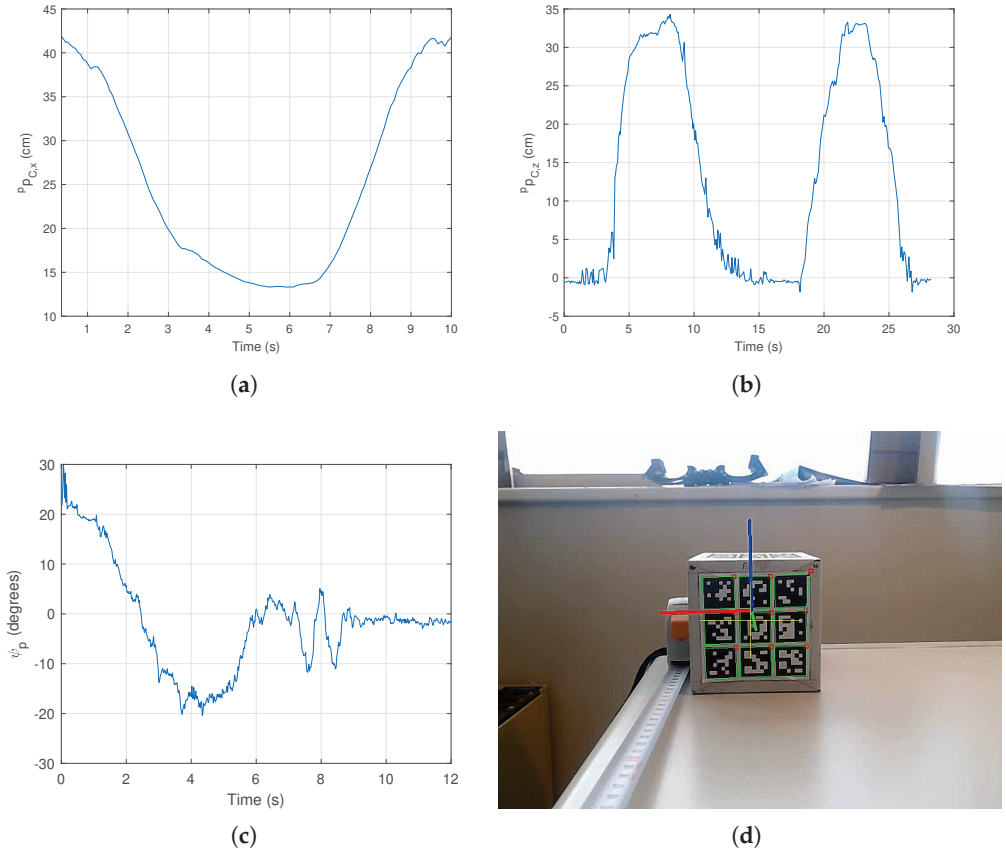
considering the parcel orientation in frame $\mathcal{W}$ given by ${}^{W}_{P}R = {}^{W}_{B}R\,{}^{B}_{C}R\,{}^{C}_{P}R$.

https://doi.org/10.3390/s26020653

3.3. ArUco Pose Estimation Evaluation

To evaluate the quality of the discussed method and its applicability in our proposed scenario, a group of tests were made, resembling the expected working conditions. The camera used for these tests was a C290 (by Logitech International S.A., Lausanne, Switzerland) with a stated image resolution of 800 × 600. The sample rate of the pose estimation algorithm is strongly dependent on the camera frame rate and on the computer processing capabilities, which in this case were both able to properly function at 30 Hz.

To better assess algorithm performance and interpret results, outputs are presented as camera pose relative to the parcel frame, ${}^{P}p_{C}$. Assuming horizontal parcel placement, $\mathcal{P}$ coincides with $\mathcal{W}$, allowing direct interpretation as ground distances. A first test involved horizontal approach along the x axis, a second examined z axis motion, where non-perpendicular observation increases motion blur susceptibility, and a third test the rotation about the z axis was tested. These tests are depicted in Figure 11, along with a picture of the testing environment.



(**a**) (**b**) (**c**) (**d**)

Figure 11. Pose estimation tests: (**a**) x axis approach; (**b**) z axis approach; (**c**) yaw rotation; (**d**) testing environment.

While no groundthrough was available, qualitative assessment confirms adequate accuracy of the pose estimation strategy with minimal noise, attributed to the perpendicular observation angle reducing motion blur. It is also noticeable that z axis motion reveals increased noise, though remaining acceptable at approximately 1 cm magnitude, which can be mitigated throught simple filtering techniques.

4. Hybrid Grasping Model Predictive Control

This section develops dynamic models for both the vehicle and gripper subsystems, subsequently integrating them into a unified hybrid model capable of representing multiple

https://doi.org/10.3390/s26020653

operational scenarios and modes. The objective is to formulate a Hybrid Model Predictive Controller (HMPC) that enforces all critical constraints with minimal deviation. Figure 12 illustrates the control architecture employed in this work, wherein the HMPC computes reference signals for both the gripper and vehicle controllers.

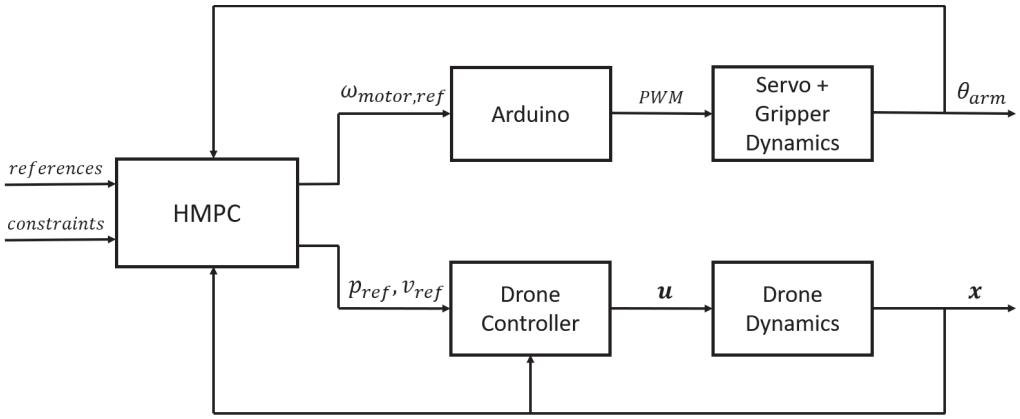


Figure 12. Control loop including the hybrid MPC combining the drone and gripper dynamics.

4.1. Drone Dynamics

Developing a fully autonomous grasping system for aerial parcel delivery necessitates establishing appropriate quadrotor dynamics and control models, for which the approach followed in [29] is adopted. The rotation matrix of the vehicle body frame $\mathcal{B}$ relative to the world frame $\mathcal{W}$, denoted as ${}^{W}_{B}R$ or simply as R, can be parameterized using, for instance, the ZYX Euler angles ϕ (roll), θ (pitch), and ψ (yaw), respectively. This rotation matrix can also be recovered by composing three simple rotations based on the Euler angles, according to the ZYX sequence. The angular velocity of $\mathcal{B}$ relative to $\mathcal{W}$ expressed in $\mathcal{B}$ is denoted as $\boldsymbol{\omega} \in \mathbb{R}^3$, which can also be related to the time derivatives of Euler angles through an appropriate transformation. Additionally, the position of the origin of $\mathcal{B}$ relative to $\mathcal{W}$ is denoted by $\mathbf{p} \in \mathbb{R}^3$, whereas its linear velocity is $\mathbf{v} \in \mathbb{R}^3$. Thus, the kinematics and dynamic differential equations that describe the motion of the vehicle can be written as

$$\dot{\mathbf{p}} = \mathbf{v} \tag{12}$$

$$\dot{\mathbf{v}} = -g\mathbf{z}_W + \frac{u_1}{m}\mathbf{z}_B \tag{13}$$

$$\dot{R} = RS(\boldsymbol{\omega})_B \tag{14}$$

$$\dot{\boldsymbol{\omega}} = J^{-1}(-S(\boldsymbol{\omega})J\boldsymbol{\omega} + \boldsymbol{u}_\tau) \tag{15}$$

where m is the mass of the vehicle, g is the gravitational acceleration, J is the inertia matrix of the vehicle, u_1 is the total thrust generated by the rotors, and $\boldsymbol{u}_\tau = [u_2\ u_3\ u_4]^T$ is the vector of moments applied to the vehicle in its body frame. Also, vectors $\mathbf{z}_w$ and $\mathbf{z}_B$ are the z axis in $\mathcal{W}$ and $\mathcal{B}$, respectively. These two vectors are related by $\mathbf{z}_B = R\mathbf{z}_w$. An appropriate control law, capable of providing the motor input vector $\boldsymbol{u}$ that is able to follow a desired trajectory based on position, velocity, and orientation references is thoroughly described in [29].

Considering a hierarchical control strategy, a drone might use several control loops that account, progressively, local control laws for angular velocity, attitude, linear velocity, and position. As typical autopilots provide such inner-loop control laws, a high-level controller such as the one considered here can assume that, if a velocity reference is computed, the autopilot inner loops can easily follow that reference. Thus, the high-level drone model used for integrated guidance and control can be greatly simplified, simply considering $\dot{\mathbf{p}} = \mathbf{v}$ and $\dot{\psi} = \omega_z$, where the inputs are now considered to be the drone

https://doi.org/10.3390/s26020653

linear velocity, $\mathbf{v}$, and the angular rate about the z axis, ω_z. A discrete-time version of this model can also be defined, considering normalized velocity and yaw-rate inputs, $\boldsymbol{u}_v \in [-1,1]^3$ and $u_\psi \in [-1,1]$, respectively, as well as one sample time delay, T_s, on the velocity input, yielding

$$\boldsymbol{p}(k+1) = \boldsymbol{p}(k) + T_s C_v \boldsymbol{u}_v(k-1) \tag{16}$$

$$\psi(k+1) = \psi(k) + T_s C_\psi u_\psi(k) \tag{17}$$

where C_v and C_ψ are constant parameters. Considering the drone state vector $\boldsymbol{x}_d(k) = [\boldsymbol{p}(k)\ \boldsymbol{u}_v(k-1)\ \psi(k)]^T$ and respective input vector $\boldsymbol{u}_d(k) = [\boldsymbol{u}_v(k)\ u_\psi(k)]^T$, this model can be rewritten as

$$\boldsymbol{x}_d(k+1) = \boldsymbol{A}_d \boldsymbol{x}_d(k) + \boldsymbol{B}_d \boldsymbol{u}_d(k) \tag{18}$$

where

$$\boldsymbol{A}_d = \begin{bmatrix} \boldsymbol{I} & T_s C_v \boldsymbol{I} & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \qquad \boldsymbol{B}_d = \begin{bmatrix} \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 1} \\ \boldsymbol{I} & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & T_s C_\psi \end{bmatrix} \tag{19}$$

4.2. Dynamic Model of the Gripper

The only actuator in the gripper system is a servo motor modified to be able to have an infinite rotation span, controlled by an Arduino microcontroller. This modification removes the original position feedback capability of a servo motor, but enables its position control.

To model the motor dynamics, simple identification tools where used and a 2nd-order discrete-time system is found to be sufficiently accurate, relating the motor angular velocity $\omega_m \in \mathbb{R}$ with a normalized input $u_m \in [-1,1]$, yielding

$$\omega_m(k+1) = -a_1\omega_m(k) - a_2\omega_m(k-1) + b_1 u_m(k) + b_2 u_m(k-1) \tag{20}$$

where a_1, a_2, b_1, and b_2 are constant model coefficients. The gripper arm angular velocity is defined as $\omega = S_{total}\omega_m$, where S_{total} is the combined gear ratio from the motor to the gripper arm, and the angular position of the arm can also be defined as $\theta = S_{total}\theta_m$, where θ_m is the angular position of the motor. Thus, considering a sample time T_s, $A_i = S_{total}a_i$, and $B_i = S_{total}b_i$, the discrete-time dynamics of the gripper can be defined as

$$\theta(k+1) = \theta(k) + T_s\omega(k) \tag{21}$$

$$\omega(k+1) = -A_1\omega(k) - A_2\omega(k-1) + B_1 u_m(k) + B_2 u_m(k-1) \tag{22}$$

Considering the gripper state vector $\boldsymbol{x}_g(k) = [\theta(k)\ \omega(k)\ \omega(k-1)\ u_m(k-1)]^T$, this model can be rewritten as

$$\boldsymbol{x}_g(k+1) = \boldsymbol{A}_g \boldsymbol{x}_g(k) + \boldsymbol{B}_g \boldsymbol{u}_m(k) \tag{23}$$

where

$$\boldsymbol{A}_g = \begin{bmatrix} 1 & T_s & 0 & 0 \\ 0 & -A_1 & -A_2 & B_2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \boldsymbol{B}_g = \begin{bmatrix} 0 \\ B_1 \\ 0 \\ 1 \end{bmatrix} \tag{24}$$

Based on the geometry illustrated in Figure 13, the jaw separation resulting from motor rotation is given by $d = 2(a + l\cos(\theta) - b)$, where l represents the gripper arm

https://doi.org/10.3390/s26020653

length, while a and b denote geometric parameters defining the spacing between gripper arm rotation axes.

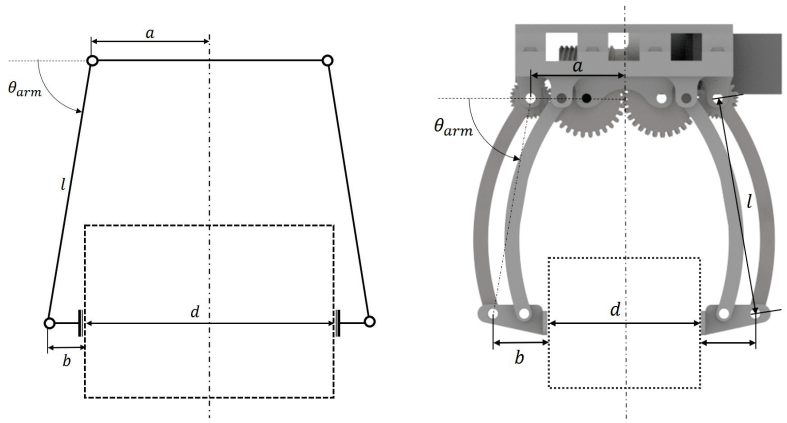


Figure 13. Closed gripper diagram and 3D rendering.

Consequently, for a parcel of width d_{ref}, the required gripper arm angle is

$$\theta_{ref} = \arccos\left(\frac{\frac{d_{ref}}{2} - a + b}{l}\right). \tag{25}$$

The actual limitations of the arms angle is constrained by the geometry of the parts, which considering a zero parcel dimension, yields $\theta_{max} = \arccos((-a+b)/l)$, whereas $\theta_{min} = 0$ due to mechanical limitations.

4.3. Hybrid Model

Given that the system does not consist only of state and input variables representing physical quantities, but also on parts described by logic and discrete evolution, a hybrid model can be formulated. To this end, additional variables are defined in order to better model the system, using a notation where binary variables are represented by a $\boldsymbol{\delta}_{var} \in \{0,1\}^{n_{var}}$ and continuous variables by a $\boldsymbol{\gamma}_{var} \in \mathbb{R}n_{var}$. The hybrid model of the drone and gripper system, based on the models introduced above, is described by a set of continuous state variables: $\theta, \omega \in \mathbb{R}$ representing the angle and angular velocity of the gripper arm; $\boldsymbol{p} \in \mathbb{R}^3$ and $\boldsymbol{v} \in \mathbb{R}^3$ denoting the drone's position and velocity in $\mathcal{W}$, and $\psi \in \mathbb{R}$ denoting the drone's yaw angle. Another important state will be the phase in which the hybrid model of the gripper is, which we can enumerate as $\{A, B, C\}$, characterizing the current mode of operation. To represent this, a binary vector can be used, such as $\boldsymbol{\delta}_{phase} = [\delta_A \ \delta_B \ \delta_C]^T \in \{0,1\}^3$.

Considering first the continuous state variables and their respective equations defined in (18) and (23), defining the continuous state vector $\boldsymbol{x}_c(k) = [\boldsymbol{x}_g(k)^T \ \boldsymbol{x}_d(k)^T]^T$ and respective input vector $\boldsymbol{u}(k) = [u_m(k) \ \boldsymbol{u}_d(k)^T]^T$, the following state equation defines the discrete-time drone and gripper continuous dynamics:

$$\boldsymbol{x}_c(k+1) = \boldsymbol{A}_c\boldsymbol{x}_c(k) + \boldsymbol{B}_c\boldsymbol{u}(k) \tag{26}$$

where

$$\boldsymbol{A}_c = \begin{bmatrix} \boldsymbol{A}_g & \boldsymbol{0}_{4\times7} \\ \boldsymbol{0}_{7\times3} & \boldsymbol{A}_d \end{bmatrix} \qquad \boldsymbol{B}_c = \begin{bmatrix} \boldsymbol{B}_g & \boldsymbol{0}_{4\times4} \\ \boldsymbol{0}_{4\times1} & \boldsymbol{B}_d \end{bmatrix} \tag{27}$$

The evolution of the variable $\boldsymbol{\delta}_{phase}$ can be described by the diagram in Figure 14.

https://doi.org/10.3390/s26020653

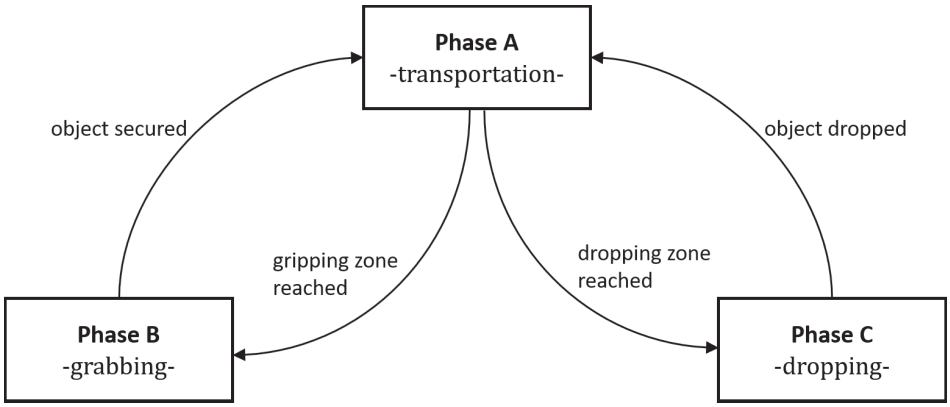


Figure 14. Diagram of the different stages of the hybrid model.

Three operational phases are defined to capture both discrete state transitions and continuous dynamics variations within (26). Phase A represents conditions where only the drone motion is affected by the controller while the gripper remains inactive, either fully closed or fully opened, which encompasses approach and transport operations, during which the vehicle navigates toward a designated target location. Phase B spans the interval from initiation of the grasping maneuver until secure package capture is achieved at the pickup location $\boldsymbol{p}_{grab} \in \mathbb{R}^3$, where both the drone motion and the gripper are actively controlled. Phase C governs package release operations at the delivery location $\boldsymbol{p}_{drop} \in \mathbb{R}^3$, which also implies the control of both gripper and drone motions. Transitions from phase A into phases B and C occurs upon entering proximity zones around the respective target locations, characterized by $\|\boldsymbol{p}_a - \boldsymbol{p}\| \leq d_{zone}$, where $\boldsymbol{p}_a$ is either $\boldsymbol{p}_{grab}$ or $\boldsymbol{p}_{drop}$ and d_{zone} is a constant parameter. To each stage corresponds a binary variable, δ_A, δ_B, or δ_C, constrained by $\delta_A + \delta_B + \delta_C = 1$, meaning that at any point in time, the system can only be in one of the phases. Concerning the gripper arm rotation span, an auxiliary variable $\delta_{close} \in \{0, 1\}$, specifying when the gripper jaws are fully closed, is created. It is defined as

$$\delta_{close} = \begin{cases} 0, & \text{if } \theta < \theta_{close} \\ 1, & \text{if } \theta \geq \theta_{close} \end{cases} \tag{28}$$

where θ_{close} is a predefined angle that varies according to the box's dimensions and can be calculated from (25). An additional binary variable δ_{force} conveys information about reaction forces applied to the gripper arms, indicating whether cargo is actively being held in these operational phases. The gripper state $\delta_{gripper}$, indicating successful parcel acquisition, is determined through logical combinations of these variables and their complements (denoted by the $\neg$ operator), expressed as

$$\delta_{gripper} = \begin{cases} 0, & \text{if } \neg\delta_{force} \wedge \delta_{open} \\ 1, & \text{if } \delta_{force} \wedge \delta_{close} \end{cases}. \tag{29}$$

Upon successful object capture, the system must immediately transition back to phase A, whereas an identical transition occurs following gripper opening. These constraints are expressed as

$$\delta_{gripper}(k) \wedge \delta_B(k) \implies \delta_A(k+1), \tag{30}$$

$$\neg\delta_{gripper}(k) \wedge \delta_C(k) \implies \delta_A(k+1). \tag{31}$$

https://doi.org/10.3390/s26020653

A final binary auxiliary variable is necessary to indicate if the drone has reached passed its target location, denoted as δ_{passed}.

With state and auxiliary variables established, the hybrid model can be formulated as a Discrete Hybrid Automaton (DHA). Expressing the model as a DHA enhances comprehension and provides rigorous formalization of the relationships among dynamics and constraints. This formulation further enables systematic conversion to alternative representations, such as Mixed Logical Dynamical (MLD) systems, which characterize the system through linear difference equations incorporating both continuous and binary variables alongside linear inequality constraints, making it well-suited for hybrid model representation and optimization-based control implementations. The HYbrid System DEscription Language (HYSDEL), introduced in [30], provides a modeling framework for specifying DHA models, whereas the methodology for converting to MLD systems and associated language constructs is detailed in [24].

4.4. Hybrid Model Predictive Controller

Satisfying the hybrid model's requirements, constraints, and objectives is most effectively accomplished through Model Predictive Control (MPC), wherein finite-horizon optimal control problems are solved iteratively at each time step. Selecting the prediction horizon, N_p, necessitates understanding the system's dominant dynamical behavior, as the controller must anticipate critical operational events with sufficient foresight to enable appropriate corrective actions. Idealy, N_p should be greater than the number of time steps necessary to fully close the gripper mechanism to its gripping angle θ_{close}. To prevent the need to use exceedingly large values for N_p, the gripping strategy consists of first moving the gripper to an intermediate closing angle θ_{pre}, after it reaches a predefined safety zone. This way, the final gripping maneuver requires a considerably smaller prediction horizon and, depending on the choice of parameters, for $T_s = 0.1$ s, the most efficient prediction horizon is between 5 and 7 time intervals.

The optimization problem to be solved in each iteration of the MPC is formulated as the Mixed Integer Quadratic Programing (MIQP) problem:

$$\begin{aligned} \min_{q_0} \quad & J_0(x(k), q_0) \\ s.t. \quad & x(k+1) = Ax(k) + B_1u(k) + B_2\delta(k) + B_3z(k) + B_5, \\ & y(k) = Cx(k) + D_1u(k) + D_2\delta(k) + D_3z(k) + D_5, \\ & E_2\delta(k) + E_3z(k) \leq E_1u(k) + E_4x(k) + E_5 \end{aligned} \tag{32}$$

where the cost function J_0 is given by

$$J_0(x(k), q_0) = \|x(N)\|^2_{Q_{x_N}} + \sum_{k=0}^{N-1} [\|y(k)\|^2_{Q_y} + \|u(k)\|^2_{Q_u} + \|z(k)\|^2_{Q_z} + \|x(k)\|^2_{Q_x}] \tag{33}$$

with $\|x\|^2_Q = x^TQx$, q_0 denoting the vector consisting of the optimization slack variables, the sequence of future values of the inputs u, and the auxiliary and output variables δ, z, and y. The cost functional matrix weights, Q_x, Q_y, Q_u, Q_z, and Q_{x_N}, can be chosen according to the relative weights of the current state of the system x, the outputs y, the auxiliary variables z, and the state vector at the prediction horizon x_N, respectively. The Hybrid Toolbox for MATLAB, presented in [31], provides a mechanism to convert the formulation expressed before into a more desirable compact form, accepted by the most common solvers, such as IBM's CPLEX or MATLAB's optimization toolbox.

The Q_y matrix takes advantage of a combination of auxiliary variables to compute a more reliable functional cost. First, it is necessary to define the $y(k)$ vector containing

https://doi.org/10.3390/s26020653

the auxiliary variables. Since there are different objectives for each stage, some auxiliary variables are only relevant when the system is at a specific phase. The notation $\gamma_x^{ph} \backslash \delta_x^{ph}$ is hereby equivalent to $\gamma_x \backslash \delta_x \wedge phase = ph$. The auxiliary variables used to computed the cost of the optimization problem are

$$\boldsymbol{y}(k) = \begin{bmatrix} \delta^B_{\neg close \& passed}(k) & \gamma^B_{close}(k) & \gamma^B_{zone}(k) & \delta^B_{motion}(k) & \gamma^C_{drop}(k) \end{bmatrix}^T. \tag{34}$$

Several auxiliary variables contribute to the cost function formulation. The variable γ_{close} yields the remaining distance to the target when the gripper has achieved full closure, while $\delta_{\neg close \& passed}$ equals unity when the predicted trajectory includes instances where the vehicle has overshot the target without completing gripper closure. The servo motor command u_m is provided by γ_{zone} when the vehicle remains beyond a designated safety radius from the target location. Additionally, δ_{motion} assumes a value of one when the gripper occupies an intermediate configuration rather than a limit position, and γ_{drop} supplies the servo motor command only prior to the vehicle reaching $\boldsymbol{p}_{grab}$.

5. Experimental Results

This section presents the results obtained from both simulations using the developed gripper prototype integrated with the hybrid MPC framework described in Section 4, as well as experimental trials using the parcel perception system outlined in Section 3.

5.1. Hybric MPC Simulation Results

Simulation results were obtained using the hybrid MPC framework described in Section 4, implemented in MATLAB R2018a and using the HYSDEL 3.0 to generate the MIQP problem, which was then solved using IBM CPLEX solver 12.8 (in particular, the solver "cplexmiqp").

The control parameters were determined through a systematic tuning procedure, initially selecting these based on the system's physical properties and time constants, the prediction horizon was chosen to balance computational load and performance, whereas the cost function matrices were adjusted iteratively to achieve desired tracking accuracy and control effort trade-offs. The final parameter values are summarized in Table 2.

Table 2. Hybrid Model simulation parameters.

Parameter	Value
θ_{close}	57.3 deg
θ_{pre}	17.2 deg
k_p	0.9
T_s	0.1 s
N_p	6
Q_x	$\mathrm{diag}([30, 1, 20, 1, 20, 1, 10, 1, 10, 0, 0])$
Q_{x_N}	$\mathrm{diag}([30, 1, 20, 1, 50, 1, 10, 1, 10, 0, 0])$
Q_y	$\mathrm{diag}([10, 100, 1, 50, 10])$
Q_u	$\mathrm{diag}([10, 1, 1, 1, 1])$
$\boldsymbol{p}_{ini}$	$[0, 0, 0]'$
$\boldsymbol{p}_{grab}$	$[5, 10, 1]'$
$\boldsymbol{p}_{drop}$	$[8, 11, 1]'$
$\boldsymbol{p}_{end}$	$[13, 16, 2]'$

After defining the model parameters and calibrating the optimization cost weight matrices, some simulation tests were used to assess the overall performance of the system. To provide the hybrid model and MPC with the required information to simulate a

https://doi.org/10.3390/s26020653

scenario with the 3 different phases (altough phase A repeats after phases B and C), it is necessary to define the state references for each phase. Assuming the initial conditions to be $x_0 = [0, \mathbf{0}_{1\times3}, \boldsymbol{p}'_{ini}, \mathbf{0}_{1\times3}, 0]'$, a scenario could be devised where the state references given in each phase are as follows:

- *Phase* A_1: $x_{ref} = [0, \mathbf{0}_{1\times3}, \boldsymbol{p}'_{grab}, \mathbf{0}_{1\times3}, \psi_{grab}]'$
- *Phase* B: $x_{ref} = [\theta_{close}, \mathbf{0}_{1\times3}, \boldsymbol{p}_{grab}, \mathbf{0}_{1\times3}, \psi_{grab}]'$
- *Phase* A_2: $x_{ref} = [\theta_{close}, \mathbf{0}_{1\times3}, \boldsymbol{p}_{drop}, \mathbf{0}_{1\times3}, \psi_{grab}]'$
- *Phase* C: $x_{ref} = [0, \mathbf{0}_{1\times3}, \boldsymbol{p}_{drop}, \mathbf{0}_{1\times3}, 0]'$
- *Phase* A_3: $x_{ref} = [0, \mathbf{0}_{1\times3}, \boldsymbol{p}_{end}, \mathbf{0}_{1\times3}, 0]'$

where ψ_{grab} is the yaw angle of the parcel at the grasping position to be acquired by the camera.

Figure 15 shows the results of a simulation using a reference structure as the one described above.

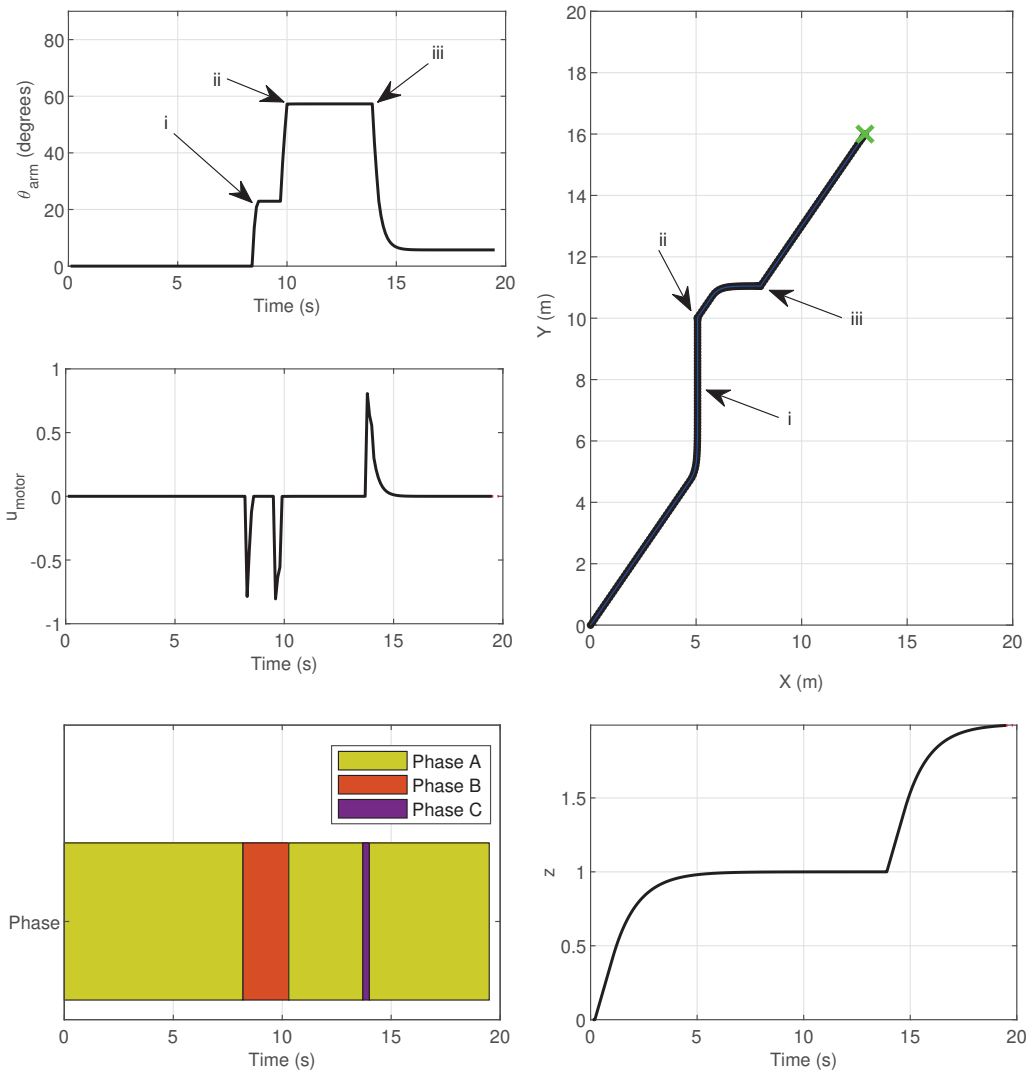


Figure 15. Hybrid MPC simulation. (i) Gripper closes to an intermediate angle θ_{pre} when the drone reaches the safety zone around $\boldsymbol{p}_{grab}$; (ii) Gripper fully closes when the drone reaches $\boldsymbol{p}_{grab}$; (iii) Gripper opens when the drone reaches $\boldsymbol{p}_{drop}$, (green marker).

The hybrid MPC outer-loop controller will relay the reference values of ω to the gripper mechanism as well as v and ψ as desired references to the drone inner-loop controllers.

https://doi.org/10.3390/s26020653

It is possible to observe that the gripper arm fully closes at the exact moment (ii) then it reaches the desired position $\boldsymbol{p}_{grab}$. It accomplishes this through the strategy described in Section 4.4, where the gripper arm first rotates to θ_{pre} before fully rotating to θ_{close}. The MPC is able to compute a trajectory that passes through the specified target locations according to the gripper arm angle θ_arm. The dropping of the parcel, represented by the opening of the gripper arm, also occurs at the exact expected moment (iii). These gripper arm movements are enforced by the motor inputs u_{motor} computed with the MPC. It is also confirmed that the hybrid model phases evolve as expected. The system phases transition according to Figure 14. Phase *B*, or the grasping motion sequence, happens between moments (i) and (ii) and is enabled when the drone enter the neighborhood of $\boldsymbol{p}_{grab}$. Phase *C*, or the parcel dropping motion sequence, occurs almost instantaneously after moment (iii), when the drone is at $\boldsymbol{p}_{drop}$. The simulation results indicate that the hybrid MPC is capable of effectively managing the gripper mechanism's operation in conjunction with the drone's movement, ensuring precise timing for grasping and releasing the parcel.

5.2. Grasping Experimental Results

For the experimental trials described hereafter, only the gripper mechanism is actuated, considering the simulated drone dynamics using the model described in Section 4.1 while manually moving the gripper system. The experimental setup that includes the gripper hardware and the HMPC is depicted in Figure 16.

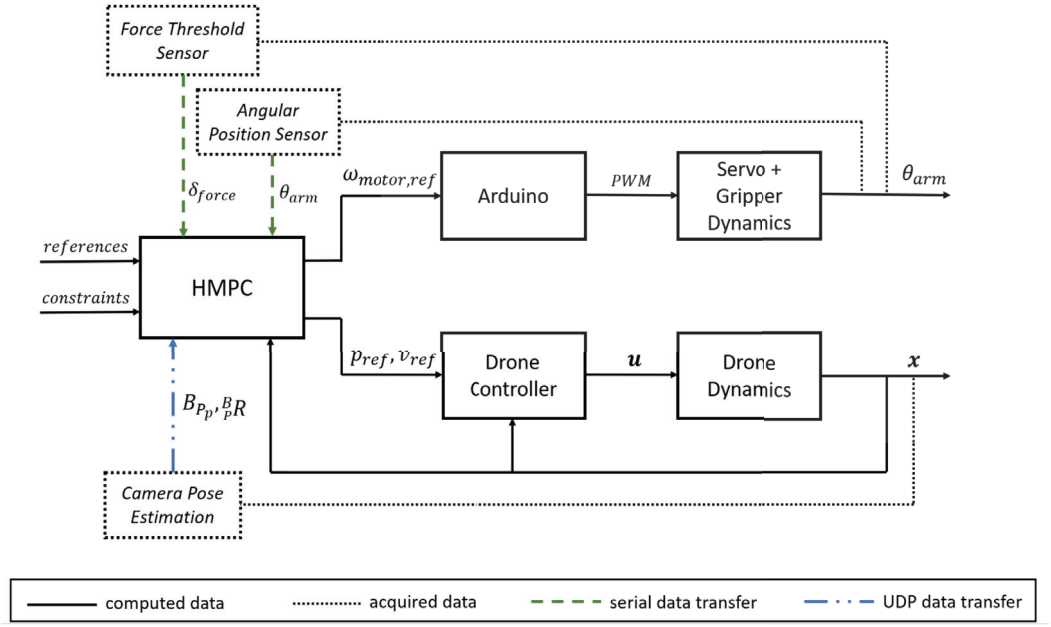


Figure 16. Experimental setup of gripper prototype integrated with the HMPC framework.

The data acquired from the angular position sensor and force threshold sensor was relayed to the main processing unit by an Arduino microcontroller through serial communication at 100 Hz, whereas the camera module provided the parcel position and yaw angle at 30 Hz, relayed to the HMPC solver through a UDP socket. With this information, the HMPC computes at 10 Hz the motor control input u_m as well as the desired drone velocities, $\boldsymbol{v}_ref = \boldsymbol{u}_v$, and yaw rate, $\omega_{z,ref} = u_\psi$, to provide the inner loops for drone motion control and the gripper arm actuation.

Both the yaw angle and position of the parcel are shown in Figures 17 and 18. The drone's position is given by the distance to the parcel where $p_{grab,x} = 1$ m. For this scenario, the gripper arm angle references are defined by $\theta_{pre} = 18°$ and $\theta_{close} = 28°$. This trial is also illustrated in a video (Video S1 in Supplementary Materials).

https://doi.org/10.3390/s26020653

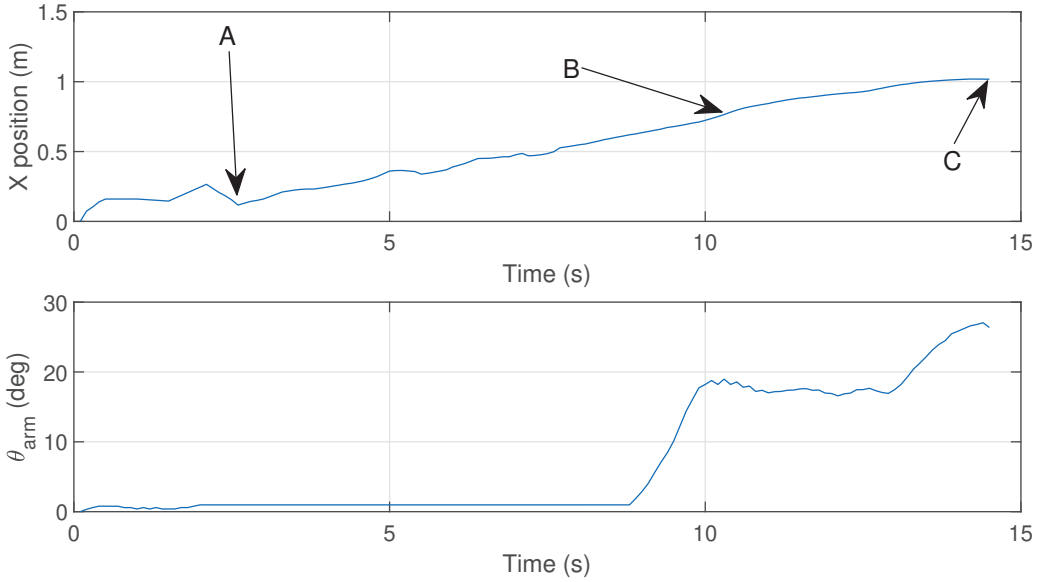


Figure 17. Hybrid MPC experiment: x axis position and gripper arm angle, considering the three moments, also depicted below in Figure 19: (A) Approaching the parcel; (B) Pre-closing the gripper arm; (C) Grasping the parcel.

It is observable that the gripper arm rotates as expected, performing the pre-closing maneuver in order to fully prehend the object in the correct time. The confirmation of a secure parcel prehension comes from the distance computed from the camera in conjunction with the binary variable δ_{force} obtained from the electric current sensor described above.

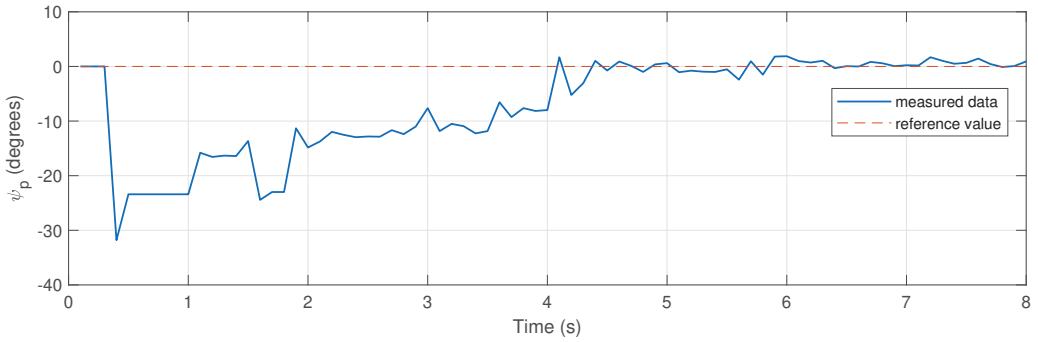


Figure 18. Hybrid MPC experiment: drone–parcel alignment measurements.

Figure 18 shows the yaw angle alignment between the gripper and the parcel during the experimental trial, which is drived towards an acceptable bound during the maneuver, ensuring a successful grasp.

Figure 19 presents snapshots taken by the camera module at different moments during the prehension maneuver, also identified as (A, B, C) in Figure 17.

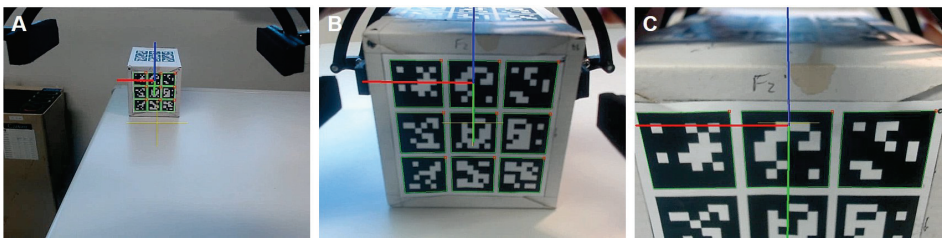


Figure 19. Hybrid MPC experiment: Camera snapshots from different moments during the prehension maneuver: (A) Approaching the parcel; (B) Pre-closing the gripper arm; (C) Grasping the parcel.

https://doi.org/10.3390/s26020653

The presented experimental results validate the proposed hybrid MPC framework's capability to coordinate the gripper mechanism's operation with the drone's positioning, ensuring accurate timing for grasping maneuvers based on real-time parcel pose estimation. Nonetheless, it is important to acknowledge that these trials were conducted under controlled conditions, with the drone's motion simulated and the parcel remaining stationary, and further experimental validation is necessary to assess the system's performance in dynamic scenarios involving actual drone flight and moving parcels.

6. Conclusions

This paper described the development and experimental assessment of an autonomous grasping system designed for integration with aerial vehicles in parcel delivery and exchange operations. The implementation involved constructing an operational gripper prototype equipped with angular position sensors and force detection capabilities, alongside a pose estimation method for packages, with all components coordinated through a Hybrid Model Predictive Controller that determines optimal vehicle trajectories and gripper actuation commands. Individual subsystems underwent successful testing, and the complete integrated system was validated through both simulated and physical experiments, where package pose was determined via the vision-based estimation algorithm using the onboard camera.

Extensions to agile maneuvers and multi-vehicle scenarios within the hybrid modeling framework represent a potential avenue for further research, as do alternative approaches for hybrid MPC formulation and implementation, given that the HYSDEL formalism has limitations in representing complex nonlinear hybrid dynamics with sufficient fidelity. Additionally, future research could focus on establishing theoretical stability foundations to ensure robust performance across a wider range of operating conditions as well as to use robust MPC techniques for disturbance rejection.

While the presented results provide an initial validation of the proposed approach, further research is necessary to advance toward a fully operational aerial grasping solution. This line of work could conduct extensive experimental trials in real-world scenarios with onboard computation, evaluating the system's performance under varying environmental conditions and with different parcel types to validate its robustness and adaptability, particularly to moving parcels of increasing agile motions.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/s26020653/s1. Video S1: experimental validation of the proposed techniques.

Author Contributions: B.J.G.: Conceptualization, Supervision, Investigation, Writing—Original draft; F.A.: Software, Validation, Investigation, Writing—Original draft; P.O.: Supervision, Writing—Reviewing and Editing; R.C.: Supervision, Writing—Reviewing and Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by FCT project REPLACE (PTDC/EEIAUT/32107/2017) which includes Lisboa 2020 and PIDDAC funds, project CAPTURE (PTDC/EEI-AUT/1732/2020), as well as projects CTS (UIDB/00066/2020), LARSYS (UIDB/50009/2020), and LAETA (UIDB/EMS/50022/2020).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

https://doi.org/10.3390/s26020653

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Guerreiro, B. REPLACE Project: Fast Delivery in Urban Environments Using Drone RElays: PLAnning, Control, and Estimation, 2022. Available online: http://replace.isr.tecnico.ulisboa.pt/ (accessed on 12 January 2026).
2. Pinto, J.; Guerreiro, B.J.; Cunha, R. Planning parcel relay manoeuvres for quadrotors. In Proceedings of the 2021 International Conference on Unmanned Aircraft Systems (ICUAS), Athens, Greece, 15–18 June 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 137–145. [CrossRef]
3. Pinto, J.; Guerreiro, B.J.; Cunha, R. Planning Aggressive Drone Manoeuvres: A Geometric Backwards Integration Approach. *J. Intell. Robot. Syst.* **2025**, *111*, 16. [CrossRef]
4. Pio Belfiore, N.; Pennestrì, E. An atlas of linkage-type robotic grippers. *Mech. Mach. Theory* **1997**, *32*, 811–833. [CrossRef]
5. Monkman, G.J. *Robot Grippers*; Wiley Online Library: Hoboken, NJ, USA, 2007.
6. Pounds, P.E.; Bersak, D.R.; Dollar, A.M. Grasping from the air: Hovering capture and load stability. In Proceedings of the IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 2491–2498. [CrossRef]
7. Mellinger, D.; Lindsey, Q.; Shomin, M.; Kumar, V. Design, modeling, estimation and control for aerial grasping and manipulation. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011. [CrossRef]
8. Mellinger, D.; Shomin, M.; Kumar, V. Control of quadrotors for robust perching and landing. In Proceedings of the International Powered Lift Conference, Philadelphia, PA, USA, 5–7 October 2010; pp. 205–225.
9. Ruggiero, F.; Lippiello, V.; Ollero, A. Aerial manipulation: A literature review. *IEEE Robot. Autom. Lett.* **2018**, *3*, 1957–1964. [CrossRef]
10. Ollero, A.; Tognon, M.; Suarez, A.; Lee, D.; Franchi, A. Past, Present, and Future of Aerial Robotic Manipulators. *IEEE Trans. Robot.* **2022**, *38*, 626–645. [CrossRef]
11. Meng, J.; Buzzatto, J.; Liu, Y.; Liarokapis, M. On Aerial Robots with Grasping and Perching Capabilities: A Comprehensive Review. *Front. Robot. AI* **2022**, *8*, 739173. [CrossRef] [PubMed]
12. Gentili, L.; Naldi, R.; Marconi, L. Modeling and control of VTOL UAVs interacting with the environment. In Proceedings of the IEEE Conference on Decision and Control, Cancun, Mexico, 9–11 December 2008; pp. 1231–1236. [CrossRef]
13. Mohta, K.; Kumar, V.; Daniilidis, K. Vision-based control of a quadrotor for perching on lines. In Proceedings of the IEEE International Conference on Robotics and Automation, Hong Kong, China, 31 May–7 June 2014; pp. 3130–3136. [CrossRef]
14. Garrido-Jurado, S.; Muñoz-Salinas, R.; Madrid-Cuevas, F.J.; Marín-Jiménez, M.J. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognit.* **2014**, *47*, 2280–2292. [CrossRef]
15. Suarez, A.; Jimenez-Cano, A.E.; Vega, V.M.; Heredia, G.; Rodriguez-Castaño, A.; Ollero, A. Design of a lightweight dual arm system for aerial manipulation. *Mechatronics* **2018**, *50*, 30–44. [CrossRef]
16. Hingston, L.; Mace, J.; Buzzatto, J.; Liarokapis, M. Reconfigurable, adaptive, lightweight grasping mechanisms for aerial robotic platforms. In Proceedings of the 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Abu Dhabi, United Arab Emirates, 4–6 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 169–175. [CrossRef]
17. Bo, V.; Franco, L.; Turco, E.; Pozzi, M.; Malvezzi, M.; Prattichizzo, D.; Salvietti, G. Design and Control of Soft-Rigid Grippers for Food Handling. In Proceedings of the ICRA2024 Workshop on Cooking Robotics: Perception and Motion Planning, Yokohama, Japan, 13–17 May 2024. Available online: https://openreview.net/forum?id=EVJbAtEUSf (accessed on 12 January 2026).
18. Hu, H.; Xie, Z.; Liu, Y.; Liu, Y.; Yang, G.; Xia, J.; Liu, H. A Variable Stiffness Actuation Based Robotic Hand Designed for Interactions. *IEEE/ASME Trans. Mechatronics* **2024**, *29*, 249–259. [CrossRef]
19. Saunders, J.; Saeedi, S.; Li, W. Autonomous aerial robotics for package delivery: A technical review. *J. Field Robot.* **2024**, *41*, 3–49. [CrossRef]
20. Witsenhausen, H.S. A Class of Hybrid-State Continuous-Time Dynamic Systems. *IEEE Trans. Autom. Control* **1966**, *11*, 161–167. [CrossRef]
21. Alur, R.; Courcoubetis, C.; Henzinger, T.A.; Ho, P.H. Hybrid automata: An algorithmic approach to the specification and verification of hybrid systems. In *Proceedings of the International Hybrid Systems Workshop*; Springer: Berlin/Heidelberg, Germany, 1991; pp. 209–229. [CrossRef]
22. Henzinger, T.A. The Theory of Hybrid Automata. In *Verification of Digital and Hybrid Systems*; Springer: Berlin/Heidelberg, Germany, 2000; Volume 300, pp. 265–292. [CrossRef]
23. Chai, J.; Casau, P.; Sanfelice, R.G. Analysis and design of event-triggered control algorithms using hybrid systems tools. *Int. J. Robust Nonlinear Control* **2020**, *30*, 5936–5965. [CrossRef]

https://doi.org/10.3390/s26020653

24. Bemporad, A.; Morari, M. Control of systems integrating logic, dynamics, and constraints. *Automatica* **1999**, *35*, 407–427. [CrossRef]
25. Ma, R.R.; Odhner, L.U.; Dollar, A.M. A modular, open-source 3D printed underactuated hand. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 2737–2743. [CrossRef]
26. Kruse, L.; Bradley, J. A hybrid, actively compliant manipulator/gripper for aerial manipulation with a multicopter. In Proceedings of the 2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Philadelphia, PA, USA, 6–8 August 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–8. [CrossRef]
27. Zhang, H.; Sun, J.; Zhao, J. Compliant Bistable Gripper for Aerial Perching and Grasping. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 1248–1253. [CrossRef]
28. Romero-Ramirez, F.J.; Muñoz-Salinas, R.; Medina-Carnicer, R. Speeded up detection of squared fiducial markers. *Image Vis. Comput.* **2018**, *76*, 38–47. [CrossRef]
29. Mellinger, D. Trajectory Generation and Control for Quadrotors. Ph.D. Thesis, University of Pennsylvania, Philadelphia, PA, USA, 2012.
30. Torrisi, F.D.; Bemporad, A. HYSDEL—A tool for generating computational hybrid models. *IEEE Trans. Control Syst. Technol.* **2004**, *12*, 235–249. [CrossRef]
31. Bemporad, A. Hybrid Toolbox-User's Guide. 2004. Available online: http://cse.lab.imtlucca.it/~bemporad/hybrid/toolbox (accessed on 12 January 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

https://doi.org/10.3390/s26020653

sensors

MDPI

Article

Pose Stabilization Control for Base of Combined System Using Feedforward Compensation PD Control During Target Satellite Transposition

Zhonghua Hu [1,2,*], Jinlong Yang [1], Wenfu Xu [2], Hengtai Chen [1], Longkun Xu [1] and Deshan Meng [3,*]

[1] School of Mechanical & Automotive Engineering, Liaocheng University, Liaocheng 252000, China; yjl_lcu@163.com (J.Y.); cht_lcu@163.com (H.C.); xulongkun@lcu.edu.cn (L.X.)
[2] School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, China; wfxu@hit.edu.cn
[3] School of Aeronautics and Astronautics, Shenzhen Campus, Sun Yat-sen University, Shenzhen 518107, China
* Correspondence: huzhonghua@lcu.edu.cn (Z.H.); mengdsh3@mail.sysu.edu.cn (D.M.)

Abstract

During the transposition of a target satellite, dynamic coupling between the target satellite, the manipulators, and the base frequently leads to disturbances in the base's attitude. To deal with the issue, this paper proposed a pose stabilization method for the base of the post-capture combined system using the feedforward compensation PD control. Firstly, the mission sequence for repositioning a target satellite using a discrete-serpentine heterogeneous dual-arm space robot (DSHDASR) was analyzed. The dynamics model of the combined system, composed of the DSHDASR and a target satellite, was established based on the Newton–Euler recursive formulation. Then, the pose stabilization method integrating dynamic feedforward compensation and PD control was developed to stabilize the base of the combined system. Finally, the mission of target satellite transposition was simulated through the co-simulation model. Compared with the traditional control algorithms, the position accuracy and attitude accuracy for the proposed method showed an overall improvement. The results demonstrated that the proposed method significantly reduced base pose errors under high-load and disturbed conditions.

Keywords: space robot–target satellite combined system; target satellite transposition; dynamics modeling; pose stabilization control

1. Introduction

Space robotic manipulators are essential for on-orbit servicing (OOS) [1,2], including debris removal, structure assembly, and satellite repair [3]. Based on the number of joints and link characteristics, space manipulators can be categorized into three types: discrete arms [4], serpentine arms [5], and continuum arms [6]. Unlike continuum arms, the discrete and serpentine manipulators present fewer challenges in terms of accurate modeling and precise control. Therefore, they are most widely adopted to carry out OOS. Furthermore, compared to a single discrete or serpentine arm, the DSHDASR combines the advantages of both types of arms, which ultimately overcomes specific limitations of individual manipulators [7]. Before conducting OOS, the space robotic arms installed on the DSHDASR should grasp the target satellite firstly, and then transpose it from the current location to the desired position for the subsequent operations [8]. Once the target satellite has been captured, the combined system is formed. However, during the target

satellite transposition, strong dynamic coupling between the target, the manipulators, and the base can disturb the base's pose (position and attitude). These disturbances not only reduce the tracking accuracy of manipulators but also bring potential risks to subsequent operations [9]. To validate space robotic control strategies under near-realistic microgravity conditions, various ground-based microgravity emulation platforms have been developed in recent years [10,11]. Therefore, suppressing these disturbances poses a great challenge for achieving precise tracking during target transposition, particularly under the condition of strong dynamic coupling.

To address this challenge, considerable research focused on pose stabilization control. Papadopoulos and Dubowsky [12] derived the dynamics model of space robotic manipulators using a Lagrangian framework, enabling coordinated control between the manipulator and the base. To achieve the desired attitude transposition, Yamada [13] proposed a joint path-planning method to achieve a desired change in attitude. By combining the Backstepping method and the radial basis function neural networks, Zhan et al. [14] designed an adaptive controller to ensure the trajectory tracking and base stabilization. Taking the end-effector as a Virtual Base, Zong et al. [15] established a new dynamics model of free-floating space manipulators to accomplish end-effector trajectory tracking and minimum disturbances to the base. Subsequently, Gong et al. [16] developed a dynamics model with parametric uncertainties to deal with the reactionless control of pose tracking. Based on the extended Jacobian model, Xu [17] designed a passivity adaptive controller to achieve the pose attitude adjustment and trajectory tracking. Considering the rigid body modes, Kuck et al. [18] developed a benchmark feedback controller to minimize trajectory tracking errors. While aforementioned methods can achieve base pose control, they do not account for the presence of a payload on the arm's end-effector.

A payload significantly intensifies the dynamic coupling between the manipulators and the base. This brings a significant challenge to base stabilization control. To deal with the problem, Raina et al. [19] proposed a unified framework for impact dynamics modeling and post-capture reactionless control to guarantee attitude stabilization of the base. Wang et al. [20] developed a coordinated control method for post-capture operations, enabling a redundant space robot to maintain the attitude stability of the combined system. It should be noted that in both the methods above, the payload-to-base mass ratio is under 10%. For high proportion payloads, Chang et al. [21] presented an adaptive Backstepping-based nonsingular fast integral terminal sliding mode control method for the attitude stabilization of space robots after capturing a tumbling target. Woodward et al. [22] designed a Lyapunov-based control law to drive the base's angular velocity attitude error asymptotically to zero. Based on dynamic coupling control with an iterative extended Kalman filter, Zhang et al. [23] proposed a rapid attitude stabilization strategy for highly robust base-attitude stabilization. Huang et al. [24] developed a robust attitude control method based on a disturbance observer and dynamic control allocation for the post-capture attitude control of the combined system, which ensured the prescribed attitude performance under uncertainties and external disturbances. Song et al. [25] applied a model-free control method for the attitude and orbital control of a post-capture combined spacecraft. Although these methods address base stabilization with large payloads on the end-effector, they do not consider the end-effector's trajectory tracking performance.

Recently, the rapid development of intelligent algorithms has led to their application in base stabilization control. Liu et al. [26] investigated a trajectory planning and coordination control method to minimize the target's attitude and disturbances to the base using the multi-objective particle swarm optimization (MOPSO) algorithm. Based on the deep Koopman operator-based modeling approach with model predictive control, Reference [27] achieved high-precision joint tracking and effective suppression of base attitude distur-

https://doi.org/10.3390/s26010206

bances. According to reinforcement learning, Wang et al. [28] and Dong et al. [29] proposed a multi-target trajectory optimization method and a sequential optimization approach, respectively, to control the base's dynamic response and stabilize its attitude. In addition, Ma et al. [30] applied deep reinforcement learning to restore a satellite's attitude, which is a task analogous to base attitude control.

As discussed above, although numerous methods for base pose stabilization have been proposed, each method has certain limitations. Some approaches established the relationship between base pose disturbance and manipulator motion to achieve stabilization control of the base, yet did not account for the case of a payload on the end-effector. Other methods addressed base stabilization with a payload, but they did not account for end-effector pose tracking. Furthermore, while some intelligent algorithms could effectively handle base pose stabilization, their computational demands often made them impractical for real-time control on spacecraft hardware. To tackle these issues, a pose stabilization method based on feedforward compensation PD control was proposed for the base of the combined system, while the DSHDASR was transposing the target satellite from the initial pose to the desired pose. Compared with the PD control method and Backstepping method, this method introduces the reaction force and torque acting on the base in the control law, effectively decreasing the strong dynamic coupling effects on the base. In addition, compared with existing approaches based on adaptive schemes or intelligent algorithms, the proposed method has advantages including a clear computational structure, low computational complexity, and high computational efficiency. This feature makes the proposed strategy more suitable for spacecraft platforms within limited onboard computational resources and offers favorable potential for real-time implementation.

The remainder of this paper is organized as follows. Section 2 analyzes the overall process of the target satellite transposition. In Section 3, the dynamics model of the combined system is established by the recursive Newton–Euler formulation. Based on the PD control with feedforward compensation, Section 4 proposes the pose stabilization method for the base during the target satellite transposition. In Section 5, the proposed method is verified through the comparison simulations including the proposed method, the traditional PD method, and the Backstepping method. A comprehensive discussion of the challenges and simplifications is given in Section 6. Based on the analysis of previous sections, the last section summarizes the work of this study.

2. Overall Task Description

After capture, the target satellite often must be moved to a suitable location for subsequent on-orbit operations. This movement process is known as the target satellite transposition, as shown in Figure 1. At the initial state, the target satellite has been captured by the DSHDASR. The DSHDASR is composed of a base, a non-offset discrete arm (namely Arm-a) and a segmented cable-driven hyper-redundant serpentine arm (namely Arm-b). Together, the DSHDASR and the target satellite form the combined system. During transposition, Arm-a serves as the primary manipulator for moving the target satellite from the initial pose to the final pose (i.e., desired position and attitude) along a predefined trajectory, while Arm-b is used as the auxiliary manipulator, providing visual monitoring of Arm-a's operation. Due to dynamic coupling between the base and the two manipulators, the base moves as Arm-a and Arm-b execute the coordinated transposition. As a result, the motion of the base alters the end-effector trajectories of both arms.

https://doi.org/10.3390/s26010206

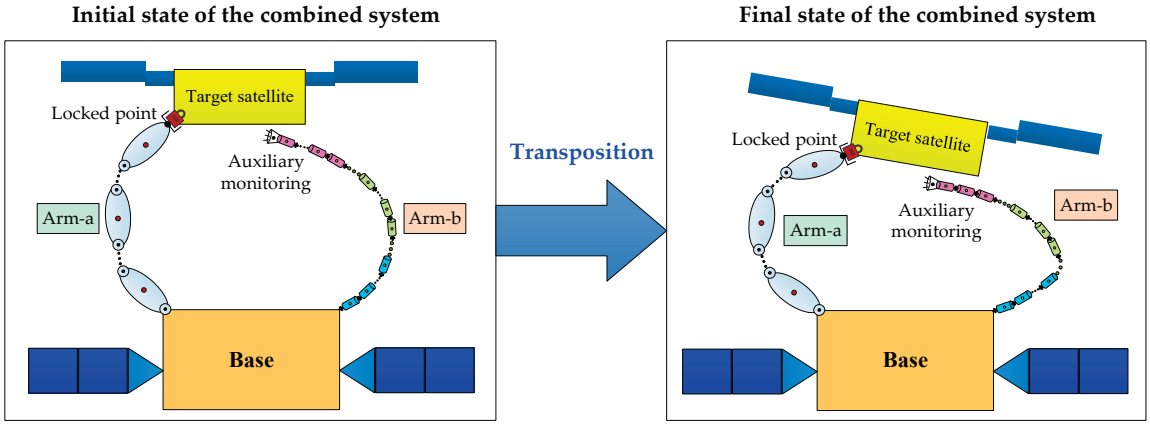


Figure 1. The illustration for target satellite transposition.

To ensure that the end-effectors of the two arms follow their predefined trajectories, it is necessary to maintain the base in a stable state. Therefore, this paper investigates control for stabilizing the base of the combined system during target satellite transposition. Some assumptions are given as follows.

(1) The mass and geometric parameters of the target satellite and the DSHDASR are assumed to be known and are considered to be exact values in this study.
(2) The base pose control can be achieved by thrusters and reaction wheels, whose installation positions and directions are known.
(3) The target satellite and the end-effector of the non-offset discrete arm remain in a securely locked state during the transposing process.
(4) Joint friction is neglected in the dynamic modeling and control design, and all joints are assumed to be ideal frictionless revolute joints.
(5) The structural flexibility of the links and joints for the DSHDASR, as well as mechanical backlash and other non-ideal effects, are not taken into account.

3. Dynamics Modeling of the Combined System

As shown in Figure 2, the combined system is formed after the target satellite is captured by the DSHDASR.

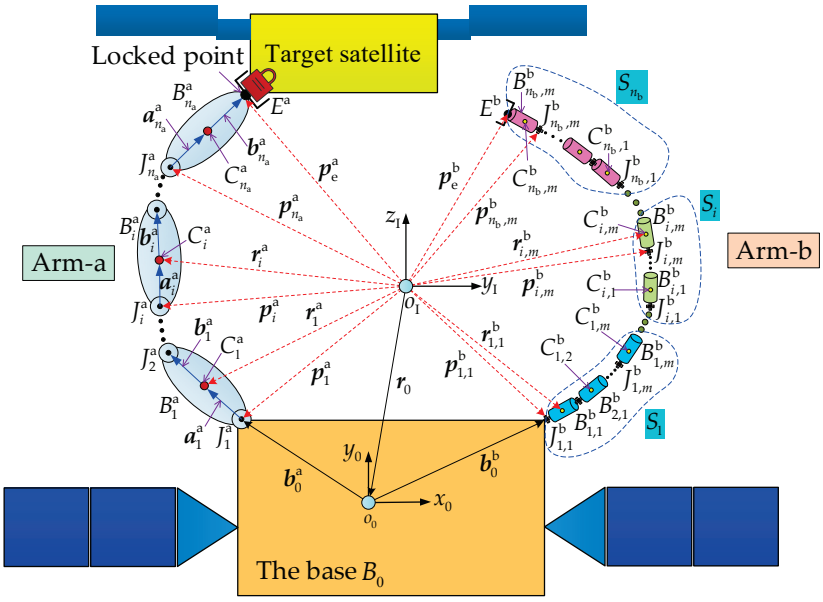


Figure 2. The generalized model of the combined system.

https://doi.org/10.3390/s26010206

As discussed above, the DSHDASR consists of a base, a non-offset discrete arm (Arm-a), and a segmented cable-driven serpentine arm (Arm-b). For Arm-a, each joint only has one rotational axis that is driven by a motor. The total number of joints is denoted by n_{a}. In contrast, Arm-b is composed of n_{b} segments with identical mechanical structures. Each segment has m associated universal joints, which are driven by several cables. An associated universal joint possesses a Pitch-axis and a Yaw-axis. For each segment, all associated universal joints are hybrid active–passive joints [31]. Specifically, the first universal joint in a segment is active and follows planned values, while the remaining joints are passive and move synchronously with it. $\{x_{\mathrm{I}}, y_{\mathrm{I}}, z_{\mathrm{I}}\}$ and $\{x_0, y_0, z_0\}$ are the inertial frame and the frame fixed on the base center of mass, respectively. Unless otherwise stated, all the variables are represented in the inertial frame.

The Newton–Euler formulation is a well-known method that offers advantages in computational efficiency and numerical stability, making it suitable for real-time control. As a result, the Newton–Euler formulation is adopted to establish the dynamic model of the combined system.

3.1. Forward Recursion for Velocities and Accelerations Computation

As shown in Figure 2, J_i^{a} and B_i^{a} represent the ith joint and link of Arm-a. C_i^{a} is the center of mass of B_i^{a} and its position vector is denoted by $\boldsymbol{r}_i^{\mathrm{a}}$. $\boldsymbol{a}_i^{\mathrm{a}}$ represents the vector from the joint J_i^{a} to the center C_i^{a} and $\boldsymbol{b}_i^{\mathrm{a}}$ is the vector from the center C_i^{a} to the joint J_{i+1}^{a}. For the ith joint of Arm-a, the position vector of J_i^{a} is obtained as follows:

$$\boldsymbol{p}_i^{\mathrm{a}} = \boldsymbol{r}_0 + \boldsymbol{b}_0^{\mathrm{a}} + \sum_{j=1}^{i-1}\left(\boldsymbol{a}_j^{\mathrm{a}} + \boldsymbol{b}_j^{\mathrm{a}}\right) \in \mathbf{R}^{3\times 1} \tag{1}$$

where $\boldsymbol{r}_0$ is the position vector of the center of mass for the base. $\boldsymbol{b}_0^{\mathrm{a}}$ denotes the vector from the base's center of mass to the first joint of Arm-a. According to the principle of relative motion, the angular velocity and the linear velocity of J_i^{a} expressed in the ith joint frame are obtained as follows:

$$\begin{cases} {}^{i}\boldsymbol{\omega}_i^{\mathrm{a}} = {}^{i}\boldsymbol{R}_{i-1}^{\mathrm{a}}\,{}^{i-1}\boldsymbol{\omega}_{i-1}^{\mathrm{a}} + {}^{i}\boldsymbol{k}_i^{\mathrm{a}}\cdot\dot{q}_i^{\mathrm{a}} \\ {}^{i}\boldsymbol{v}_i^{\mathrm{a}} = {}^{i}\boldsymbol{R}_{i-1}^{\mathrm{a}}\left[{}^{i-1}\boldsymbol{v}_{i-1}^{\mathrm{a}} + {}^{i-1}\boldsymbol{\omega}_{i-1}^{\mathrm{a}} \times \left({}^{i-1}\boldsymbol{a}_{i-1}^{\mathrm{a}} + {}^{i-1}\boldsymbol{b}_{i-1}^{\mathrm{a}}\right)\right] \end{cases} \tag{2}$$

where ${}^{i}\boldsymbol{R}_{i-1}^{\mathrm{a}}$ stands for the transformation matrix from the $(i-1)$th joint frame to the ith joint frame for Arm-a. Regarding the other parameters in Equation (2), the left superscript means that the vector is described in the frame indicated by the superscript, i.e., ${}^{i-1}\boldsymbol{v}_{i-1}^{\mathrm{a}}$, ${}^{i-1}\boldsymbol{\omega}_{i-1}^{\mathrm{a}}$, ${}^{i-1}\boldsymbol{a}_{i-1}^{\mathrm{a}}$, and ${}^{i-1}\boldsymbol{b}_{i-1}^{\mathrm{a}}$ are all expressed in the $(i-1)$th joint frame of Arm-a. The subsequent notations follow the same convention. ${}^{i}\boldsymbol{k}_i^{\mathrm{a}}$ and $\dot{q}_j^{\mathrm{a}}$ refer to the rotation axis and the angular velocity magnitude of the ith joint for Arm-a, respectively.

Differentiating Equation (2), the angular acceleration and linear acceleration of J_i^{a} are derived as follows:

$$\begin{cases} {}^{i}\dot{\boldsymbol{\omega}}_i^{\mathrm{a}} = {}^{i}\boldsymbol{R}_{i-1}^{\mathrm{a}}\,{}^{i-1}\dot{\boldsymbol{\omega}}_{i-1}^{\mathrm{a}} + {}^{i}\boldsymbol{R}_{i-1}^{\mathrm{a}}\,{}^{i-1}\boldsymbol{\omega}_{i-1}^{\mathrm{a}} \times \left({}^{i}\boldsymbol{k}_i^{\mathrm{a}}\cdot\dot{q}_i^{\mathrm{a}}\right) + {}^{i}\boldsymbol{k}_i^{\mathrm{a}}\cdot\ddot{q}_i^{\mathrm{a}} \\ {}^{i}\dot{\boldsymbol{v}}_i^{\mathrm{a}} = {}^{i}\boldsymbol{R}_{i-1}^{\mathrm{a}}\left({}^{i-1}\dot{\boldsymbol{v}}_{i-1}^{\mathrm{a}} + {}^{i-1}\dot{\boldsymbol{\omega}}_{i-1}^{\mathrm{a}} \times \left({}^{i-1}\boldsymbol{a}_{i-1}^{\mathrm{a}} + {}^{i-1}\boldsymbol{b}_{i-1}^{\mathrm{a}}\right) + {}^{i-1}\boldsymbol{\omega}_{i-1}^{\mathrm{a}} \times \left({}^{i-1}\boldsymbol{\omega}_{i-1}^{\mathrm{a}} \times \left({}^{i-1}\boldsymbol{a}_{i-1}^{\mathrm{a}} + {}^{i-1}\boldsymbol{b}_{i-1}^{\mathrm{a}}\right)\right)\right) \end{cases} \tag{3}$$

where $\ddot{q}_i^{\mathrm{a}}$ denotes the angular acceleration magnitude of J_i^{a}.

Based on Equation (3), the angular and linear accelerations of C_i^{a} are obtained as follows.

$$\begin{cases} {}^{i}\dot{\boldsymbol{\omega}}_{\mathrm{C}i}^{\mathrm{a}} = {}^{i}\dot{\boldsymbol{\omega}}_i^{\mathrm{a}} = {}^{i}\boldsymbol{R}_{i-1}^{\mathrm{a}}\,{}^{i-1}\dot{\boldsymbol{\omega}}_{i-1}^{\mathrm{a}} + {}^{i}\boldsymbol{R}_{i-1}^{\mathrm{a}}\,{}^{i-1}\boldsymbol{\omega}_{i-1}^{\mathrm{a}} \times \left({}^{i}\boldsymbol{k}_i^{\mathrm{a}}\cdot\dot{q}_i^{\mathrm{a}}\right) + {}^{i}\boldsymbol{k}_i^{\mathrm{a}}\cdot\ddot{q}_i^{\mathrm{a}} \\ {}^{i}\dot{\boldsymbol{v}}_{\mathrm{C}i}^{\mathrm{a}} = {}^{i}\dot{\boldsymbol{v}}_i^{\mathrm{a}} + {}^{i}\dot{\boldsymbol{\omega}}_{\mathrm{C}i}^{\mathrm{a}} \times {}^{i}\boldsymbol{a}_i^{\mathrm{a}} + {}^{i}\boldsymbol{\omega}_i^{\mathrm{a}} \times \left({}^{i}\boldsymbol{\omega}_i^{\mathrm{a}} \times {}^{i}\boldsymbol{a}_i^{\mathrm{a}}\right) \end{cases} \tag{4}$$

https://doi.org/10.3390/s26010206

For Arm-b, $J_{i,j}^{\mathrm{b}}$ and $B_{i,j}^{\mathrm{b}}$ are the jth joint and the jth link in the ith segment. $C_{i,j}^{\mathrm{b}}$ refers the center of mass of $B_{i,j}^{\mathrm{b}}$ and its position vector is named by $\boldsymbol{r}_{i,j}^{\mathrm{b}}$. The position vector of the jth joint in the ith segment of Arm-b can be calculated as follows:

$$\boldsymbol{p}_{i,j}^{\mathrm{b}} = \boldsymbol{r}_0 + \boldsymbol{b}_0^{\mathrm{b}} + \sum_{h=1}^{i-1}\sum_{n=1}^{m}\left(\boldsymbol{a}_{h,n}^{\mathrm{b}} + \boldsymbol{b}_{h,n}^{\mathrm{b}}\right) + \sum_{n=1}^{j-1}\left(\boldsymbol{a}_{i,n}^{\mathrm{b}} + \boldsymbol{b}_{i,n}^{\mathrm{b}}\right) \in \mathbf{R}^{3\times 1} \tag{5}$$

where $\boldsymbol{b}_0^{\mathrm{b}}$ is the vector from the base's center of mass to the first joint of Arm-b. $\boldsymbol{a}_{i,j}^{\mathrm{b}}$ represents the vector from the joint $J_{i,j}^{\mathrm{b}}$ to the center $C_{i,j}^{\mathrm{b}}$. $\boldsymbol{b}_{i,j}^{\mathrm{b}}$ is the vector from the center $C_{i,j}^{\mathrm{b}}$ to the joint $J_{i,j+1}^{\mathrm{b}}$.

For the jth joint within the ith segment of Arm-b, its two rotational axes are denoted as $\boldsymbol{k}_{i,\,j,1}^{\mathrm{b}}$ and $\boldsymbol{k}_{i,\,j,2}^{\mathrm{b}}$, and the corresponding angular velocities are $\dot{q}_{i,\,j,1}^{\mathrm{b}}$ and $\dot{q}_{i,\,j,2}^{\mathrm{b}}$. Since the universal joints in the same segment are hybrid active and passive, the movement law of universal joints in the ith segment can be expressed as follows.

$$\begin{cases} \dot{q}_{i,j,1}^{\mathrm{b}} = \dot{q}_{i,j+1,2}^{\mathrm{b}} \\ \dot{q}_{i,j,2}^{\mathrm{b}} = \dot{q}_{i,j+1,1}^{\mathrm{b}} \end{cases} \quad i = 1,2,\ldots,n_{\mathrm{b}}; \quad j = 1,2,\ldots,m-1 \tag{6}$$

Based on the kinematics of relative motion, the angular velocity and the linear velocity of the jth joint in the ith segment are obtained as follows:

$$\begin{cases} {}^{i,j,2}\boldsymbol{\omega}_{i,j}^{\mathrm{b}} = {}^{i,j,2}\boldsymbol{R}_{i,j-1,2}^{\mathrm{b}}\,{}^{i,j-1,2}\boldsymbol{\omega}_{i,j-1}^{\mathrm{b}} + {}^{i,j,2}\boldsymbol{R}_{i,j,1}^{\mathrm{b}}\left({}^{i,j,1}\boldsymbol{k}_{i,j,1}^{\mathrm{b}}\cdot\dot{q}_{i,j,1}^{\mathrm{b}}\right) + {}^{i,j,2}\boldsymbol{k}_{i,j,2}^{\mathrm{b}}\cdot\dot{q}_{i,j,2}^{\mathrm{b}} \\ {}^{i,j,2}\boldsymbol{v}_{i,j}^{\mathrm{b}} = {}^{i,j,2}\boldsymbol{R}_{i,j-1,2}^{\mathrm{b}}\left({}^{i,j-1,2}\boldsymbol{v}_{i,j-1}^{\mathrm{b}} + {}^{i,j-1,2}\boldsymbol{\omega}_{i,j-1}^{\mathrm{b}} \times \left({}^{i,j-1,2}\boldsymbol{a}_{i,j-1}^{\mathrm{b}} + {}^{i,j-1,2}\boldsymbol{b}_{i,j-1}^{\mathrm{b}}\right)\right) \end{cases} \tag{7}$$

where ${}^{i,j,2}\boldsymbol{R}_{i,j-1,2}^{\mathrm{b}}$ is the transformation matrix from the second rotational axis of the $(j-1)$th joint to the jth joint in the ith segment. With respect to the remaining parameters, the left superscript represents that the vector is expressed in the frame attached to the second rotational axis of the $(j-1)$th joint in the ith segment. ${}^{i,j,1}\boldsymbol{k}_{i,j,1}^{\mathrm{b}}$ and ${}^{i,j,2}\boldsymbol{k}_{i,j,2}^{\mathrm{b}}$ are the first and second rotational axes of the jth joint. Their corresponding angular velocities are $\dot{q}_{i,j,1}^{\mathrm{b}}$ and $\dot{q}_{i,j,2}^{\mathrm{b}}$, respectively.

Differentiating Equation (7), the angular and linear accelerations of $J_{i,j}^{\mathrm{b}}$ are computed as follows.

$$\begin{cases} \begin{aligned} {}^{i,j,2}\dot{\boldsymbol{\omega}}_{i,j}^{\mathrm{b}} = {} & \left({}^{i,j,2}\boldsymbol{R}_{i,j-1,2}^{\mathrm{b}}\,{}^{i,j-1,2}\boldsymbol{\omega}_{i,j-1}^{\mathrm{b}}\right) \times \left({}^{i,j,2}\boldsymbol{R}_{i,j,1}^{\mathrm{b}}\,{}^{i,j,1}\boldsymbol{z}_{i,j,1}^{\mathrm{b}}\cdot\dot{q}_{i,j,1}^{\mathrm{b}} + {}^{i,j,2}\boldsymbol{z}_{i,j,2}^{\mathrm{b}}\cdot\dot{q}_{i,j,2}^{\mathrm{b}}\right) \\ & + \left({}^{i,j,2}\boldsymbol{R}_{i,j,1}^{\mathrm{b}}\,{}^{i,j,1}\boldsymbol{z}_{i,j,1}^{\mathrm{b}}\cdot\dot{q}_{i,j,1}^{\mathrm{b}}\right) \times \left({}^{i,j,2}\boldsymbol{z}_{i,j,2}^{\mathrm{b}}\cdot\dot{q}_{i,j,2}^{\mathrm{b}}\right) + {}^{i,j,2}\boldsymbol{R}_{i,j,1}^{\mathrm{b}}\,{}^{i,j,1}\boldsymbol{z}_{i,j,1}^{\mathrm{b}}\cdot\ddot{q}_{i,j,1}^{\mathrm{b}} \\ & + {}^{i,j,2}\boldsymbol{z}_{i,j,2}^{\mathrm{b}}\cdot\ddot{q}_{i,j,2}^{\mathrm{b}} + {}^{i,j,2}\boldsymbol{R}_{i,j-1,2}^{\mathrm{b}}\,{}^{i,j-1,2}\dot{\boldsymbol{\omega}}_{i,j-1}^{\mathrm{b}} \end{aligned} \\ {}^{i,j,2}\dot{\boldsymbol{v}}_{i,j}^{\mathrm{b}} = {}^{i,j,2}\boldsymbol{R}_{i,j-1,2}^{\mathrm{b}}\cdot\begin{pmatrix} {}^{i,j-1,2}\dot{\boldsymbol{v}}_{i,j-1}^{\mathrm{b}} + {}^{i,j-1,2}\dot{\boldsymbol{\omega}}_{i,j-1}^{\mathrm{b}} \times \left({}^{i,j-1,2}\boldsymbol{a}_{i,j-1}^{\mathrm{b}} + {}^{i,j-1,2}\boldsymbol{b}_{i,j-1}^{\mathrm{b}}\right) \\ + {}^{i,j-1,2}\boldsymbol{\omega}_{i,j-1}^{\mathrm{b}} \times \left({}^{i,j-1,2}\boldsymbol{\omega}_{i,j-1}^{\mathrm{b}} \times \left({}^{i,j-1,2}\boldsymbol{a}_{i,j-1}^{\mathrm{b}} + {}^{i,j-1,2}\boldsymbol{b}_{i,j-1}^{\mathrm{b}}\right)\right) \end{pmatrix} \end{cases} \tag{8}$$

According to Equation (8), the angular acceleration and linear acceleration of $C_{i,j}^{\mathrm{b}}$ can be obtained as follows.

$$\begin{cases} {}^{i,j,2}\dot{\boldsymbol{\omega}}_{\mathrm{C}i,j}^{\mathrm{b}} = {}^{i,j,2}\dot{\boldsymbol{\omega}}_{i,j}^{\mathrm{b}} \\ {}^{i,j,2}\dot{\boldsymbol{v}}_{\mathrm{C}i,j}^{\mathrm{b}} = {}^{i,j,2}\dot{\boldsymbol{v}}_{i,j}^{\mathrm{b}} + {}^{i,j,2}\dot{\boldsymbol{\omega}}_{i,j}^{\mathrm{b}} \times {}^{i,j,2}\boldsymbol{a}_{i,j}^{\mathrm{b}} + {}^{i,j,2}\boldsymbol{\omega}_{i,j}^{\mathrm{b}} \times \left({}^{i,j,2}\boldsymbol{\omega}_{i,j}^{\mathrm{b}} \times {}^{i,j,2}\boldsymbol{a}_{i,j}^{\mathrm{b}}\right) \end{cases} \tag{9}$$

3.2. Backward Recursion for Joint Forces and Torques Calculation

The base's motion including angular velocity, angular acceleration, linear velocity, and linear acceleration can be obtained from measurement by gyroscopes, accelerometers, or other sensors. In other words, ${}^{0}\boldsymbol{\omega}_0^{\mathrm{a}}$, ${}^{0}\dot{\boldsymbol{\omega}}_0^{\mathrm{a}}$, ${}^{0}\boldsymbol{v}_0^{\mathrm{a}}$, and ${}^{0}\dot{\boldsymbol{v}}_0^{\mathrm{a}}$ are known. On the basis of

https://doi.org/10.3390/s26010206

Equations (4) and (9), the angular and linear accelerations of each link's center of mass for Arm-a and Arm-b can be determined.

For both arms, joint forces and torques are calculated in reverse order, propagating from the end-effector back toward the base. For Arm-a, the dynamic relationship of the *i*th link is shown in Figure 3.

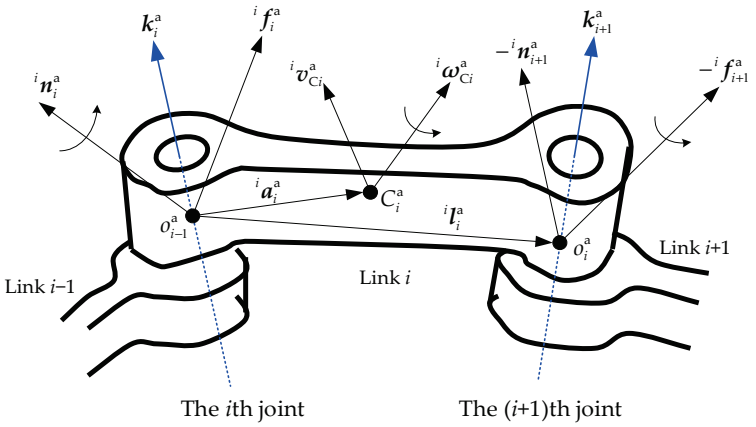


Figure 3. The force analysis for the *i*th link of Arm-a.

The gravitational force acting on each link of Arm-a can be considered as zero, due to the microgravity environment where the arm is. For the *i*th link of Arm-a, the force and torque balance equations are derived as follows.

$$\begin{cases} {}^i f_{Ci}^a - \left({}^i f_i^a - {}^i f_{i+1}^a\right) = \mathbf{0} \\ {}^i n_{Ci}^a - \left({}^i n_i^a - {}^i n_{i+1}^a + {}^i f_i^a \times {}^i a_i^a - {}^i f_{i+1}^a \times \left({}^i a_i^a - {}^i l_i^a\right)\right) = \mathbf{0} \end{cases} \tag{10}$$

where ${}^i f_{Ci}^a$ and ${}^i n_{Ci}^a$ are the inertia force and torque of the *i*th link. ${}^i f_i^a$ and ${}^i n_i^a$ refer to the force and torque exerted by the $(i-1)$th link on the *i*th link. ${}^i f_{i+1}^a$ and ${}^i n_{i+1}^a$ are the force and torque exerted by the *i*th link on the $(i+1)$th link. ${}^i l_i^a$ is the position vector from the *i*th joint to the $(i+1)$th joint. They are all expressed in the frame fixed on the *i*th joint.

According to Newton's laws and Euler's equations of motion, the inertia force and torque of the *i*th link can be obtained as follows:

$$\begin{cases} {}^i f_{Ci}^a = m_i^a \cdot {}^i \dot{v}_{Ci}^a \\ {}^i n_{Ci}^a = {}^i I_i^a \cdot {}^i \dot{\omega}_{Ci}^a + {}^i \omega_{Ci}^a \times \left({}^i I_i^a \cdot {}^i \omega_{Ci}^a\right) \end{cases} \tag{11}$$

where m_i^a and ${}^i I_i^a$ are the mass and inertia tensor of the *i*th link.

Substituting Equation (11) into Equation (10), the relationship between the motion and the applied force and torque for the *i*th link can be obtained as follows.

$$\begin{cases} {}^i f_i^a - {}^i f_{i+1}^a = m_i^a \cdot {}^i \dot{v}_{Ci}^a \\ {}^i n_i^a - {}^i n_{i+1}^a + {}^i f_i^a \times {}^i a_i^a - {}^i f_{i+1}^a \times \left({}^i a_i^a - {}^i l_i^a\right) = {}^i I_i^a \cdot {}^i \dot{\omega}_{Ci}^a + {}^i \omega_{Ci}^a \times \left({}^i I_i^a \cdot {}^i \omega_{Ci}^a\right) \end{cases} \tag{12}$$

Here, $\dot{v}_{Ci}^a$ and ${}^i \dot{\omega}_{Ci}^a$ have been calculated based on Equation (4). Substituting Equation (4) into Equation (12), the following equations hold.

$$\begin{cases} {}^i f_i^a = {}^i R_{i+1}^a \, {}^{i+1} f_{i+1}^a + m_i^a \left({}^i \dot{v}_i^a + {}^i \dot{\omega}_{Ci}^a \times {}^i r_{Ci}^a + {}^i \omega_i^a \times \left({}^i \omega_i^a \times {}^i r_{Ci}^a\right)\right) \\ {}^i n_i^a = {}^i R_{i+1}^a \, {}^{i+1} n_{i+1}^a - {}^i f_i^a \times {}^i a_i^a + {}^i f_{i+1}^a \times \left({}^i a_i^a - {}^i l_i^a\right) + {}^i I_i^a \cdot {}^i \dot{\omega}_{Ci}^a + {}^i \omega_{Ci}^a \times \left({}^i I_i^a \cdot {}^i \omega_{Ci}^a\right) \end{cases} \tag{13}$$

Notably, the issue of the virtual link $n_a + 1$ has occurred during the backward recursion for calculating forces and torques. To address this, the external environment of Arm-

https://doi.org/10.3390/s26010206

a is treated as the link $n_a + 1$. If there is no contact between the end-effector and the external environment, the external force and torque at the end-effector are set to zero, i.e., ${}^{n_a+1}\boldsymbol{f}^{a}_{n_a+1} = \mathbf{0}$ and ${}^{n_a+1}\boldsymbol{n}^{a}_{n_a+1} = \mathbf{0}$.

The recursive process for calculating the applied forces and torques on each joint of Arm-b is given as follows. The dynamic relationship of the *j*th link in the *i*th segment is shown in Figure 4.

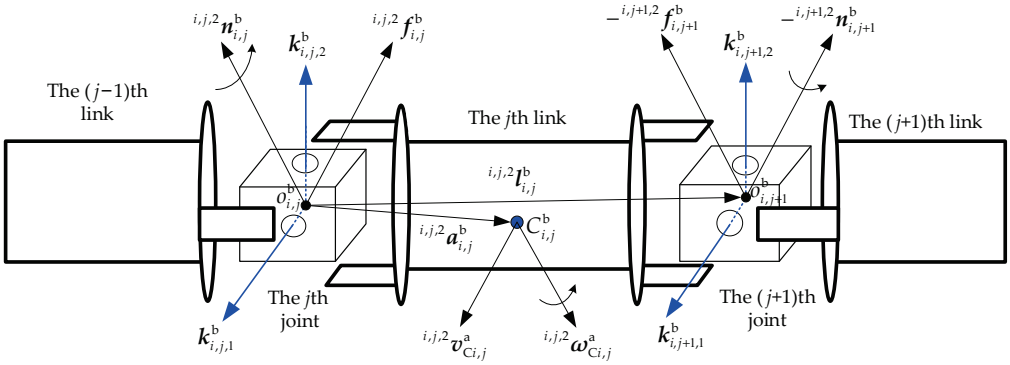


Figure 4. The force analysis for the *i*th link of Arm-b.

The force and torque balance equations of the *j*th link in the *i*th segment are obtained as follows:

$$\begin{cases} {}^{i,j,2}\boldsymbol{f}^{b}_{i,j} = {}^{i,j+1,2}\boldsymbol{f}^{b}_{i,j+1} + m^{b}_{i,j}\cdot{}^{i,j,2}\dot{\boldsymbol{v}}^{b}_{Ci,j} \\ {}^{i,j,2}\boldsymbol{n}^{b}_{i,j} = {}^{i,j+1,2}\boldsymbol{n}^{b}_{i,j+1} - {}^{i,j,2}\boldsymbol{f}^{b}_{i,j} \times {}^{i,j,2}\boldsymbol{a}^{b}_{i,j} + {}^{i,j+1,2}\boldsymbol{f}^{b}_{i,j+1} \times \left({}^{i,j,2}\boldsymbol{a}^{b}_{i,j} - {}^{i,j,2}\boldsymbol{l}^{b}_{i,j}\right) \\ \quad + {}^{i,j,2}\boldsymbol{I}^{b}_{i,j}\cdot{}^{i,j,2}\dot{\boldsymbol{\omega}}^{b}_{Ci,j} + {}^{i,j,2}\boldsymbol{\omega}^{b}_{Ci,j} \times \left({}^{i,j,2}\boldsymbol{I}^{b}_{i,j}\cdot{}^{i,j,2}\boldsymbol{\omega}^{b}_{Ci,j}\right) \end{cases} \tag{14}$$

where ${}^{i,j,2}\boldsymbol{f}^{b}_{i,j}$ and ${}^{i,j,2}\boldsymbol{n}^{b}_{i,j}$ denote the force and torque exerted by the $(j-1)$th link on the *j*th link. ${}^{i}\boldsymbol{f}^{a}_{i+1}$ and ${}^{i}\boldsymbol{n}^{a}_{i+1}$ are the force and torque exerted by the *j*th link on the $(j+1)$th link. ${}^{i,j,2}\boldsymbol{l}^{b}_{i,j}$ is the position vector from the *j*th joint to the $(j+1)$th joint. They are all expressed in the frame attached to the second rotation axis of the *j*th joint.

Substituting Equation (9) into Equation (14), Equation (14) can be rewritten as follows.

$$\begin{cases} {}^{i,j,2}\boldsymbol{f}^{b}_{i,j} = {}^{i,j+1,2}\boldsymbol{f}^{b}_{i,j+1} + m^{b}_{i,j}\cdot\left({}^{i,j,2}\dot{\boldsymbol{v}}^{b}_{i,j} + {}^{i,j,2}\dot{\boldsymbol{\omega}}^{b}_{i,j} \times {}^{i,j,2}\boldsymbol{a}^{b}_{i,j} + {}^{i,j,2}\boldsymbol{\omega}^{b}_{i,j} \times \left({}^{i,j,2}\boldsymbol{\omega}^{b}_{i,j} \times {}^{i,j,2}\boldsymbol{a}^{b}_{i,j}\right)\right) \\ {}^{i,j,2}\boldsymbol{n}^{b}_{i,j} = {}^{i,j+1,2}\boldsymbol{n}^{b}_{i,j+1} - {}^{i,j,2}\boldsymbol{f}^{b}_{i,j} \times {}^{i,j,2}\boldsymbol{a}^{b}_{i,j} + {}^{i,j+1,2}\boldsymbol{f}^{b}_{i,j+1} \times \left({}^{i,j,2}\boldsymbol{a}^{b}_{i,j} - {}^{i,j,2}\boldsymbol{l}^{b}_{i,j}\right) \\ \quad + {}^{i,j,2}\boldsymbol{I}^{b}_{i,j}\cdot{}^{i,j,2}\dot{\boldsymbol{\omega}}^{b}_{i,j} + {}^{i,j,2}\boldsymbol{\omega}^{b}_{i,j} \times \left({}^{i,j,2}\boldsymbol{I}^{b}_{i,j}\cdot{}^{i,j,2}\boldsymbol{\omega}^{b}_{i,j}\right) \end{cases} \tag{15}$$

Similarly, a virtual link appears in the above iteration, which does not exist for the actual structure of Arm-b. The treatment for this case is consistent with that of Arm-a.

4. Pose Stabilization of the Base by the PD Control with Feedforward Compensation

4.1. The Strategy for Pose Stabilization During Transposition

During target satellite transposition, the pose stabilization strategy for the base of the combined system is depicted in Figure 5. In this process, the non-offset discrete arm (Arm-a) is the mission manipulator to move the target satellite from its initial pose to its desired pose. At the same time, the segmented cable-driven hyper-redundant serpentine arm (Arm-b) serves as the auxiliary manipulator for monitoring the operation of Arm-a. Firstly, the desired pose of the target satellite was determined and the time (i.e., t_f) was set based on transposition requirements. Accordingly, the target's movement trajectory is planned using an interpolation method with a sampling time dt. After that, the pose of the target satellite is known at each moment.

https://doi.org/10.3390/s26010206

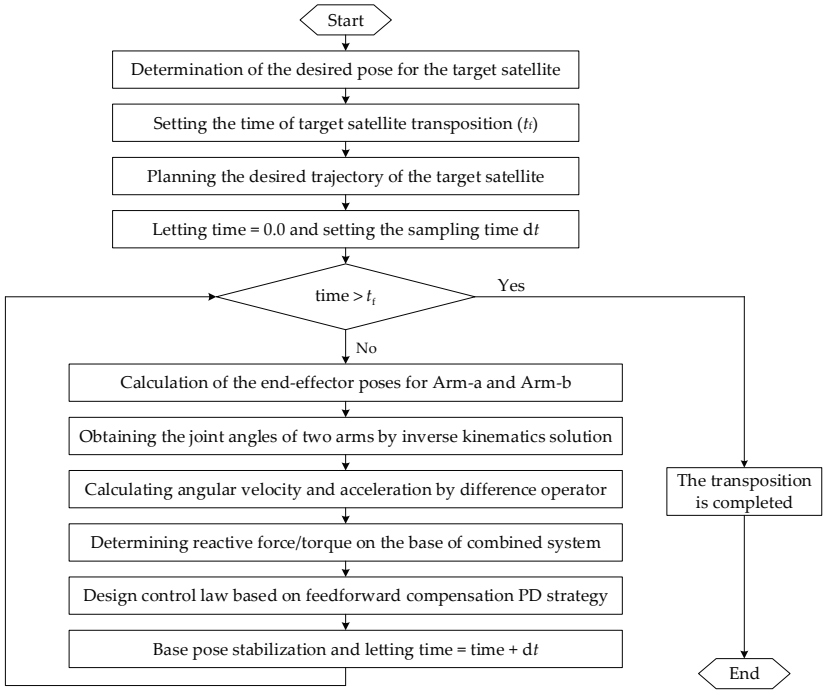


Figure 5. The flowchart of pose stabilization for the base of the combined system.

Then, the system checked whether the current time exceeds the allotted mission time. If it is true, it means that the target satellite has been moved to the desired pose. Otherwise, the expected poses of end-effectors are calculated according to the geometric constraints between the target satellite and the two arms. Specifically, the relative poses between the target satellite's center of mass, the capture point (Arm-a's end-effector), and the monitoring point (Arm-b's end-effector) remains fixed during transposition. Joint angular velocities and accelerations are determined by the inverse kinematics solution. The angular velocity and acceleration are obtained by the difference method. Furthermore, the PD control law integrated with dynamic feedforward compensation is designed to maintain the base stability. The feedforward compensation term is used to compensate for the base reaction force and torque generated by the motion of Arm-a and Arm-b, which can be calculated by the Newton–Euler recursion discussed in Section 3. The proportional-derivative control component is obtained according to the pose deviation of the base. Finally, the desired control force and torque are achieved by thrusters and reaction wheels installed on the base. This ensures the base remains stable while the DSHDASR performed the cooperative transposition-and-monitoring task.

4.2. Trajectory Planning of the Target Satellite

The schematic diagram for the target satellite transposition process is given in Figure 6. Before the transposition, the target satellite has been captured by Arm-a and the locked point is the capture point that is on the docking ring. During transposition, a monitoring device on Arm-b's end-effector provided auxiliary monitoring of Arm-a's operation. Therefore, the poses of Arm-a tip and Arm-b tip are dependent on the pose of the target satellite's center of mass. Since the transposition is to move the target satellite from the initial pose to the desired pose, it can be taken as point-to-point motion planning. To ensure smooth start and stop transitions, the trajectory for the target satellite's center of mass was generated using quintic polynomial interpolation.

https://doi.org/10.3390/s26010206

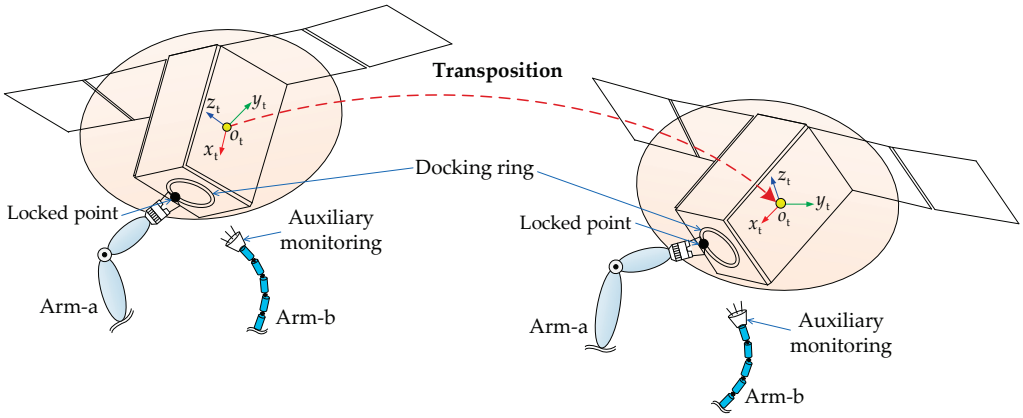


Figure 6. Target satellite transposition from the initial pose to the desired pose.

The initial pose and the desired pose of the target are set as follows:

$$\begin{cases} \boldsymbol{X}_{t_0} = \left[\begin{array}{cccccc} p_{t_0_x}, & p_{t_0_y}, & p_{t_0_z}, & \alpha_{t_0_z}, & \alpha_{t_0_y}, & \alpha_{t_0_x} \end{array} \right]^{\mathrm{T}} \\ \boldsymbol{X}_{t_f} = \left[\begin{array}{cccccc} p_{t_f_x}, & p_{t_f_y}, & p_{t_f_z}, & \alpha_{t_f_z}, & \alpha_{t_f_y}, & \alpha_{t_f_x} \end{array} \right]^{\mathrm{T}} \end{cases} \tag{16}$$

where $p_{t_0_x}$, $p_{t_0_y}$ and $p_{t_0_z}$ are initial positions of the target satellite's centroid along x-axis, y-axis, and z-axis, respectively. The desired positions are $p_{t_f_x}$, $p_{t_f_y}$, and $p_{t_f_z}$. The initial attitude angles (Z-Y-X Euler angles) are denoted by $\alpha_{t_0_z}$, $\alpha_{t_0_y}$, and $\alpha_{t_0_x}$. Their desired attitude angles are $\alpha_{t_f_z}$, $\alpha_{t_f_y}$, and $\alpha_{t_f_x}$, respectively. The velocity and acceleration are set to zeros at the initial and terminal states. Therefore, the desired motion trajectory of the target satellite is obtained as follows.

$$\boldsymbol{X}_{td}(t) = \boldsymbol{X}_{t_0} + \frac{10(\boldsymbol{X}_{t_f} - \boldsymbol{X}_{t_0})}{{t_f}^3} t^3 - \frac{15(\boldsymbol{X}_{t_f} - \boldsymbol{X}_{t_0})}{{t_f}^4} t^4 + \frac{6(\boldsymbol{X}_{t_f} - \boldsymbol{X}_{t_0})}{{t_f}^5} t^5, 0 \le t \le t_f \tag{17}$$

According to the pose relationship among the target satellite's centroid, Arm-a tip and Arm-b tip, the desired trajectories of the two arms are also calculated. Specifically, the desired poses of Arm-a tip and Arm-b tip are obtained, i.e., $\boldsymbol{P}_{ed}^{a}$, $\boldsymbol{R}_{ed}^{a}$, $\boldsymbol{P}_{ed}^{b}$, and $\boldsymbol{R}_{ed}^{b}$.

4.3. Feedforward Compensation PD Control Algorithm

As shown in Figure 7, the control method for the base's pose stabilization is composed of two components. The former is dynamic feedforward compensation, while the latter is proportional-derivative control. The detailed calculation process for each component is given as follows.

4.3.1. Calculation of Feedforward Compensation

As discussed above, once the motion trajectory for the target satellite's centroid was planned, the required end-effector poses for Arm-a and Arm-b became known. The joint angles, velocities, and accelerations of Arm-a and Arm-b are obtained by the inverse kinematics solution and numerical differentiation. These joint motion parameters are then fed into the Newton–Euler recursive algorithm to compute the corresponding joint forces and torques. Notably, the last link of Arm-a and the target satellite form a single rigid body after the target is captured. Therefore, the actual mass properties of the link used in the Newton–Euler recursion are changed, and they can be obtained as follows.

https://doi.org/10.3390/s26010206

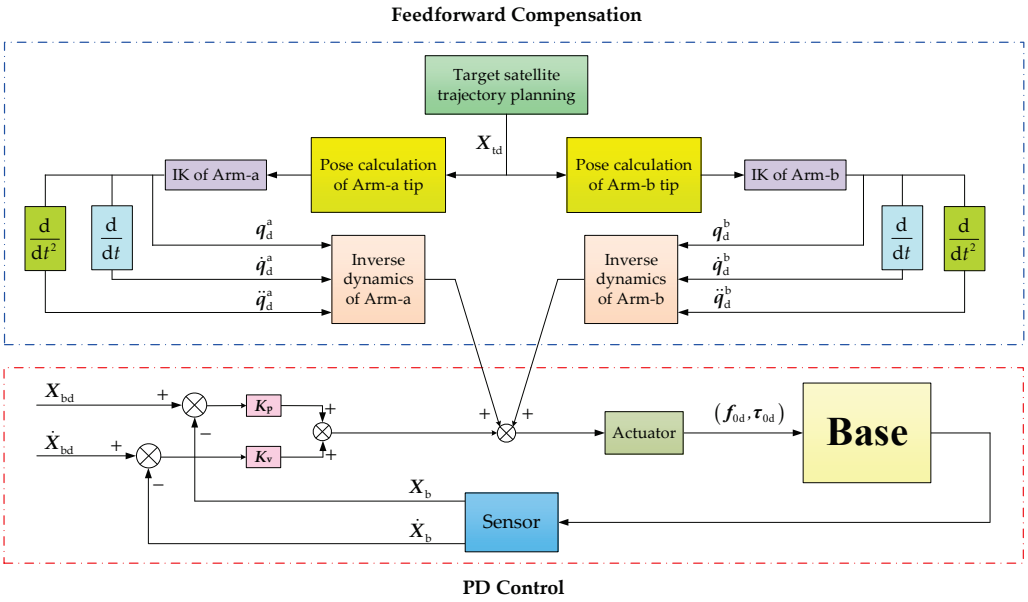


Figure 7. The flow chart of the PD control with feedforward compensation.

As shown in Figure 8, $\{x_{n_a}^a, y_{n_a}^a, z_{n_a}^a\}$ is the frame fixed on the last joint of Arm-a. $\{x_t, y_t, z_t\}$ is the frame attached to the target satellite. The equivalent centroid of the last link is denoted by C_{at}. Its associated frame is $\{x_{at}, y_{at}, z_{at}\}$. The position vector of C_{at} expressed in $\{x_{n_a}^a, y_{n_a}^a, z_{n_a}^a\}$ is obtained as follows:

$$\boldsymbol{r}_{ac} = \frac{m_{n_a}^a \boldsymbol{a}_{n_a}^a + m_t\left(\boldsymbol{a}_{n_a}^a + \boldsymbol{b}_{n_a}^a + \boldsymbol{r}_{et}\right)}{m_{n_a}^a + m_t} \tag{18}$$

where $m_{n_a}^a$ and m_t are the masses of the last link and the target, respectively. $\boldsymbol{r}_{et}$ is the position vector from Arm-a tip to the centroid of the target satellite.

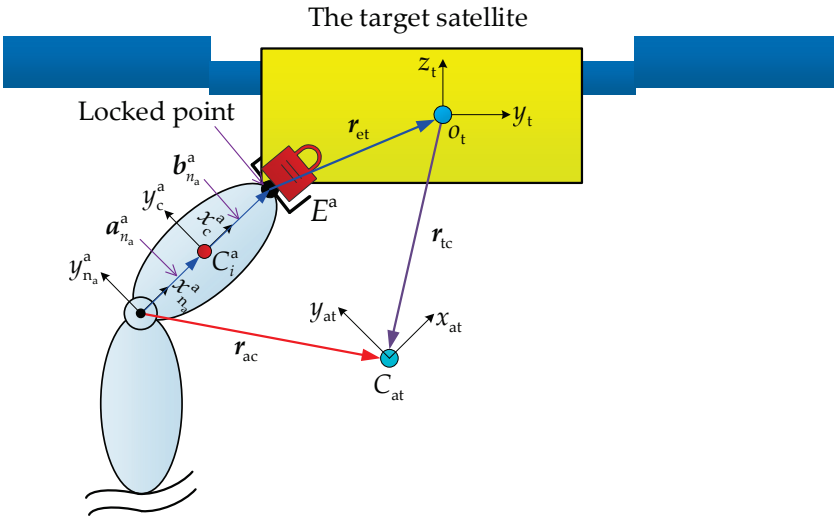


Figure 8. The equivalent last link of Arm-a constituted by the link and the target satellite.

The equivalent inertia tensor expressed in the $\{x_{n_a}^a, y_{n_a}^a, z_{n_a}^a\}$ is obtained as follows:

$$\boldsymbol{I}_{ac} = \left(\boldsymbol{I}_{n_a}^a + m_{n_a}^a\left(\boldsymbol{r}_{ac}{}^T \cdot \boldsymbol{r}_{ac}\right)\boldsymbol{E} - \boldsymbol{r}_{ac} \cdot \boldsymbol{r}_{ac}{}^T\right) + \left({}^{n_a}\boldsymbol{R}_t \boldsymbol{I}_t {}^{n_a}\boldsymbol{R}_t{}^T + m_t\left(\boldsymbol{r}_{tc}{}^T \cdot \boldsymbol{r}_{tc}\right)\boldsymbol{E} - \boldsymbol{r}_{tc} \cdot \boldsymbol{r}_{tc}{}^T\right) \tag{19}$$

where $\boldsymbol{I}_{n_a}^a$ and $\boldsymbol{I}_t$ are the inertia tensors of the last link and the target, respectively, expressed in their attached coordinate systems. $\boldsymbol{E}$ is the unit matrix. ${}^{n_a}\boldsymbol{R}_t$ is the rotation matrix from

https://doi.org/10.3390/s26010206

frame $\{x_t, y_t, z_t\}$ to frame $\{x^a_{n_a}, y^a_{n_a}, z^a_{n_a}\}$. $\boldsymbol{r}_{tc}$ is the position vector from the centroid of the target satellite to C_{at}.

Replace $m^a_{n_a}$, $\boldsymbol{a}^a_{n_a}$, and $\boldsymbol{I}^a_{n_a}$ with $m^a_{n_a} + m_t$, $\boldsymbol{r}_{ac}$, and $\boldsymbol{I}_{ac}$. These equivalent parameters, along with the joint motion data, are then substituted into (13) and (15) to yield the joint forces and torques. Subsequently, reaction forces ($-^1\boldsymbol{f}^a_1$ and $-^{1,1,2}\boldsymbol{f}^b_{1,1}$) and torques ($-^1\boldsymbol{n}^a_1$ and $-^{1,1,2}\boldsymbol{n}^b_{1,1}$) on the base induced by the two arms are obtained. By expressing these forces and torques in the inertial frame, the dynamic feedforward compensation term is derived as follows:

$$\begin{cases} \boldsymbol{f}^{DFC}_b = -\left(\boldsymbol{R}^a_1 \cdot {}^1\boldsymbol{f}^a_1 + \boldsymbol{R}^b_{1,1,2} \cdot {}^{1,1,2}\boldsymbol{f}^b_{1,1}\right) \\ \boldsymbol{n}^{DFC}_b = -\left[\boldsymbol{R}^a_1 \cdot {}^1\boldsymbol{n}^a_1 + \boldsymbol{b}^a_0 \times \left(\boldsymbol{R}^a_1 \cdot {}^1\boldsymbol{f}^a_1\right) + \boldsymbol{R}^b_{1,1,2} \cdot {}^{1,1,2}\boldsymbol{n}^b_{1,1} + \boldsymbol{b}^b_0 \times \left(\boldsymbol{R}^b_{1,1,2} \cdot {}^{1,1,2}\boldsymbol{f}^b_{1,1}\right)\right] \end{cases} \tag{20}$$

where $\boldsymbol{R}^a_1$ is the rotation matrix from the frame fixed on the first joint of Arm-a to the inertial frame. $\boldsymbol{R}^b_{1,1,2}$ refers to the rotation matrix from the frame attached to the second rotational axis of the first joint of Arm-b to the inertial frame.

4.3.2. Control Law Design

As shown in Figure 7, the pose stability control of the base consists of the dynamic feedforward compensation term and the PD control term. The former is calculated using Equation (20). The PD control term is derived from the deviation between the base's desired and measured poses. In practice, the desired pose is the base's state when the target satellite is grasped. The base's current pose is measured by onboard sensors (e.g., gyroscopes and accelerometers). Therefore, the PD control term is derived as follows:

$$\begin{cases} \boldsymbol{f}^{PD}_b = -\boldsymbol{K}_{pp} \cdot \Delta\boldsymbol{X}_p - \boldsymbol{K}_{vv} \cdot \Delta\boldsymbol{X}_v \\ \boldsymbol{n}^{PD}_b = -\boldsymbol{K}_{p\varphi} \cdot \Delta\boldsymbol{X}_\varphi - \boldsymbol{K}_{v\omega} \cdot \Delta\boldsymbol{X}_\omega \end{cases} \tag{21}$$

where $\Delta\boldsymbol{X}_p$ and $\Delta\boldsymbol{X}_\varphi$ are the position and attitude deviations for the base. Their linear and angular deviations are denoted by $\Delta\boldsymbol{X}_v$ and $\Delta\boldsymbol{X}_\omega$. $\boldsymbol{K}_{pp}$ and $\boldsymbol{K}_{p\varphi}$ are the proportional gain matrices corresponding to the position deviation and attitude deviation, respectively. $\boldsymbol{K}_{vv}$ and $\boldsymbol{K}_{v\omega}$ are the derivative gain matrices.

In this paper, the PD gains are designed following a systematic performance and robustness trade-off principle. Firstly, the gains should be proportional to the mass/inertia of the base, ensuring the physical consistency of the control force/torque. Furthermore, the gain matrix should be positive and definite to ensure system stability. Finally, to ensure system robustness, the minimum eigenvalue of the proportional gain matrix should be set in consideration of the magnitude of external disturbances and the required control accuracy. In other words, the proportional gain is expressed in a form explicitly related to the base mass, i.e., $\boldsymbol{K}_{pp} = m_0\widetilde{\boldsymbol{K}}_{pp}$. Here, m_0 denotes the base mass, and $\widetilde{\boldsymbol{K}}_{pp}$ represents a mass-independent normalized proportional gain that characterizes the desired closed-loop dynamic behavior.

Combining Equations (20) and (21), the control law of the base's pose stabilization is obtained as follows.

$$\begin{cases} \boldsymbol{f}_{bd} = \boldsymbol{f}^{DFC}_b + \boldsymbol{f}^{PD}_b = -\boldsymbol{K}_{pp}\Delta\boldsymbol{X}_p - \boldsymbol{K}_{vv}\Delta\boldsymbol{X}_v - \left(\boldsymbol{R}^{a1}_1 \boldsymbol{f}^a_1 + \boldsymbol{R}^b_{1,1,2}{}^{1,1,2}\boldsymbol{f}^b_{1,1}\right) \\ \boldsymbol{n}_{bd} = \boldsymbol{n}^{DFC}_b + \boldsymbol{n}^{PD}_b \\ \quad = -\boldsymbol{K}_{p\varphi}\Delta\boldsymbol{X}_\varphi - \boldsymbol{K}_{v\omega}\Delta\boldsymbol{X}_\omega - \left[\boldsymbol{R}^{a1}_1 \boldsymbol{n}^a_1 + \boldsymbol{b}^a_0 \times \left(\boldsymbol{R}^{a1}_1 \boldsymbol{f}^a_1\right) + \boldsymbol{R}^b_{1,1,2}{}^{1,1,2}\boldsymbol{n}^b_{1,1} + \boldsymbol{b}^b_0 \times \left(\boldsymbol{R}^b_{1,1,2}{}^{1,1,2}\boldsymbol{f}^b_{1,1}\right)\right] \end{cases} \tag{22}$$

https://doi.org/10.3390/s26010206

4.3.3. Generation of Control Force and Torque

The control forces and torques calculated from Equation (22) are generated by thrusters and reaction wheels. As shown in Figure 9, the layout of the thrusters (see Figure 9a) and reaction wheels (see Figure 9b) installed on the base is illustrated. According to Figure 9a, twelve thrusters are mounted on the base, organized into three groups aligned with the x_0, y_0, and z_0 axes of the base-fixed frame. Group 1 includes four thrusters that are numbered 1, 2, 3, and 4, respectively. These thrusters are arranged on the x_0-z_0 plane. Thrusters 1 and 2 produce thrust along the positive x_0 axis direction, while Thrusters 3 and 4 generate thrust along the negative x_0 axis direction. Group 2 consists of Thrusters 5, 6, 7, and 8, which are placed on the x_0-y_0 plane and Group 3 includes Thrusters 9, 10, 11, and 12, which are arranged on the y_0-z_0 plane. The definition of the thrust direction for these thrusters in Group 2 and Group 3 is the same as that of the thrusters in Group 1.

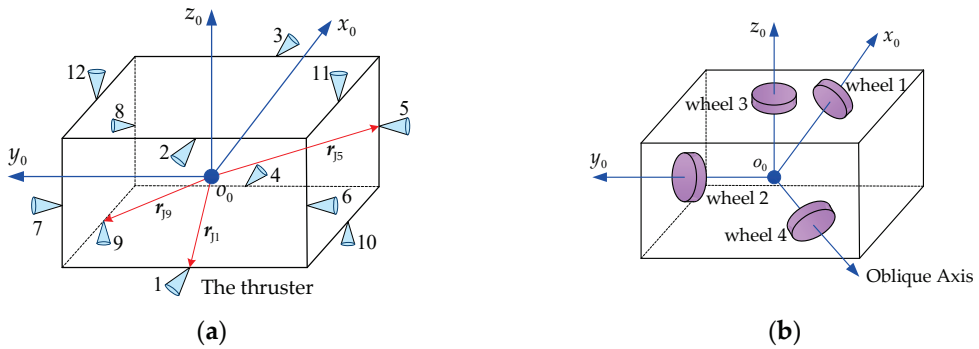


Figure 9. The configuration of thrusters and reaction wheels on the base. (**a**) The layout of thrusters. (**b**) The layout of reaction wheels.

The position vector of the ith thruster in the frame fixed on the base centroid is denoted as $\boldsymbol{r}_{\mathrm{J}i}$. To generate a desired control force along a principal axis, two opposing thrusters from the corresponding group should be fired. For example, Thrusters 1 and 4 are selected to generate the desired control force along the x_0 axis; Thrusters 5 and 8 are used for y_0 axis; and Thrusters 9 and 12 are selected for the z_0 axis. Thus, the desired control force from Equation (22) can be implemented as follows:

$$\begin{cases} f_{\mathrm{J}1} = \boldsymbol{f}_{\mathrm{bd}}(1), & f_{\mathrm{J}4} = 0, & if\ \boldsymbol{f}_{\mathrm{bd}}(1) > 0 \\ f_{\mathrm{J}1} = 0, & f_{\mathrm{J}4} = \boldsymbol{f}_{\mathrm{bd}}(1), & if\ \boldsymbol{f}_{\mathrm{bd}}(1) < 0 \\ f_{\mathrm{J}1} = 0, & f_{\mathrm{J}4} = 0, & if\ \boldsymbol{f}_{\mathrm{bd}}(1) = 0 \end{cases} \tag{23}$$

$$\begin{cases} f_{\mathrm{J}5} = \boldsymbol{f}_{\mathrm{bd}}(2), & f_{\mathrm{J}8} = 0, & if\ \boldsymbol{f}_{\mathrm{bd}}(2) > 0 \\ f_{\mathrm{J}5} = 0, & f_{\mathrm{J}8} = \boldsymbol{f}_{\mathrm{bd}}(2), & if\ \boldsymbol{f}_{\mathrm{bd}}(2) < 0 \\ f_{\mathrm{J}5} = 0, & f_{\mathrm{J}8} = 0, & if\ \boldsymbol{f}_{\mathrm{bd}}(2) = 0 \end{cases} \tag{24}$$

$$\begin{cases} f_{\mathrm{J}9} = \boldsymbol{f}_{\mathrm{bd}}(3), & f_{\mathrm{J}12} = 0, & if\ \boldsymbol{f}_{\mathrm{bd}}(3) > 0 \\ f_{\mathrm{J}9} = 0, & f_{\mathrm{J}12} = \boldsymbol{f}_{\mathrm{bd}}(3), & if\ \boldsymbol{f}_{\mathrm{bd}}(3) < 0 \\ f_{\mathrm{J}9} = 0, & f_{\mathrm{J}12} = 0, & if\ \boldsymbol{f}_{\mathrm{bd}}(3) = 0 \end{cases} \tag{25}$$

where $f_{\mathrm{J}i}$ is the thrust generated by the ith thruster. $\boldsymbol{f}_{\mathrm{bd}}(1)$, $\boldsymbol{f}_{\mathrm{bd}}(2)$, and $\boldsymbol{f}_{\mathrm{bd}}(3)$ are the components of the vector $\boldsymbol{f}_{\mathrm{bd}}$ along the x_0, y_0, and z_0 axes, respectively.

The layout of the reaction wheels is depicted in Figure 9b. Wheels 1, 2, and 3 are installed along the x_0, y_0, and z_0 axes, respectively. Each wheel generates the control torque to adjust the attitude about its corresponding axis. Wheel 4 is oriented along an oblique axis to serve as a backup. In case of a failure in any one of Wheels 1–3, it is activated to maintain base attitude control.

https://doi.org/10.3390/s26010206

Once the desired control torque in the feedforward compensation PD control algorithm is determined, it can be generated through rotation of the reaction wheels. Since the thrust direction of each thruster does not pass through the center of mass of the base, their operation introduces disturbance torques to the base. Therefore, the total torque output from the reaction wheels has two components, namely the nominal control torque from the algorithm and a compensatory torque to cancel thruster-induced disturbances. Taking Thrusters 1, 4, 5, 8, 9, and 12 as an example, the disturbance torque generated by their thrusts is obtained as follows.

$$\begin{aligned} \boldsymbol{n}_{\text{ext}} &= \boldsymbol{r}_{\text{J1}} \times \begin{bmatrix} f_{\text{J1}} & 0 & 0 \end{bmatrix}^{\text{T}} + \boldsymbol{r}_{\text{J4}} \times \begin{bmatrix} -f_{\text{J4}} & 0 & 0 \end{bmatrix}^{\text{T}} \\ &+ \boldsymbol{r}_{\text{J5}} \times \begin{bmatrix} 0 & f_{\text{J5}} & 0 \end{bmatrix}^{\text{T}} + \boldsymbol{r}_{\text{J8}} \times \begin{bmatrix} 0 & -f_{\text{J8}} & 0 \end{bmatrix}^{\text{T}} \\ &+ \boldsymbol{r}_{\text{J9}} \times \begin{bmatrix} 0 & 0 & f_{\text{J9}} \end{bmatrix}^{\text{T}} + \boldsymbol{r}_{\text{J12}} \times \begin{bmatrix} 0 & 0 & -f_{\text{J12}} \end{bmatrix}^{\text{T}} \end{aligned} \tag{26}$$

Combining Equations (22) and (26) yields the total three-axis control torque required from the reaction wheels.

$$\boldsymbol{n}_{\text{rw}} = \boldsymbol{n}_{\text{bd}} + \boldsymbol{n}_{\text{ext}} \tag{27}$$

Assuming that Reaction Wheels 1, 2, and 3 are operating nominally, the relationship between the control torque and its angular acceleration is given as follows:

$$\boldsymbol{n}_{\text{rw}} = \begin{bmatrix} I_{\text{rwx}} & 0 & 0 \\ 0 & I_{\text{rwy}} & 0 \\ 0 & 0 & I_{\text{rwz}} \end{bmatrix} \begin{bmatrix} \dot{\omega}_{\text{rwx}} \\ \dot{\omega}_{\text{rwy}} \\ \dot{\omega}_{\text{rwz}} \end{bmatrix} \tag{28}$$

where I_{fwx}, I_{rwy}, and I_{rwz} stand for moments of inertia of Reaction Wheels 1, 2, and 3, respectively. $\dot{\omega}_{\text{rwx}}$, $\dot{\omega}_{\text{rwy}}$, and $\dot{\omega}_{\text{rwz}}$ refer to angular accelerations of Reaction Wheels 1, 2, and 3, respectively.

Substituting Equation (27) into Equation (28), the desired angular acceleration of each reaction wheel is calculated as follows.

$$\begin{bmatrix} \dot{\omega}_{\text{rwx}} \\ \dot{\omega}_{\text{rwy}} \\ \dot{\omega}_{\text{rwz}} \end{bmatrix} = \begin{bmatrix} I_{\text{rwx}} & 0 & 0 \\ 0 & I_{\text{rwy}} & 0 \\ 0 & 0 & I_{\text{rwz}} \end{bmatrix}^{-1} (\boldsymbol{n}_{\text{bd}} + \boldsymbol{n}_{\text{ext}}) \tag{29}$$

Finally, the pose stabilization control of the combined system's base is achieved by operating thrusters and reaction wheels according to Equations (23)–(25) and (29).

It is worth noting that actuator saturation can disrupt the controller's linearity and may impair stability. To address this issue, the saturation limits are discussed as follows. From the perspective of control force and torque, their magnitude are determined by the feedforward compensation and the PD control term. Among these, the feedforward compensation contributes the majority, while the PD term accounts for a smaller portion. Therefore, to ensure that the feedforward compensation does not induce actuator saturation, a limiter has been applied to it. The specific settings are given as follows:

$$\boldsymbol{f}_{\text{b}}^{\text{PD}} = \begin{cases} \left(f_{\text{b,max}}^{\text{PD}} / \left\| \boldsymbol{f}_{\text{b}}^{\text{PD}} \right\| \right) \cdot \left\| \boldsymbol{f}_{\text{b}}^{\text{PD}} \right\|, & \text{if } \left\| \boldsymbol{f}_{\text{b}}^{\text{PD}} \right\| > f_{\text{b,max}}^{\text{PD}} \\ \boldsymbol{f}_{\text{b}}^{\text{PD}}, & \text{else} \end{cases} \tag{30}$$

$$\boldsymbol{n}_{\text{b}}^{\text{PD}} = \begin{cases} \left(n_{\text{b,max}}^{\text{PD}} / \left\| \boldsymbol{n}_{\text{b}}^{\text{PD}} \right\| \right) \cdot \boldsymbol{n}_{\text{b}}^{\text{PD}}, & \text{if } \left\| \boldsymbol{n}_{\text{b}}^{\text{PD}} \right\| > n_{\text{b,max}}^{\text{PD}} \\ \boldsymbol{n}_{\text{b}}^{\text{PD}}, & \text{else} \end{cases} \tag{31}$$

where $f_{\text{b,max}}^{\text{PD}}$ and $n_{\text{b,max}}^{\text{PD}}$ are the maximum output force and torque of the actuator.

https://doi.org/10.3390/s26010206

4.4. Analysis of the Proposed Algorithm

4.4.1. Stability Analysis

The movement of the base is composed of the translation and rotation. According to Newton–Euler formulation, the dynamics model of the base can be obtained as follows:

$$m_0\dot{\boldsymbol{v}}_0 = \boldsymbol{f}_{\mathrm{bd}} + \boldsymbol{f}_{\mathrm{arm}} + \boldsymbol{f}_{\mathrm{ext}} \tag{32}$$

$$\boldsymbol{I}_0\dot{\boldsymbol{\omega}}_0 + \boldsymbol{\omega}_0 \times \boldsymbol{I}_0\boldsymbol{\omega}_0 = \boldsymbol{n}_{\mathrm{bd}} + \boldsymbol{n}_{\mathrm{arm}} + \boldsymbol{n}_{\mathrm{ext}} \tag{33}$$

where m_0 and $\boldsymbol{I}_0$ are the mass and inertia of the base. Its linear velocity and angular velocity are $\boldsymbol{v}_0$ and $\boldsymbol{\omega}_0$, respectively. $\boldsymbol{f}_{\mathrm{arm}}$ and $\boldsymbol{n}_{\mathrm{arm}}$ are the reaction force and reaction torque on the base induced by the two arms. $\boldsymbol{f}_{\mathrm{ext}}$ and $\boldsymbol{n}_{\mathrm{ext}}$ are the external disturbance force and torque, respectively.

Based on Equation (22), the force and torque of PD control with feedforward compensation are obtained. Substituting Equation (22) into Equations (32) and (33), it is calculated as follows:

$$m_0\dot{\boldsymbol{v}}_0 = -\boldsymbol{K}_{\mathrm{pp}}\Delta\boldsymbol{X}_{\mathrm{p}} - \boldsymbol{K}_{\mathrm{vv}}\Delta\boldsymbol{X}_{\mathrm{v}} + \boldsymbol{f}_{\mathrm{b}}^{\mathrm{DFC}} + \boldsymbol{f}_{\mathrm{arm}} + \boldsymbol{f}_{\mathrm{ext}} \tag{34}$$

$$\boldsymbol{I}_0\dot{\boldsymbol{\omega}}_0 + \boldsymbol{\omega}_0 \times \boldsymbol{I}_0\boldsymbol{\omega}_0 = -\boldsymbol{K}_{\mathrm{p}\varphi}\Delta\boldsymbol{X}_{\varphi} - \boldsymbol{K}_{\mathrm{v}\omega}\Delta\boldsymbol{X}_{\omega} + \boldsymbol{n}_{\mathrm{b}}^{\mathrm{DFC}} + \boldsymbol{n}_{\mathrm{arm}} + \boldsymbol{n}_{\mathrm{ext}} \tag{35}$$

where $\Delta\boldsymbol{X}_{\mathrm{p}} = \boldsymbol{p}_0 - \boldsymbol{p}_{0\mathrm{d}}$, $\Delta\boldsymbol{X}_{\mathrm{v}} = \dot{\boldsymbol{p}}_0 - \dot{\boldsymbol{p}}_{0\mathrm{d}} = \dot{\boldsymbol{p}}_0$. Because the base undergoes only small variations in its attitude, $\Delta\boldsymbol{X}_{\varphi}$ and $\Delta\boldsymbol{X}_{\omega}$ can be calculated by $\Delta\boldsymbol{X}_{\varphi} = \boldsymbol{\varphi}_0 - \boldsymbol{\varphi}_{0\mathrm{d}}$ and $\Delta\boldsymbol{X}_{\omega} = \dot{\boldsymbol{\varphi}}_0 - \dot{\boldsymbol{\varphi}}_{0\mathrm{d}} = \dot{\boldsymbol{\varphi}}_0$. $\boldsymbol{p}_0$ and $\boldsymbol{\varphi}_0$ are the current position and attitude of the base. $\boldsymbol{p}_{0\mathrm{d}}$ and $\boldsymbol{\varphi}_{0\mathrm{d}}$ are the desired position and attitude. Thus, $\dot{\boldsymbol{p}}_0$ and $\dot{\boldsymbol{\varphi}}_0$ are the current linear velocity and angular velocity of the base. $\dot{\boldsymbol{p}}_{0\mathrm{d}}$ and $\dot{\boldsymbol{\varphi}}_{0\mathrm{d}}$ are the desired linear velocity and angular velocity, and they are set as zeros in this paper. Therefore, $\boldsymbol{v}_0 = \dot{\boldsymbol{p}}_0$, $\dot{\boldsymbol{v}}_0 = \ddot{\boldsymbol{p}}_0$, $\boldsymbol{\omega}_0 = \dot{\boldsymbol{\varphi}}_0$, and $\dot{\boldsymbol{\omega}}_0 = \ddot{\boldsymbol{\varphi}}_0$ can be obtained.

From a reformulation of Equations (34) and (35), the closed-loop dynamics for the base's position and attitude can be derived as follows:

$$m_0\ddot{\boldsymbol{p}}_0 + \boldsymbol{K}_{\mathrm{pp}}\Delta\boldsymbol{X}_{\mathrm{p}} + \boldsymbol{K}_{\mathrm{vv}}\dot{\boldsymbol{p}}_0 = \boldsymbol{\delta}_{\mathrm{f}} \tag{36}$$

$$\boldsymbol{I}_0\ddot{\boldsymbol{\varphi}}_0 + \boldsymbol{K}_{\mathrm{p}\varphi}\Delta\boldsymbol{X}_{\varphi} + \boldsymbol{K}_{\mathrm{v}\omega}\dot{\boldsymbol{\varphi}}_0 = \boldsymbol{\delta}_{\tau} \tag{37}$$

where $\boldsymbol{\delta}_{\mathrm{f}} = \boldsymbol{f}_{\mathrm{b}}^{\mathrm{DFC}} + \boldsymbol{f}_{\mathrm{arm}} + \boldsymbol{f}_{\mathrm{ext}}$ and $\boldsymbol{\delta}_{\tau} = \boldsymbol{n}_{\mathrm{b}}^{\mathrm{DFC}} + \boldsymbol{n}_{\mathrm{arm}} + \boldsymbol{n}_{\mathrm{ext}} - \boldsymbol{\omega}_0 \times \boldsymbol{I}_0\boldsymbol{\omega}_0$. $\boldsymbol{\delta}_{\mathrm{f}}$ and $\boldsymbol{\delta}_{\tau}$ are defined as the lumped disturbance in the position channel and in attitude channel, respectively.

Assuming that $\boldsymbol{\delta}_{\mathrm{f}}$ and $\boldsymbol{\delta}_{\tau}$ are zeros, Equations (36) and (37) can be rewritten as follows.

$$\ddot{\boldsymbol{p}}_0 = \left(-\boldsymbol{K}_{\mathrm{pp}}\Delta\boldsymbol{X}_{\mathrm{p}} - \boldsymbol{K}_{\mathrm{vv}}\dot{\boldsymbol{p}}_0\right)/m_0 \tag{38}$$

$$\ddot{\boldsymbol{\varphi}}_0 = {\boldsymbol{I}_0}^{-1}\left(-\boldsymbol{K}_{\mathrm{p}\varphi}\Delta\boldsymbol{X}_{\varphi} - \boldsymbol{K}_{\mathrm{v}\omega}\dot{\boldsymbol{\varphi}}_0\right) \tag{39}$$

The Lyapunov function of the control system is established as follows.

$$\begin{aligned} V &= V_{\mathrm{p}} + V_{\varphi} = \tfrac{1}{2}\left(\Delta\boldsymbol{X}_{\mathrm{v}}{}^{\mathrm{T}} m_0\Delta\boldsymbol{X}_{\mathrm{v}} + \Delta\boldsymbol{X}_{\mathrm{p}}{}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{pp}}\Delta\boldsymbol{X}_{\mathrm{p}}\right) + \tfrac{1}{2}\left(\Delta\boldsymbol{X}_{\omega}{}^{\mathrm{T}}\boldsymbol{I}_0\Delta\boldsymbol{X}_{\omega} + \Delta\boldsymbol{X}_{\varphi}{}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{p}\varphi}\Delta\boldsymbol{X}_{\varphi}\right) \\ &= \tfrac{1}{2}\left(\dot{\boldsymbol{p}}_0{}^{\mathrm{T}} m_0\dot{\boldsymbol{p}}_0 + \Delta\boldsymbol{X}_{\mathrm{p}}{}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{pp}}\Delta\boldsymbol{X}_{\mathrm{p}}\right) + \tfrac{1}{2}\left(\dot{\boldsymbol{\varphi}}_0{}^{\mathrm{T}}\boldsymbol{I}_0\dot{\boldsymbol{\varphi}}_0 + \Delta\boldsymbol{X}_{\varphi}{}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{p}\varphi}\Delta\boldsymbol{X}_{\varphi}\right) \end{aligned} \tag{40}$$

Owing to the fact that m_0, $\boldsymbol{K}_{\mathrm{pp}}$, $\boldsymbol{I}_0$, and $\boldsymbol{K}_{\mathrm{p}\varphi}$ are all positive definite matrix values, V is greater than or equal to zero. Furthermore, only when $\dot{\boldsymbol{p}}_0$, $\Delta\boldsymbol{X}_{\mathrm{p}}$, $\dot{\boldsymbol{\varphi}}_0$, and $\Delta\boldsymbol{X}_{\varphi}$ are zeros, V is equal to zero. Differentiating Equation (40), it can be obtained as follows.

$$\dot{V} = \dot{\boldsymbol{p}}_0{}^{\mathrm{T}} m_0\ddot{\boldsymbol{p}}_0 + \Delta\boldsymbol{X}_{\mathrm{p}}{}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{pp}}\dot{\boldsymbol{p}}_0 + \dot{\boldsymbol{\varphi}}_0{}^{\mathrm{T}}\boldsymbol{I}_0\ddot{\boldsymbol{\varphi}}_0 + \Delta\boldsymbol{X}_{\varphi}{}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{p}\varphi}\dot{\boldsymbol{\varphi}}_0 \tag{41}$$

https://doi.org/10.3390/s26010206

Substituting Equations (38) and (39) into Equation (41), we can obtain

$$\dot{V} = -\dot{\boldsymbol{p}}_0{}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{vv}}\dot{\boldsymbol{p}}_0 - \dot{\boldsymbol{\varphi}}_0{}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{v}\omega}\dot{\boldsymbol{\varphi}}_0 \leq 0 \tag{42}$$

According to Equation (42), the control system is globally asymptotically stable.

The stability analysis presented above is derived under nominal conditions, where the dynamic feedforward compensation is assumed to exactly cancel the base coupling disturbances induced by the motion of the manipulators and the captured target satellite. Under this ideal assumption, the closed-loop system errors are guaranteed to asymptotically converge to zero.

In practical implementations, however, the perfect feedforward cancellation of coupling effects is difficult to achieve due to modeling uncertainties, parameter deviations, and unmodeled dynamics. In this case, the system is subject to bounded residual disturbances. Based on Lyapunov stability theory, it can be shown that the closed-loop system remains uniformly ultimately bounded, where the tracking errors converge to a bounded neighborhood whose size depends on the upper bound of the disturbances, rather than strictly converging to zero.

4.4.2. Robustness Analysis

Based on Equations (36) and (37), the lumped disturbances in position and attitude channels are $\boldsymbol{\delta}_{\mathrm{f}}$ and $\boldsymbol{\delta}_{\tau}$. They can both be decomposed into two components. The former is the feedforward compensation error, which could be generated due to parameter uncertainties or modeling errors. The latter is the disturbance force (or torque) exerted by the environment. When the base reaches the stable state, $\ddot{\boldsymbol{p}}_0$, $\dot{\boldsymbol{p}}_0$, $\dot{\boldsymbol{\varphi}}_0$, and $\ddot{\boldsymbol{\varphi}}_0$ are all zeros. Therefore, its position and attitude errors can be obtained as follows.

$$\Delta\boldsymbol{X}_{\mathrm{p}0} = \left(\boldsymbol{K}_{\mathrm{pp}}\right)^{-1}\cdot\boldsymbol{\delta}_{\mathrm{f}}, \quad \Delta\boldsymbol{X}_{\varphi} = \left(\boldsymbol{K}_{\mathrm{p}\varphi}\right)^{-1}\cdot\boldsymbol{\delta}_{\tau} \tag{43}$$

Error bounds for the pose of the base are determined as follows:

$$\begin{cases} e_{\mathrm{p}} \leq \left\| \left(\boldsymbol{K}_{\mathrm{pp}}\right)^{-1} \right\| \|\boldsymbol{\delta}_{\mathrm{f}}\| = \frac{1}{\lambda_{\min}\left(\boldsymbol{K}_{\mathrm{pp}}\right)} \|\boldsymbol{\delta}_{\mathrm{f}}\| \\ e_{\mathrm{a}} \leq \left\| \left(\boldsymbol{K}_{\mathrm{p}\varphi}\right)^{-1} \right\| \|\boldsymbol{\delta}_{\tau}\| = \frac{1}{\lambda_{\min}\left(\boldsymbol{K}_{\mathrm{p}\varphi}\right)} \|\boldsymbol{\delta}_{\tau}\| \end{cases} \tag{44}$$

where $\lambda_{\min}(\cdot)$ is the minimum eigenvalue of a matrix.

According to Equation (44), the position and attitude errors are related to external disturbances and proportional gain. Additionally, the larger minimum eigenvalue of the proportional gain results in a smaller error. As a result, we can enhance the robustness of the system by increasing the minimum eigenvalue of the proportional gain matrix.

4.4.3. Gain Margins Analysis

The system can be decoupled into six independent second-order systems.

$$\begin{cases} m_0\ddot{\boldsymbol{p}}_0(i) + \boldsymbol{K}_{\mathrm{pp}}(i,i)\Delta\boldsymbol{X}_{\mathrm{p}}(i) + \boldsymbol{K}_{\mathrm{vv}}(i,i)\dot{\boldsymbol{p}}_0(i) = \boldsymbol{\delta}_{\mathrm{f}}(i) \\ \boldsymbol{I}_0(i,i)\ddot{\boldsymbol{\varphi}}_0(i) + \boldsymbol{K}_{\mathrm{p}\varphi}(i,i)\Delta\boldsymbol{X}_{\varphi}(i) + \boldsymbol{K}_{\mathrm{v}\omega}(i,i)\dot{\boldsymbol{\varphi}}_0(i) = \boldsymbol{\delta}_{\tau}(i) \end{cases} \quad i = 1,2,3 \tag{45}$$

For the position channel, the open-loop transfer function for each axis is given as follows.

$$G_i^p(s) = \frac{\boldsymbol{K}_{\mathrm{vv}}(i,i)s + \boldsymbol{K}_{\mathrm{pp}}(i,i)}{ms^3}e^{-\tau s} \tag{46}$$

https://doi.org/10.3390/s26010206

The phase crossover frequency ω_i^{pc} satisfies the following equation.

$$\arctan\left(\frac{\boldsymbol{K}_{\mathrm{vv}}(i,i)\omega_i^{\mathrm{pc}}}{\boldsymbol{K}_{\mathrm{pp}}(i,i)}\right)-\frac{\pi}{2}-\tau\omega_i^{\mathrm{pc}}=-\pi \tag{47}$$

As derived from Equation (47), ω_i^{pc} can be determined by using numerical methods such as the Newton–Raphson method or the bisection method. The corresponding gain at this frequency are calculated as follows.

$$\left|L_i^p\left(j\omega_i^{\mathrm{pc}}\right)\right|=\frac{\sqrt{\left[\boldsymbol{K}_{\mathrm{pp}}(i,i)\right]^2+\left[\boldsymbol{K}_{\mathrm{vv}}(i,i)\omega_i^{\mathrm{pc}}\right]^2}}{m_0\left(\omega_i^{\mathrm{pc}}\right)^2}\cdot\frac{1}{\sqrt{1+\left(\tau\omega_i^{\mathrm{pc}}\right)^2}} \tag{48}$$

Thus, the gain margin of the position channel is obtained as follows.

$$G_{\mathrm{M},i}^{\mathrm{p}}=\frac{1}{\left|L_i^p\left(j\omega_i^{\mathrm{pc}}\right)\right|},\quad G_{\mathrm{M,dB},i}^{\mathrm{p}}=-20\lg\left(\left|L_i^p\left(j\omega_i^{\mathrm{pc}}\right)\right|\right) \tag{49}$$

Similarly, the gain margin for the attitude channel can also be calculated, which is omitted here for brevity. It should be noted that the gain margin analysis presented in this section is conducted around the nominal operating point based on a linearized model. As such, it serves primarily as a theoretical stability reference and has inherent limitations when applied to the nonlinear and time-varying dynamics of the base–manipulator–target combined system, particularly under large payload conditions.

4.4.4. Sensitivity Analysis

During the transposition, the desired velocity and acceleration of the base are zeros, i.e., $\dot{\boldsymbol{p}}_{0\mathrm{d}}=0$ and $\dot{\boldsymbol{\varphi}}_{0\mathrm{d}}=0$. Substituting them into Equation (45), the closed-loop control system for the base is established as follows.

$$\begin{cases} m_0\Delta\ddot{\boldsymbol{X}}_{\mathrm{p}}(i)+\boldsymbol{K}_{\mathrm{vv}}(i,i)\Delta\dot{\boldsymbol{X}}_{\mathrm{p}}(i)+\boldsymbol{K}_{\mathrm{pp}}(i,i)\Delta\boldsymbol{X}_{\mathrm{p}}(i)=\boldsymbol{\delta}_{\mathrm{f}}(i) \\ \boldsymbol{I}_0(i,i)\Delta\ddot{\boldsymbol{X}}_{\varphi}(i)+\boldsymbol{K}_{\mathrm{v}\omega}(i,i)\Delta\dot{\boldsymbol{X}}_{\varphi}(i)+\boldsymbol{K}_{\mathrm{p}\varphi}(i,i)\Delta\boldsymbol{X}_{\varphi}(i)=\boldsymbol{\delta}_{\tau}(i) \end{cases} \quad i=1,2,3 \tag{50}$$

Based on Equation (50), the position or attitude channels can be treated as a second-order system that is fully defined by two characteristic parameters (damping ratio and natural frequency). Since the inherent parameters of a second-order system define its intrinsic dynamic characteristics, which remain unchanged by external disturbances, the disturbance force and torque can be neglected in Equation (50).

Taking the sensitivity analysis of parameter uncertainty as an example, assume that the true mass parameter of the base is $m_0=(1+\lambda_m)\widetilde{m}_0$. $\widetilde{m}_0$ is the estimated mass of the base and λ_m is the uncertainty. From Equation (50), the closed-loop characteristic equation for the position channel can be derived as

$$s_i^2+\frac{\boldsymbol{K}_{\mathrm{vv}}(i,i)}{(1+\lambda_m)\widetilde{m}_0}s_i+\frac{\boldsymbol{K}_{\mathrm{pp}}(i,i)}{(1+\lambda_m)\widetilde{m}_0}=0\quad i=1,2,3 \tag{51}$$

From Equation (51), it can be seen that the uncertainty λ_m affects the distribution of the system poles. When m_0 increases (i.e., $\lambda_m>0$), the poles move toward the imaginary axis, potentially making the response more oscillatory. In this case, the natural frequency decreases and the damping ratio reduces. When m_0 decreases (i.e., $\lambda_m<0$), the poles move away from the imaginary axis, leading to an increase in both the natural frequency and

https://doi.org/10.3390/s26010206

damping ratio. However, if m_0 becomes too small, the poles might approach the imaginary axis again and exacerbate oscillations.

The damping ratio and natural frequency can be calculated as follows.

$$\zeta_i^{\mathrm{p}} = \frac{\boldsymbol{K}_{\mathrm{vv}}(i,i)}{2\sqrt{m_0 \boldsymbol{K}_{\mathrm{pp}}(i,i)}}, \quad \omega_{\mathrm{n},i}^{\mathrm{p}} = \sqrt{\frac{\boldsymbol{K}_{\mathrm{pp}}(i,i)}{m_0}} \tag{52}$$

For the position channel, the sensitivity of the damping ratio and natural frequency with respect to the mass parameter uncertainty are then obtained as follows.

$$S_{\zeta}^{\mathrm{p}} = \frac{\partial \zeta_i^{\mathrm{p}}}{\partial m_0} \cdot \frac{m_0}{\zeta_i^{\mathrm{p}}} = -\frac{1}{2}, \quad S_{\omega}^{\mathrm{p}} = \frac{\partial \omega_{\mathrm{n},i}^{\mathrm{p}}}{\partial m_0} \cdot \frac{m_0}{\omega_{\mathrm{n},i}^{\mathrm{p}}} = -\frac{1}{2} \tag{53}$$

From Equation (53), it implies that a 10% increase in mass results in approximately a 5% decrease in both the damping ratio and natural frequency.

Similarly, the closed-loop characteristic function for the attitude channel can be established based on Equation (50). Its damping ratio and natural frequency can also be calculated. The corresponding sensitivity functions of the damping ratio and natural frequency with respect to the inertia parameter uncertainty are obtained as follows.

$$S_{\zeta}^{\varphi} = \frac{\partial \zeta_i^{\varphi}}{\partial \boldsymbol{I}_0(i,i)} \cdot \frac{\boldsymbol{I}_0(i,i)}{\zeta_i^{\varphi}} = -\frac{1}{2}, \quad S_{\omega}^{\varphi} = \frac{\partial \omega_{\mathrm{n},i}^{\varphi}}{\partial \boldsymbol{I}_0(i,i)} \cdot \frac{\boldsymbol{I}_0(i,i)}{\omega_{\mathrm{n},i}^{\varphi}} = -\frac{1}{2} \tag{54}$$

where $\zeta_i^{\varphi} = \boldsymbol{K}_{\mathrm{v}\omega}(i,i)/\sqrt{4\boldsymbol{I}_0(i,i)\boldsymbol{K}_{\mathrm{p}\varphi}(i,i)}$ and $\omega_{\mathrm{n},i}^{\varphi} = \sqrt{\boldsymbol{K}_{\mathrm{p}\varphi}(i,i)/\boldsymbol{I}_0(i,i)}$. A 10% increase in inertia reduces both the natural frequency and damping ratio by about 5%. This trend is fully analogous to the position channel analysis.

4.4.5. Discussion of the Challenges

Regarding the time delay issue, system delays from sensor sampling, communication, and computation could introduce phase errors, particularly degrading the performance of the model-dependent feedforward compensation. This would place a greater burden on the robustness of the PD feedback loop. We will study some strategies like predictive filtering or high-speed data protocols to address this issue.

On the problem of computational cost, the real-time execution of the recursive Newton–Euler algorithm for feedforward dynamics, while efficient ($o(n)$), must be rigorously validated on representative flight processor hardware.

About the actuator saturation, the physical limits of the thrusters and reaction wheels are the critical constraint. Each actuator requires integrating a saturation management block and a control allocation strategy to distribute torque among these actuators, preventing performance degradation or instability when commanded values exceed physical limits.

5. Simulation Study

To verify the proposed method, a co-simulation model is built using Matlab2016 and Adams2018 software. As shown in Figure 10, the DSHDASR for cooperatively transposing the target satellite is taken as an example in the simulation. At the initial moment of the simulation, the target has been grasped by the DSHDASR. The capture point is on the docking ring of the target, and it is referred to as the locked point. During the transposition, the discrete arm (Arm-a) is moving the target satellite from the initial pose to the desired pose, while the serpentine arm (Arm-b) is offering auxiliary monitoring for the operation of Arm-a.

https://doi.org/10.3390/s26010206

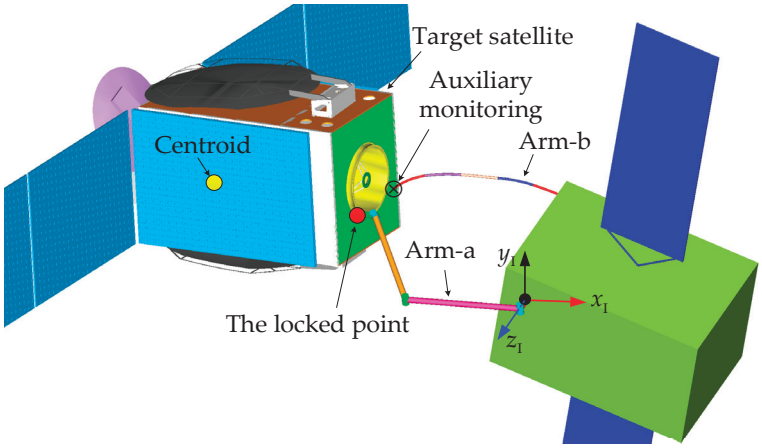


Figure 10. The initial state of the combined system in the simulation.

5.1. The Parameters of Simulation Model

5.1.1. The Parameters of the DSHDASR

The DSHDASR is composed of the base, Arm-a, and Arm-b. Arm-a is a non-offset discrete manipulator with seven degrees of freedom (DoFs). Its D-H frames and parameters are given in Figure 11 and Table 1. The mass parameters are listed in Table 2.

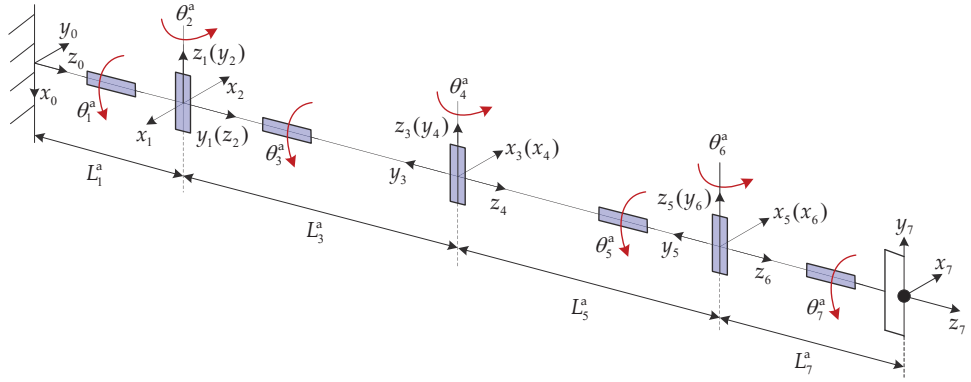


Figure 11. The D-H frames of Arm-a.

Table 1. The D-H parameters of Arm-a.

Link i	θ_i (°)	α_i (°)	a_i (m)	d_i (m)
1	−90	90	0	$L_1^{\mathrm{a}} = 0.255$
2	180	90	0	0
3	0	−90	0	$L_3^{\mathrm{a}} = 1.9$
4	0	90	0	0
5	0	−90	0	$L_5^{\mathrm{a}} = 1.9$
6	0	90	0	0
7	0	0	0	$L_7^{\mathrm{a}} = 0.255$

Arm-b is a segmented serpentine arm with five identical segments. Each segment has six associated universal joints. All segments share an identical mechanical design. The D-H frames and parameters for a representative segment are shown in Figure 12 and provided in Table 3. The length and mass property of each link for Arm-b are given as follows.

$$L_{i,j}^{\mathrm{b}} = 0.1\ \mathrm{m},\ \ m_{i,j}^{\mathrm{b}} = 0.1\ \mathrm{kg},\ \ \boldsymbol{I}_{i,j}^{\mathrm{b}} = diag(0.00011, 0.00011, 0.0005)\ \ \mathrm{kg{\cdot}m^2} \tag{55}$$

https://doi.org/10.3390/s26010206

Table 2. The mass parameters of Arm-a.

Link		B_1^a	B_2^a	B_3^a	B_4^a	B_5^a	B_6^a	B_7^a
Mass (kg)		4.24	4.24	8.675	4.24	8.675	4.24	4.24
${}^i\boldsymbol{a}_i$ (m)	x	0	0	0	0	0	0	0
	y	0	0	0	0	0	0	0
	z	0.1275	0	0.95	0	0.95	0	0.1275
${}^i\boldsymbol{b}_i$ (m)	x	0	0	0	0	0	0	0
	y	0	0	0	0	0	0	0
	z	0.1275	0	0.95	0	0.95	0	0.1275
${}^i\boldsymbol{I}_i$ (kg·m^2)	I_{xx}	0.018	0.018	3.846	0.018	3.846	0.018	0.0055
	I_{yy}	0.018	0.018	3.846	0.018	3.846	0.018	0.0055
	I_{zz}	0.0075	0.0075	0.393	0.0075	0.393	0.0075	0.0065
	I_{xy}	0	0	0	0	0	0	0
	I_{xz}	0	0	0	0	0	0	0
	I_{yz}	0	0	0	0	0	0	0

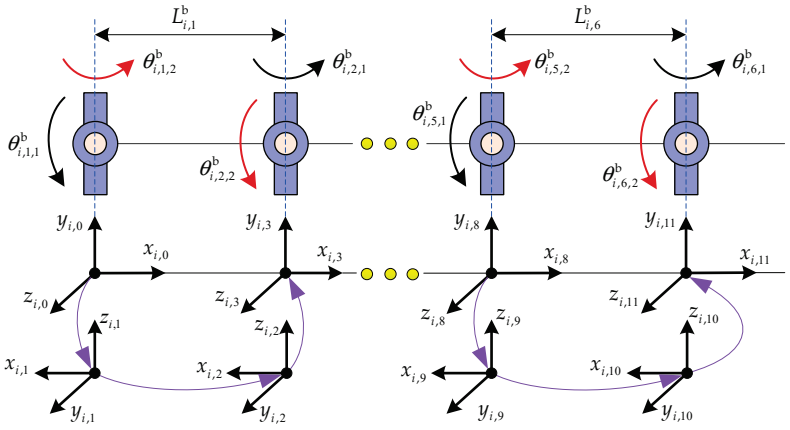


Figure 12. The D-H frames of the ith segment for Arm-b.

Table 3. The D-H parameters of the ith segment for Arm-b.

Segment	Rotation Axis	$\theta_{i,j,k}$ (°)	$\alpha_{i,j,k}$ (°)	$a_{i,j,k}$ (m)	$d_{i,j,k}$ (m)
i	1	0	90	0	0
	2	0	0	$L^b_{i,1} = 0.1$	0
	3	0	−90	0	0
	4	0	0	$L^b_{i,2} = 0.1$	0
	⋮	⋮	⋮	⋮	⋮
	9	0	90	0	0
	10	0	0	$L^b_{i,5} = 0.1$	0
	11	0	−90	0	0
	12	0	0	$L^b_{i,6} = 0.1$	0

The position and attitude (X-Y-Z Euler angle) of the base's centroid with respect to the inertial frame are as follows.

$$\boldsymbol{r}_0 = [-4.71,\ -3.29,\ 70.11]\ \text{mm},\quad \boldsymbol{\Psi}_0 = [1.5^\circ,\ 0.8^\circ,\ 2.1^\circ] \tag{56}$$

The mass and inertia tensor matrix of the base are as follows.

$$m_0 = 3000\ \text{kg},\quad \boldsymbol{I}_0 = diag(3500,\ 3500,\ 3200)\ \text{kg·m}^2 \tag{57}$$

https://doi.org/10.3390/s26010206

The installation position and corresponding attitude (X-Y-Z Euler angle) for Arm-a and Arm-b, which are expressed in the frame attached to the base, are given as follows.

$$\begin{cases} \boldsymbol{b}_0^{\mathrm{a}} = [1.3,\ 0.585,\ -1.15]\mathrm{mm}, & \boldsymbol{\Psi}_{\mathrm{a}} = [-180^\circ,\ 0^\circ,\ -180^\circ] \\ \boldsymbol{b}_0^{\mathrm{b}} = [1.75,\ 0.55,\ 1.3]\mathrm{mm}, & \boldsymbol{\Psi}_{\mathrm{b}} = [0^\circ,\ 0^\circ,\ 0^\circ] \end{cases} \tag{58}$$

5.1.2. The Parameters of the Target Satellite

The mass and inertia tensor matrix of the target satellite are as follows.

$$m_{\mathrm{t}} = 450\ \mathrm{kg}, \quad \boldsymbol{I}_{\mathrm{t}} = \begin{bmatrix} 385.5 & 0.54 & 17.7 \\ 0.54 & 327.75 & 19.65 \\ 17.7 & 19.65 & 230.4 \end{bmatrix} \ \mathrm{kg{\cdot}m^2} \tag{59}$$

When the target satellite is grasped, the position and attitude (X-Y-Z Euler angle) of the target satellite centroid with respect to the frame attached to the last joint of Arm-a are as follows.

$$\boldsymbol{r}_{\mathrm{et}} = [-594.95,\ 9.33,\ 2662.65]\mathrm{mm}, \quad \boldsymbol{\Psi}_{\mathrm{t}} = [-0.0289^\circ,\ -0.0172^\circ,\ 138.6069^\circ] \tag{60}$$

5.2. The Desired Trajectory of the Target Satellite

To achieve target satellite transposition, the desired trajectory of the target satellite centroid is planned. Following Section 4.2, quintic polynomial interpolation ensured smooth acceleration and deceleration, generating the desired trajectory. In the simulation, the initial and final poses of the target satellite centroid are as follows.

$$\begin{cases} \boldsymbol{X}_0 = \begin{bmatrix} -5.3135 & -0.9097 & -1.2121 & -68.70 & -20.31 & 90.76 \end{bmatrix} \\ \boldsymbol{X}_{\mathrm{f}} = \begin{bmatrix} -3.8366 & 2.8592 & -2.0086 & -121.49 & -2.88 & 101.27 \end{bmatrix} \end{cases} \tag{61}$$

Substituting Equation (61) into Equation (17), the variations of position and attitude for the centroid of the target satellite are obtained, as shown in Figure 13.

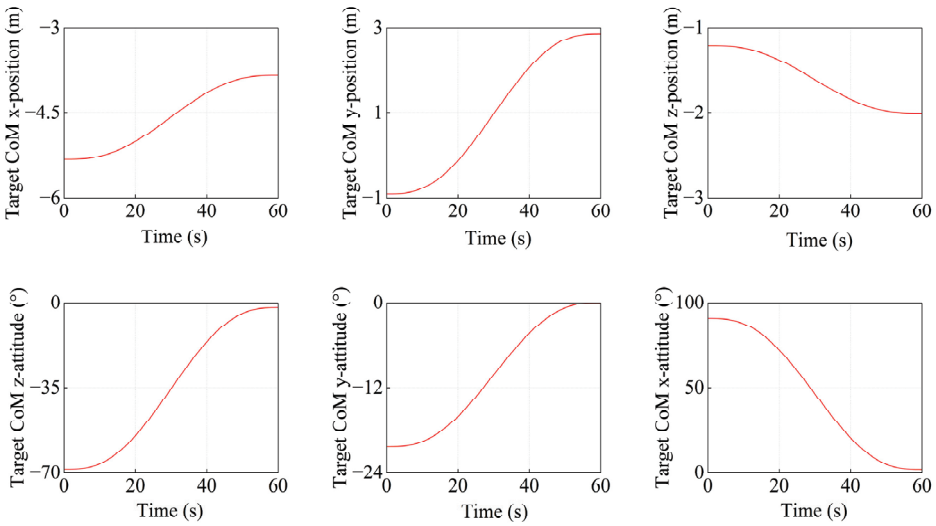


Figure 13. The variation curves of the pose for the target satellite centroid.

5.3. The Simulation for the Pose Stabilization of the Combined System Base

5.3.1. The Simulation Results for the Proposed Method

At the initial moment, the joint angles for Arm-a and Arm-b are set as follows.

https://doi.org/10.3390/s26010206

$$\begin{aligned}\boldsymbol{\Phi}_{\mathrm{a}} &= [1.7486,\ 1.2895,\ -0.1394,\ 1.1473,\ -0.3021,\ -0.9054,\ -0.1692]\mathrm{rad}\\ \boldsymbol{\Phi}_{\mathrm{b}} &= [-0.011,\ 0.006,\ -0.0046,\ 0.0503,\ -0.0072,\ -0.0059,\ -0.0227,\ 0.0263,\ 0.0217,\ 0.1402]\mathrm{rad}\end{aligned} \tag{62}$$

From the target centroid trajectory (as indicated in Figure 13), the desired trajectories of the Arm-a tip and Arm-b tip can be obtained. Furthermore, the joint angles of both arms are calculated by the inverse kinematics solution. The variation curves of the joint angles are shown in Figures 14 and 15, respectively.

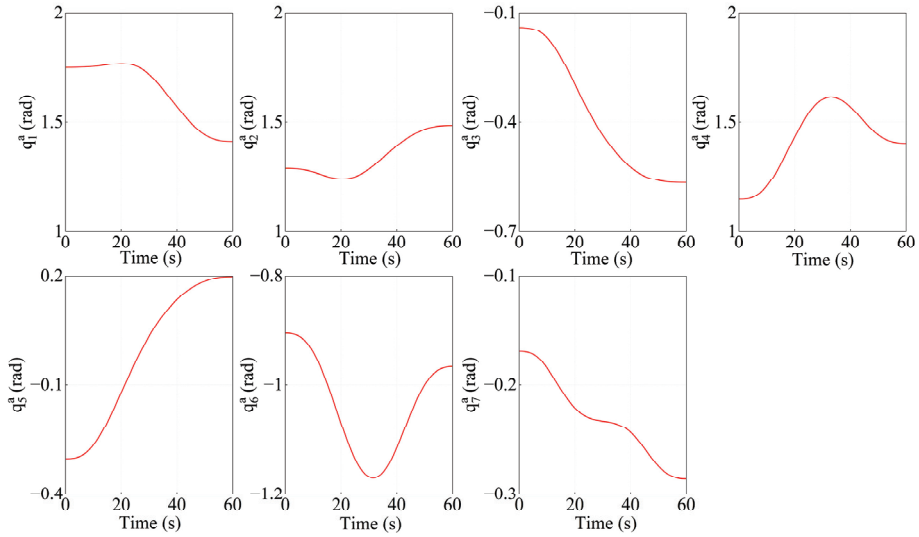


Figure 14. The variation curves of joint angles for Arm-a.

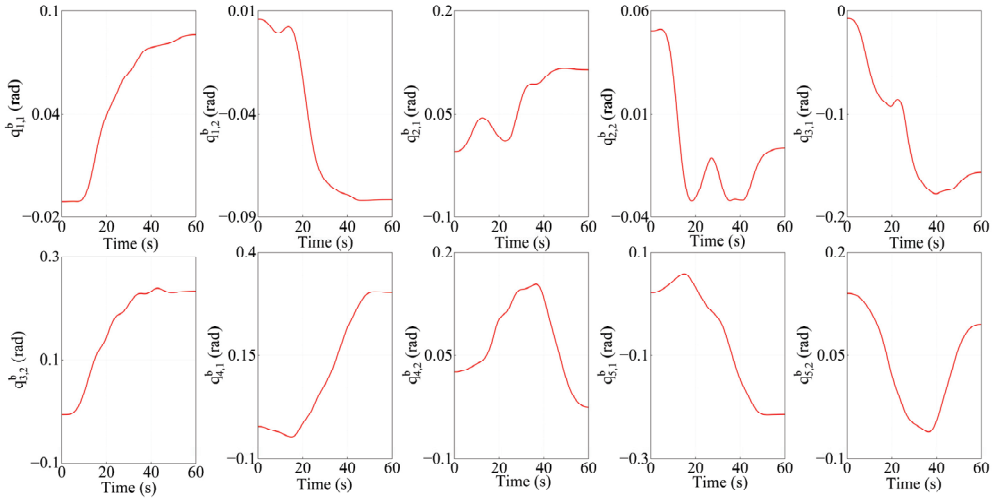


Figure 15. The variation curves of joint angles for Arm-b.

In the simulation, the proportional and derivative coefficient matrices corresponding to the position and attitude of the base are set as follows.

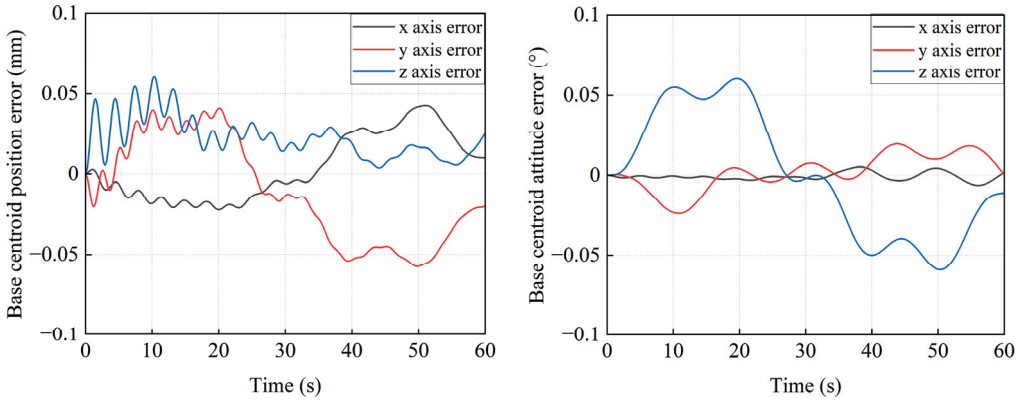
$$\begin{aligned}\boldsymbol{K}_{\mathrm{pp}} &= \begin{bmatrix}12000 & 0 & 0\\ 0 & 19000 & 0\\ 0 & 0 & 14000\end{bmatrix}, \quad \boldsymbol{K}_{\mathrm{vv}} = \begin{bmatrix}300 & 0 & 0\\ 0 & 400 & 0\\ 0 & 0 & 400\end{bmatrix}\\ \boldsymbol{K}_{\mathrm{p}\varphi} &= \begin{bmatrix}6400 & 0 & 0\\ 0 & 5600 & 0\\ 0 & 0 & 6400\end{bmatrix}, \quad \boldsymbol{K}_{\mathrm{v}\omega} = \begin{bmatrix}160 & 0 & 0\\ 0 & 320 & 0\\ 0 & 0 & 320\end{bmatrix}\end{aligned} \tag{63}$$

https://doi.org/10.3390/s26010206

Here, external disturbance forces and torques are introduced to simulate external environment influences, such as solar radiation pressure. The forces and torques are set as follows.

$$\boldsymbol{F}_{\mathrm{r}} = \begin{bmatrix} -0.05, & 0.22, & -0.33 \end{bmatrix}^{\mathrm{T}} \mathrm{N}, \quad \boldsymbol{\tau}_{\mathrm{r}} = \begin{bmatrix} 0.076, & -0.11, & 0.047 \end{bmatrix}^{\mathrm{T}} \mathrm{N{\cdot}m^{-1}} \tag{64}$$

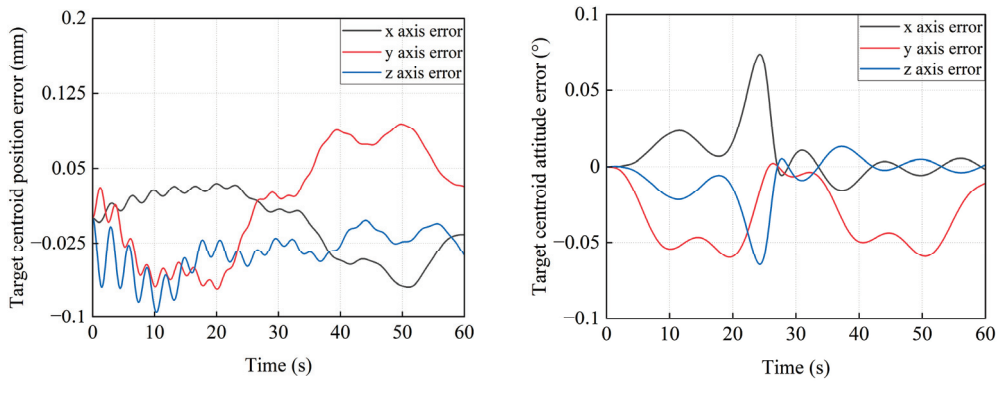
Figure 16 illustrates the base's pose variations during transposition under the proposed control method. The maximum position deviations of the base centroid are 0.043 mm, −0.057 mm, and 0.06 mm along the x_{I}, y_{I}, and z_{I} axes (see Figure 16a), respectively. The maximum deviations of attitude (X-Y-Z Euler angles, Figure 16b) are −0.006°, −0.024°, and 0.06°, respectively. The simulation results indicate that pose error is maintained within 0.1 mm and 0.1° by the proposed method, whose accuracy meets the engineering requirements.



(**a**) (**b**)

Figure 16. The pose error curves for the base centroid of the combined system. (**a**) The position error; (**b**) The attitude error.

Due to base pose deviations, the actual trajectory tracking of the target satellite could deviate from the desired trajectory. The trajectory tracking errors are shown in Figure 17. It can be observed that the maximum position errors of the target satellite centroid are −0.07 mm, 0.094 mm, and −0.094 mm (Figure 17a) corresponding to the x_{I}, y_{I}, and z_{I} axes, while the maximum attitude errors during this process are 0.083°, –0.06°, and −0.072° (Figure 17b), respectively. Both position and attitude errors are limited within 0.1 mm and 0.1°, which further validate the correctness of the proposed method.



(**a**) (**b**)

Figure 17. The trajectory tracking error curves of the target satellite. (**a**) The position tracking error. (**b**) The attitude tracking error.

https://doi.org/10.3390/s26010206

The control forces and torques, which are applied to the base, are shown in Figure 18. They can be generated by thrusters and reaction wheels. Figure 19 shows the 3D simulation models at the following milestones: t = 0 s, 12 s, 24 s, 36 s, 48 s, and 60 s.

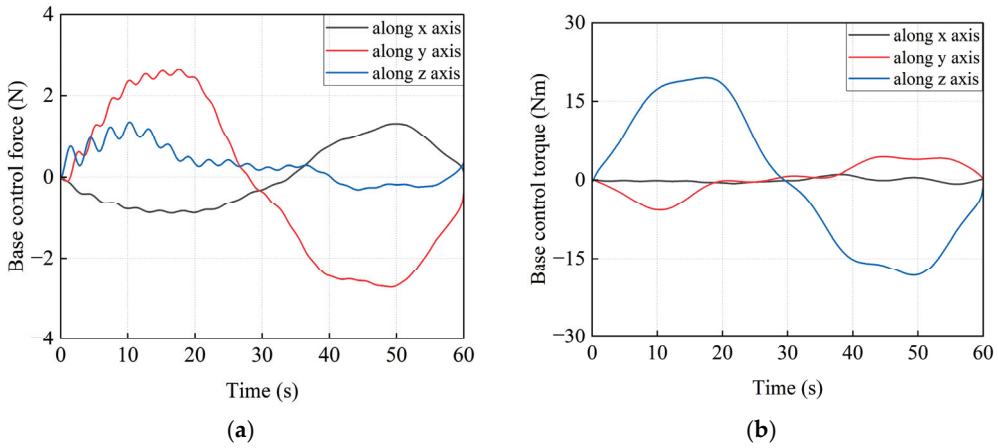


Figure 18. The curves of the control forces and torques. (**a**) The control forces on the base of the combined system. (**b**) The control torques on the base of the combined system.

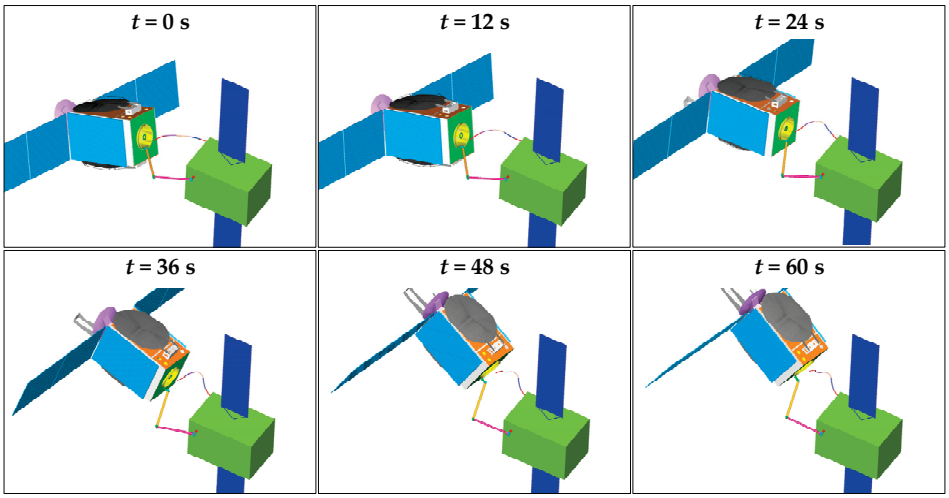
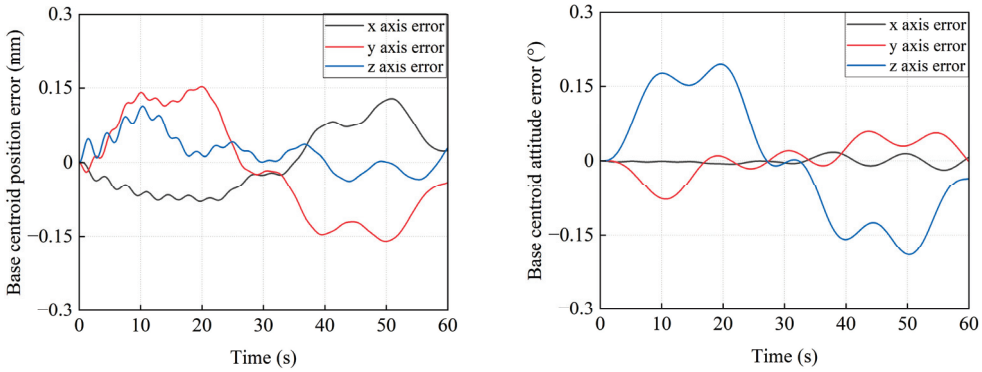


Figure 19. The different states of the combined system in the simulation.

5.3.2. The Simulation Results for the PD Control Method

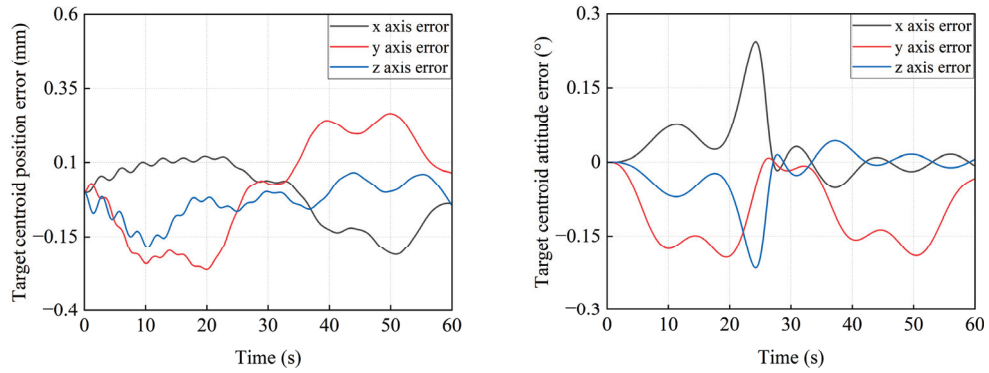
To further validate the effectiveness of the proposed control strategy in stabilizing the base's pose, a comparative simulation is conducted by using the PD control without feedforward compensation. When the traditional PD control method is applied, the pose variations for the base centroid of the combined system are given in Figure 20. As shown in Figure 20a, the simulation results indicate that the maximum position deviations of the base centroid along the x_I, y_I, and z_I axes are 0.1289 mm, −0.1599 mm, and 0.1139 mm, respectively. It can be seen that all errors exceeded 0.1 mm. The maximum attitude deviations of the base centroid are −0.0186°, −0.0785°, and 0.1945° (see Figure 20b), respectively. Notably, the z_I axis error exceeds 0.1°. In addition, the trajectory tracking deviations of the target satellite centroid are shown in Figure 21. It can be concluded that the maximum position tracking errors of the target satellite centroid are −0.2121 mm, 0.2665 mm, and −0.1857 mm, respectively. Meanwhile, they are 0.2442°, 0.1926°, and −0.2142° for the attitude tracking errors. All position errors and attitude errors are more than 0.1 mm and 0.1°, respectively. These results indirectly show the superior control accuracy of the proposed method over the traditional PD control method.

https://doi.org/10.3390/s26010206



(**a**) (**b**)

Figure 20. The pose error curves for the base centroid of the combined system by the PD control method. (**a**) The position error of the base centroid. (**b**) The attitude error of the base centroid.

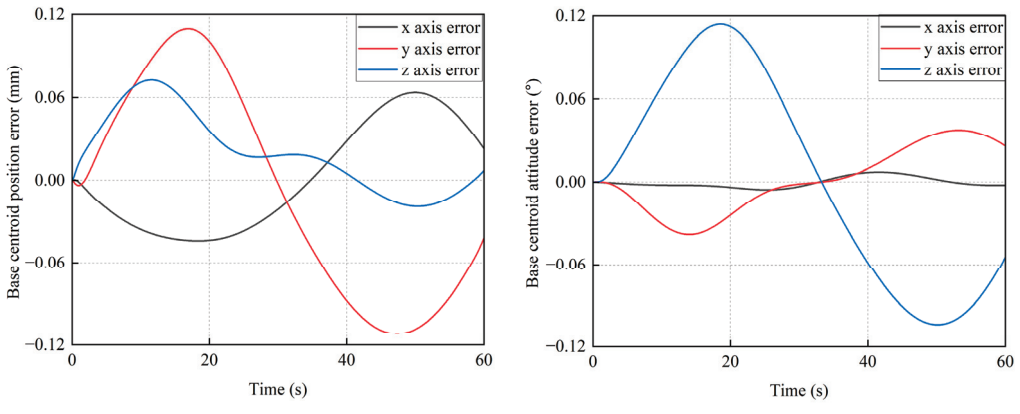


(**a**) (**b**)

Figure 21. The trajectory tracking error curves of the target satellite centroid by the PD control method. (**a**) The position tracking error. (**b**) The attitude tracking error.

5.3.3. The Simulation Results for the Backstepping Method

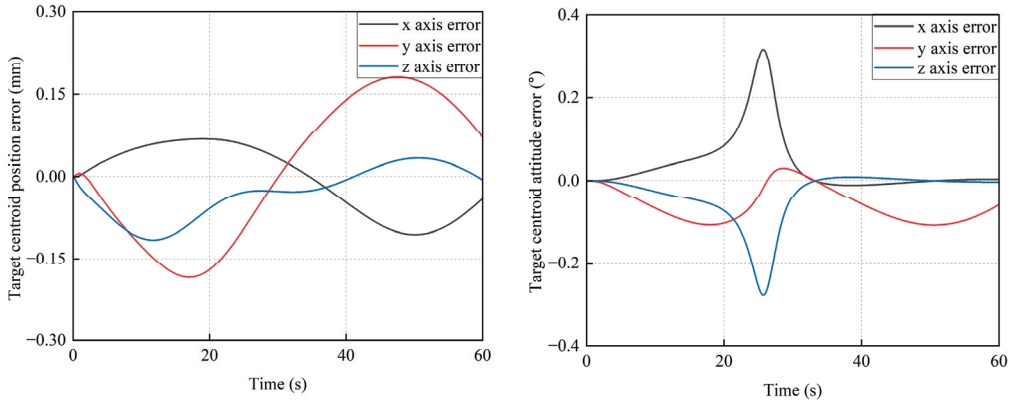
In order to further verify the advantages of the proposed method, another comparison method (i.e., the Backstepping method) is implemented. The pose control errors of the base and the pose tracking error of the target satellite are shown in Figures 22 and 23, respectively.



(**a**) (**b**)

Figure 22. The pose control error curves of the base by the Backstepping method. (**a**) The position error of the base centroid. (**b**) The attitude error of the base centroid.

https://doi.org/10.3390/s26010206



(**a**) (**b**)

Figure 23. The trajectory tracking error curves of the base by the Backstepping method. (**a**) The position tracking error. (**b**) The attitude tracking error.

From Figure 22, it can be seen that the maximum position and attitude control errors of the base along x_I, y_I, and z_I axes are 0.0636 mm, 0.1114 mm, 0.0724 mm, 0.0072°, 0.0381°, and 0.1141°, respectively. As given in Figure 23, the maximum position and attitude tracking errors of the target satellite along x_I, y_I, and z_I axes are 0.1060 mm, 0.1840 mm, 0.1158 mm, 0.3176°, 0.1093°, and 0.2774°, respectively.

5.4. Simulation Results Analysis

The simulation results for the PD control method, the Backstepping method, and the proposed method are listed in Table 4.

Table 4. The simulation results of the two methods.

Item	Coordinate Axis	PD Control	Backstepping	Proposed Method	Improvement over PD	Improvement over Backstepping
Position error of the base	x_I	0.1289	0.0636	0.0430	66.64%	32.39%
	y_I	−0.1599	0.1114	−0.0570	64.35%	48.83%
	z_I	0.1139	0.0724	0.0600	47.32%	17.12%
Attitude error of the base	x_I	−0.0186	0.0072	−0.0063	66.13%	12.50%
	y_I	−0.0785	0.0381	−0.0239	69.55%	37.27%
	z_I	0.1945	0.1141	0.0602	69.05%	47.24%
Position tracking error	x_I	−0.2121	0.1060	−0.0699	67.04%	34.06%
	y_I	0.2665	0.1840	0.0943	64.62%	48.75%
	z_I	−0.1857	0.1158	−0.0949	48.90%	18.05%
Target satellite attitude tracking error	x_I	0.2442	0.3176	0.0738	69.78%	76.76%
	y_I	−0.1926	−0.1093	−0.0599	68.90%	45.20%
	z_I	−0.2142	0.2774	−0.0646	69.84%	76.71%

According to the simulation results, the position control accuracy for the base of the combined system based on the proposed method is improved by 66.64%, 64.35%, and 47.32% along the x_I, y_I, and z_I axes, respectively, compared with the PD control method. The attitude control accuracy of the base is improved by 66.13%, 69.55%, and 69.05%, respectively. Due to the dynamic coupling between the combined system's base and the two manipulators, it will lead to the occurrence of a trajectory tracking error for the target satellite centroid. The proposed method achieves 67.04%, 64.62%, and 48.90% improvement

https://doi.org/10.3390/s26010206

in position tracking accuracy, compared with the PD control method. Additionally, they are 69.78%, 68.90%, and 69.84% for the attitude tracking accuracy.

Compared with the Backstepping method, the accuracy in both position and attitude control of the base are improved by 32.39%, 48.83%, 17.12%, 12.50%, 37.27%, and 47.24%, respectively. The position tracking accuracy of the target satellite are improved by 34.06%, 48.75%, and 18.05%, respectively. They are 76.76%, 45.20% and 76.71% for attitude tracking accuracy.

As a result, it can be concluded that the proposed method significantly improves the pose control precision of the base and the pose tracking accuracy of the target satellite. The effectiveness of the proposed method is thoroughly verified.

Regarding the comparison with the Backstepping method, it should be noted that the Backstepping controller is introduced in this study as a base stabilization strategy, with the same control objective as the conventional PD controller, namely suppressing the base disturbances induced by the manipulator and target satellite motion. The Backstepping method is not designed to directly optimize the attitude tracking performance of the target satellite. For a fair comparison, both controllers are tuned to achieve comparable settling times. In addition, relatively conservative virtual control gains are adopted for the Backstepping controller to ensure stability under strong dynamic coupling and large payload conditions.

In addition, the sensor noise is accounted for by incorporating model parameter uncertainty. The simulations are performed for parameter uncertainty levels of 2%, 5%, and 10%, respectively, to evaluate each scenario. The simulation results are listed in Table 5. It can be concluded that a larger parameter uncertainty leads to greater control errors. Furthermore, with a parameter uncertainty of 15%, the maximum position control error of the base is less than 0.09 mm, and the attitude error is below 0.11°. Meanwhile, for target satellite transposition, the position tracking error is less than 0.15 mm, and the attitude error remains under 0.12°.

Table 5. The simulation results under the condition of parameter uncertainty.

Item	Coordinate Axis	2%	5%	10%	15%
Position error of the base	x_I	0.0461	0.0510	0.0589	0.0664
	y_I	−0.0614	−0.0672	−0.0766	−0.0854
	z_I	0.0624	0.0655	0.0704	0.0750
Attitude error of the base	x_I	−0.0068	−0.0075	−0.0086	−0.0096
	y_I	−0.0261	−0.0291	−0.0341	−0.0388
	z_I	0.0651	0.0728	0.0849	0.0963
Position tracking error of the target satellite	x_I	−0.0755	−0.0836	−0.0966	−0.1089
	y_I	0.1010	0.1110	0.1261	0.1413
	z_I	−0.0985	−0.1037	−0.1120	−0.1198
Attitude tracking error of the target satellite	x_I	0.0802	0.0895	0.1045	0.1187
	y_I	−0.0649	−0.0724	−0.0843	−0.0956
	z_I	−0.0702	−0.0784	−0.0916	−0.1041

In this analysis, the parametric uncertainty scenarios listed in Table 5 primarily focus on variations in the system mass and inertial properties. It is further clarified that sensor noise effects, which introduce measurement errors in the feedback loop and in the estimation of external disturbance forces and torques, are not explicitly modeled as stochastic processes, but are instead implicitly reflected through bounded uncertainty and equivalent disturbance terms within the adopted parametric uncertainty framework.

https://doi.org/10.3390/s26010206

6. Discussion

A more comprehensive discussion of the challenges and simplifications associated with real-world implementation is given, including time delay, computational cost, and actuator saturation. Regarding the time delay issue, system delays from sensor sampling, communication, and computation could introduce phase errors, particularly degrading the performance of the model-dependent feedforward compensation. This would place a greater burden on the robustness of the PD feedback loop. We will study some strategies like predictive filtering or high-speed data protocols to address this issue. On the problem of computational cost, the real-time execution of the recursive Newton–Euler algorithm for feedforward dynamics, while efficient ($o(n)$), must be rigorously validated on representative flight processor hardware. About the actuator saturation, the physical limits of the thrusters and reaction wheels are a critical constraint. Each actuator requires integrating an actuator saturation management block and a control allocation strategy to dynamically distribute torque between actuators, preventing performance degradation or instability when commanded values exceed physical limits. It should be noted that the implementation-related aspects, including computation time per Newton–Euler recursion iteration, delay tolerance, and actuator saturation limits, are intended as qualitative considerations derived from simulation analysis. These discussions aim to provide preliminary insights into the practical feasibility of the proposed control strategy, rather than to report quantitatively validated hardware-level performance metrics. A detailed quantitative evaluation of these aspects requires hardware implementation or hardware-in-the-loop experiments, which are beyond the scope of the present study and will be addressed in future studies.

The base pose stabilization control method proposed in this study has so far been validated only in a simulation environment, and the analysis results are obtained under a set of idealized modeling assumptions. Specifically, the mass and geometric parameters of the system are assumed to be accurately known, while non-ideal factors such as joint friction, structural flexibility, and mechanical backlash are neglected in the simulation. Consequently, the simulation results mainly reflect the control performance of the proposed method in suppressing base dynamic coupling disturbances induced by large payload motions under ideal conditions. The advantage of simulation-based studies lies in their ability to evaluate the effectiveness and stability of the proposed strategy in a controllable and repeatable environment, which facilitates the investigation of system dynamics and the underlying control mechanisms. However, in practical physical implementations, factors such as sensor noise, actuator constraints, computational delays, and modeling inaccuracies may adversely affect control performance. To address these issues, future work will incorporate parameter uncertainties, joint friction, and structural flexibility, and will further assess the engineering applicability of the proposed method through higher-fidelity simulations or experimental platforms.

7. Conclusions

This paper proposed a pose stabilization algorithm based on feedforward compensation PD control to stabilize the base of the combined system during target satellite transposition. The overall transposition mission was analyzed to define the roles of each manipulator. Then, a dynamics model of the combined system was established to calculate the feedforward compensation term. A pose stabilization strategy combing feedforward compensation and PD control was developed to suppress disturbances to the base. Finally, to verify the proposed method, a co-simulation model was established byMatlab2016 and Adams2018 software. Comparative simulations were conducted for the proposed method and the traditional method. By incorporating the reaction forces and torques exerted on the base as feedforward compensation terms in the control law, the proposed method effec-

https://doi.org/10.3390/s26010206

tively mitigates the strong dynamic coupling effects during the target satellite transposition process. As a result, the errors of the base's position and attitude are reduced significantly, and the trajectory tracking accuracy is improved for the transposition operation. Simulation results indicate that the proposed method exhibits good applicability for base stabilization control under large payload conditions, providing a feasible solution for maintaining base pose stability during the target satellite transposition process.

The current research only focuses on pose stabilization of the base for the combined system that is composed of a heterogeneous multi-arm space robot and a target satellite. The application of these methods for heterogeneous multi-arm systems has not been explored. Its dynamic recursive formulation does not account for joint friction or mechanical clearance. Furthermore, the robustness validation in this study is restricted to deterministic bounded-uncertainty scenarios rather than statistical characterization. Future work will incorporate stochastic analysis methods, such as Monte Carlo simulations, to statistically evaluate the control performance under random measurement noise and time-varying disturbances. In addition, the minimum energy optimization for target satellite transposition is also not considered in this paper. In the future, we will strive to do research on these topics.

Author Contributions: Conceptualization, Z.H. and D.M.; methodology, J.Y., W.X. and Z.H.; software, J.Y., H.C. and Z.H.; validation, Z.H., J.Y. and L.X.; formal analysis, Z.H. and W.X.; investigation, D.M. and Z.H.; resources, H.C. and D.M.; data curation, J.Y., W.X. and H.C.; writing—original draft preparation, Z.H., J.Y. and L.X.; writing—review and editing, W.X., H.C. and D.M.; visualization, Z.H., L.X. and W.X.; supervision, H.C., Z.H. and J.Y.; project administration, W.X., Z.H., L.X. and D.M.; funding acquisition, Z.H., L.X. and D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (62303205), the Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515030029, project ZR2023QF137 supported by the Shandong Provincial Natural Science Foundation, project ZR2024QE464 supported by the Shandong Provincial Natural Science Foundation, and the Doctoral Startup Fund of Liaocheng University (318052231).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, W.; Cheng, D.; Liu, X.; Wang, Y.; Shi, W.; Tang, Z.; Gao, F.; Zeng, F.; Chai, H.; Luo, W.; et al. On-orbit service (OOS) of spacecraft: A review of engineering developments. *Prog. Aeosp. Sci.* **2019**, *108*, 32–120. [CrossRef]
2. Fallahiarezoodar, N.; Zhu, Z. Review of autonomous space robotic manipulators for on-orbit servicing and active debris removal. *Space Sci. Technol.* **2025**, *5*, 291. [CrossRef]
3. Moghaddam, B.M.; Chhabra, R. On the guidance, navigation and control of in-orbit space robotic missions: A survey and prospective vision. *Acta Astronaut.* **2021**, *184*, 70–100. [CrossRef]
4. Kang, G.; Zhang, Q.; Wu, J.; Zhang, H. PD-Impedance Combined Control Strategy for Capture Operations Using a 3-DOF Space Manipulator with a Compliant End-Effector. *Sensors* **2020**, *20*, 6739. [CrossRef] [PubMed]
5. Wang, T.; Ma, G.; Xu, L.; Yu, R. CP-QRRT*: A Path Planning Algorithm for Hyper-Redundant Manipulators Considering Joint Angle Constraints. *Sensors* **2025**, *25*, 1490. [CrossRef]
6. Robinson, G.; Davies, J.B.C. Continuum robots-a state of the art. In Proceedings of the IEEE International Conference on Robotics and Automation, Detroit, MI, USA, 10–15 May 1999; pp. 2849–2854.
7. Hu, Z.; Li, C.; Sun, Q.; Peng, J.; Li, W. Coordinated trajectory planning of discrete-serpentine heterogeneous multi-arm space robot for capturing tumbling targets using manipulability optimization. *Aerospace* **2025**, *12*, 944. [CrossRef]

https://doi.org/10.3390/s26010206

8. Li, S.; She, Y. Recent advances in contact dynamics and post-capture control for combined spacecraft. *Prog. Aeosp. Sci.* **2021**, *120*, 100678. [CrossRef]
9. Estable, S.; Pruvost, C.; Ferreira, E.; Telaar, J.; Fruhnert, M.; Imhof, C.; Rybus, T.; Peckover, G.; Lucas, R.; Ahmed, R.; et al. Capturing and deorbiting Envisat with an Airbus Spacetug. Results from the ESA e.Deorbit consolidation phase study. *J. Space Saf. Eng.* **2020**, *7*, 52–66. [CrossRef]
10. Olivares-Mendez, M.; Makhdoomi, M.R.; Yalçın, B.C.; Bokal, Z.; Muralidharan, V.; Ortiz Del Castillo, M.; Gaudilliere, V.; Pauly, L.; Borgue, O.; Alandihallaj, M.; et al. Zero-G Lab: A multi-purpose facility for emulating space operations. *J. Space Saf. Eng.* **2023**, *10*, 509–521. [CrossRef]
11. Muralidharan, V.; Makhdoomi, M.R.; žinys, A.; Razgus, B.; Klimavičius, M.; Olivares-Mendez, M.; Martinez, C. On-ground validation of orbital GNC: Visual navigation assessment in robotic testbed facility. *Astrodynamics* **2025**, *9*, 343–367. [CrossRef]
12. Papadopoulos, E.; Dubowsky, S. Coordinated manipulator/spacecraft motion control for space robotic systems. In Proceedings of the IEEE International Conference on Robotics and Automation, Sacramento, CA, USA, 9–11 April 1991; pp. 1696–1701.
13. Yamada, K. Attitude control of space robot by arm motion. *J. Guid. Control Dyn.* **1994**, *17*, 1050–1054. [CrossRef]
14. Zhan, B.; Jin, M.; Liu, J. Extended-state-observer-based adaptive control of flexible-joint space manipulators with system uncertainties. *Adv. Space Res.* **2022**, *69*, 3088–3102. [CrossRef]
15. Zong, L.; Emami, M.R.; Luo, J. Reactionless control of free-floating space manipulators. *IEEE Trans. Aerosp. Electron. Syst.* **2020**, *56*, 1490–1503. [CrossRef]
16. Gong, K.; Jia, Y.; Jia, Y. Reactionless control of free-floating space manipulators with parameter uncertainty and input disturbance. *IET Contr. Theory Appl.* **2024**, *18*, 960–975. [CrossRef]
17. Xu, S. Passivity based adaptive extended Jacobian control for free-floating space manipulators. In Proceedings of the International Conference on Signal and Information Processing, Networking and Computers (ICSINC), Chongqing, China, 10–13 September 2024; pp. 329–337.
18. Kuck, E.; Sands, T. Space robot sensor noise amelioration using trajectory shaping. *Sensors* **2024**, *24*, 666. [CrossRef] [PubMed]
19. Raina, D.; Gora, S.; Maheshwari, D.; Shah, S.V. Impact modeling and reactionless control for post-capturing and maneuvering of orbiting objects using a multi-arm space robot. *Acta Astronaut.* **2021**, *182*, 21–36. [CrossRef]
20. Wang, M.; Luo, J.; Yuan, J.; Walter, U. Detumbling strategy and coordination control of kinematically redundant space robot after capturing a tumbling target. *Nonlinear Dyn.* **2018**, *92*, 1023–1043. [CrossRef]
21. Chang, R.; Jia, Q.; Chu, M. Stabilization control for spacecraft-manipulator system after capturing tumbling target. *Int. J. Control. Autom. Syst.* **2022**, *20*, 3706–3717. [CrossRef]
22. Woodward, N.; Bevilacqua, R. Attitude control and online estimation of unknown mass properties for captured debris with a flexible spacecraft. *Acta Astronaut.* **2025**, *232*, 191–203. [CrossRef]
23. Zhang, T.; Shi, P.; Guo, M.; Yue, X. Enhanced rapid attitude stabilization for space robots in postcapture using an iterated extended Kalman filter. *Adv. Space Res.* **2023**, *72*, 4965–4983. [CrossRef]
24. Huang, X.; Biggs, J.D.; Duan, G. Post-capture attitude control with prescribed performance. *Aerosp. Sci. Technol.* **2020**, *96*, 105572. [CrossRef]
25. Song, T.; Zheng, Z.; Guo, Y.; Yuan, J. Application of model-free control to the operation of post-capture combined spacecraft. *Aerospace* **2023**, *10*, 90. [CrossRef]
26. Liu, Y.; Liu, X.; Cai, G.; Chen, J. Trajectory planning and coordination control of a space robot for detumbling a flexible tumbling target in post-capture phase. *Multibody Syst. Dyn.* **2021**, *52*, 281–311. [CrossRef]
27. Mao, R.; Meng, T.; Wang, K.; Lei, J.; Wang, W. Deep Koopman-operator-based model predictive control for free-floating space robots with disturbance observer. *Aerosp. Sci. Technol.* **2024**, *154*, 109515. [CrossRef]
28. Wang, S.; Cao, Y.; Zheng, X.; Zhang, T. A learning system for motion planning of free-float dual-arm space manipulator towards non-cooperative object. *Aerosp. Sci. Technol.* **2022**, *131*, 107980. [CrossRef]
29. Dong, Z.; Wang, L.; Zhu, Y.; Liu, L.; Guo, J.; Guo, Y. Motion planning of free-flying space robots for tracking relative motion tumbling targets by sequential optimization via reinforcement learning. *Acta Astronaut.* **2025**, *235*, 463–474. [CrossRef]
30. Ma, Z.; Wang, Y.; Yang, Y.; Wang, Z.; Tang, L.; Ackland, S. Reinforcement Learning-Based Satellite Attitude Stabilization Method for Non-Cooperative Target Capturing. *Sensors* **2018**, *18*, 4331. [CrossRef] [PubMed]
31. Hu, Z.; Yuan, H.; Xu, W.; Yang, T.; Liang, B. Equivalent kinematics and pose-configuration planning of segmented hyper-redundant space manipulators. *Acta Astronaut.* **2021**, *185*, 102–116. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

https://doi.org/10.3390/s26010206

sensors

MDPI

Article

Kinematic Modeling and Solutions for Cable-Driven Parallel Robots Considering Adaptive Pulley Kinematics

Zhonghua Hu [1,*], Chaowen Deng [1], Kai Wang [2] and Jianqing Peng [3,4,*]

[1] School of Mechanical & Automotive Engineering, Liaocheng University, Liaocheng 252000, China; dcw_lcu_edu@163.com
[2] College of Engineering, China Agricultural University, Beijing 100083, China; wkcau@cau.edu.cn
[3] School of Intelligent Systems Engineering, Shenzhen Campus, Sun Yat-sen University, Shenzhen 518107, China
[4] Guangdong Provincial Key Laboratory of Fire Science and Technology, Guangzhou 510275, China
* Correspondence: huzhonghua@lcu.edu.cn (Z.H.); pengjq7@mail.sysu.edu.cn (J.P.)

Abstract

Although the use of adaptive pulleys enhances the motion characteristics of cable-driven parallel robots (CDPRs), it significantly increases the complexity of the kinematics model. Conventional methods often fail to account for the influence of adaptive pulley motion on cable length variation, making it difficult to establish a precise kinematics model. To deal with the problem, this study presents a kinematic modeling and solution method for CDPRs, which incorporates adaptive pulley kinematics. First, the structural design of the CDPR driven by eight cables is analyzed. Then, the generalized kinematics model and the improved kinematics model with adaptive pulley considerations are established. Furthermore, a hybrid Levenberg–Marquardt and Genetic algorithm is proposed to achieve the efficient and high-precision solution of kinematics equations by combining the rapid global search and precise local optimization. Finally, the proposed method is validated through straight path simulation and elliptical path simulation. The simulation results indicate that the tracking accuracy of the end-effector is better than the 1×10^{-7} level for the proposed method.

Keywords: cable-driven parallel robot; kinematic modeling; adaptive pulley kinematics; kinematic solution; hybrid Levenberg–Marquardt and Genetic algorithm

1. Introduction

The cable-driven parallel robot (CDPR) is a typical parallel mechanism. Unlike traditional rigid-link parallel robots, the end-effector of the CDPR is suspended by a flexible cables, which allows the end-effector to move rapidly in the space [1–3]. In addition, CDPRs have many advantages over the rigid-link parallel robots in terms of their high load capacity, large workspace and light overall inertia [4,5]. Therefore, CDPRs are widely used in challenging tasks, such as cargo handling [6], microgravity environment simulation [7], medical rehabilitation [8], 3D printing [9,10], and many other fields. Because of this, the research on CDPR design, modeling and motion planning is gaining attention among scholars.

A CDPR is usually composed of a fixed platform, an end-effector and several cables. Generally, one end of each cable is connected to the actuator on the fixed platform, and the other end is attached to the mobile platform [11]. Based on the number of cables and degree of freedom (DoF), CDPRs can be subdivided into under-constrained CDPRs (UCCDPRs)

and fully-constrained CDPRs (FCCDPRs) [2]. If the number of cables is less than that of the end-effector's DoFs, these robots are UCCDPRs; otherwise, they are FCCDPRs. Mattioni et al. [12] designed an overconstrained planar cable-driven parallel robot to perform quasi-static non-contact operations on the plane. Job et al. [13] established the mathematical model of an UCCDPR with four cables to describe its configuration. Based on this, the range of feasible directions for the UCCDPR end-effector at different positions of the workspace was analyzed. Hwang et al. [14] designed a UCCDPR supported by n cables and proposed a novel method to generate the equilibrium configuration trajectory of the UCCDPR. For the FCCDPRs, Gagliardini et al. [15] developed a cable-driven parallel robot for industrial purposes. Moreover, the minimum size of the robot was determined by the base anchor position of each cable. To accomplish pick-and-place objects with heavy weight in large workspaces, Lin et al. [16] designed a multiple-degrees-of-freedom cable-suspended robot with a gripper. Wu et al. [17] proposed a reconfigurable cable-driven parallel robot to avoid collision and improve the flexibility of the robot. Although UCCDPRs and FCCDPRs have been both studied, FCCDPRs are a research hotspot currently due to their higher stiffness and higher dynamic characteristics compared with UCCDPRs. Therefore, this study mainly focuses on the FCCDPRs.

In the structure design of FCCDPRs, one or more pulleys are used for guiding cables and reducing abrasion. Therefore, it is highly necessary to consider the influence of the pulley factor when conducting research into FCCDPRs. Zhang et al. [18] established the kinematics of an FCCDPR, considering pulleys, and they further developed its error model and kinematic calibration method. To significantly increase the size of the wrench-feasible workspace, Martin-Parra et al. [19] designed a novel fully-constrained planar robot with three DoFs and proposed an optimization approach to solve the kineto-static equations of the robot. Reference [20] presented the conceptual design of an FCCDPR with pulleys on its base platform and established the error mapping model considering pulleys. Although the above studies considered the influence of the pulley on the parallel robot, they focused on the error estimation and kinematic calibration of the robot and failed to achieve a kinematic solution directly.

In fact, the kinematics solution issues of FCCDPRs are the inverse kinematics (IK) and the forward kinematics (FK). The IK problem refers to when the pose of the end-effector is given, and then the length of each cable is calculated. In contrast, the FK issue arises when obtaining the pose of the end-effector when the cable lengths are known. For the inverse kinematics of FCCDPRs, there is a direct unique solution, especially without considering the pulley factor [21,22]. Zi et al. [23] discussed the inverse kinematics and statics of a rehabilitation CDPR. Based on this, the optimal design was achieved for the robot's waist twist and lower limb traction devices. At present, the inverse kinematics approach is always used to analyze the workspace of FCCDPRs [24,25].

Unlike the IK problem, it is challenging to establish the forward kinematics of FCCDPRs, owing to over-determination with respect to the distribution of cables. As the number of cables in the FCCDPRs increases, the challenges become greater. To address this issue, researchers have conducted extensive studies on the forward kinematics of FCCDPRs. Pott and Schmidt [26] formulated forward kinematics as an optimization problem in which the optimal solution was obtained by minimizing the potential energy of cables. To solve the direct kinematics of a CDPR with deformable cables, Merlet [27] developed a generic numerical continuation scheme. In this method, the cable model depended on physical parameters, and it was formulated such that a deformable cable would simplify to a non-deformable one at the parameter limits. For a sagging CDPR, Briot and Merlet [28] converted the problem of solving forward kinematics into a fixed-time Bolza problem. Additionally, they demonstrated that singularities of the forward kinematics model of

https://doi.org/10.3390/s26010039

the CDPRs were the limits of stability. With respect to the same problem in [28], Baskar et al. [29] reformulated the Irvine cable model to solve the global forward kineto-statics of large CDPRs with sagging cables. Here, an iterative strategy was proposed for the root accumulation that employed random monodromy loops. This method tracked circuits instead of lone paths, thereby potentially identifying multiple roots for nonlinear, non-algebraic, and complex-analytic systems. Based on the kinematic solution, the control strategies of CDPRs have attracted more attention among researchers [30,31]. This indicates that the kinematic solution of CDPRs is a core issue.

As discussed above, the existing research commonly ignores the influence of adaptive pulleys during kinematics modeling, despite their significant effect on model accuracy. In addition to this, the mathematical model for the forward kinematics essentially constitutes a system of overdetermined equations, which poses a significant challenge for obtaining efficient solutions. To tackle this issue, the kinematics model of the cable-driven parallel robot was first established, in which the effect of the adaptive pulley on cable length variation has been considered. Furthermore, the kinematic solution method based on the hybrid Levenberg–Marquardt and Genetic algorithm is proposed for the cable-driven parallel robot. Finally, the straight trajectory and elliptic trajectory are adopted as the desired paths of the end-effector for verifying the correctness and effectiveness of the proposed method.

The remainder of this study is organized as follows. Section 2 analyzes the overall structure and design principle of the cable-driven parallel robot. In Section 3, the generalized kinematics model and the improved kinematics model with adaptive pulley considerations are both established, and the forward or inverse kinematics equations were obtained. Based on the hybrid Levenberg–Marquardt and Genetic algorithm, Section 4 proposes an effective kinematic solution method. In Section 5, the proposed method is verified through the simulation of a straight path and an elliptical path. Based on the analysis in the previous sections, the last section summarizes the study.

2. Design of the Parallel Robot

2.1. Overall Structural Design

The mechanical structure of the cable-driven parallel robot (CDPR) is depicted in Figure 1. The robot comprises a frame, an end-effector, and eight motion control modules. Each module has the same structure and consists of a servo motor, torque sensor, cable winch reel, limit and tension pulley, and adaptive pulley. These modules are symmetrically arranged at the eight vertices of the frame. The servo motor serves as the power source for the motion control module. The motor's output torque is transmitted to the cable winch reel and drive it to rotate. The torque sensor continuously monitors the actuator's output torque and sends real-time data to the control unit, ensuring the control accuracy of cables. Each cable tip is fixed to the end-effector, enabling its translation and rotation through the coordinated release and retraction of the cables. Due to the symmetrical structural design of the robot, its space utilization, control accuracy, stiffness, and load capacity are significantly enhanced. The adaptive pulley is a core component of the motion control module, capable of rotating around the central axis of its base in response to the movement of the end-effector. This rotational ability enables the cable to move synchronously with the end-effector, ensuring its translation and rotation around in three-dimensional space. The structural design principles and the component function of the CDPR are given in Figure 2.

https://doi.org/10.3390/s26010039

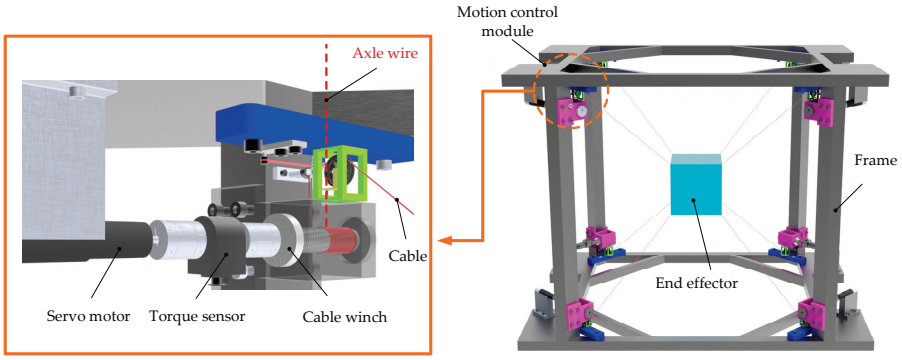


Figure 1. The overall structure of the cable-driven parallel robot.

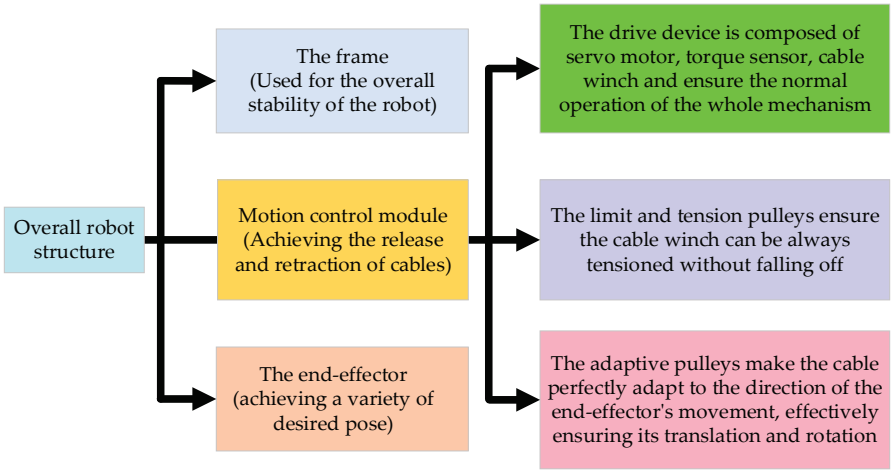


Figure 2. The design scheme of the CDPR.

2.2. Operating Principle of Motion Control Module

As shown in Figure 3, the working principle of the cable-driven parallel robot is as follows. The signal from the upper computer is received by servo motors. Then, the servo motor starts and the cable winch reel connected to the servo motor rotates in synchronization with it. The release and retraction of the cable are achieved through the rotation of the cable winch reel. Initially, the direction of the cable is determined by the limit pulley 1, after which the tension pulley enables the cable to maintain tension.

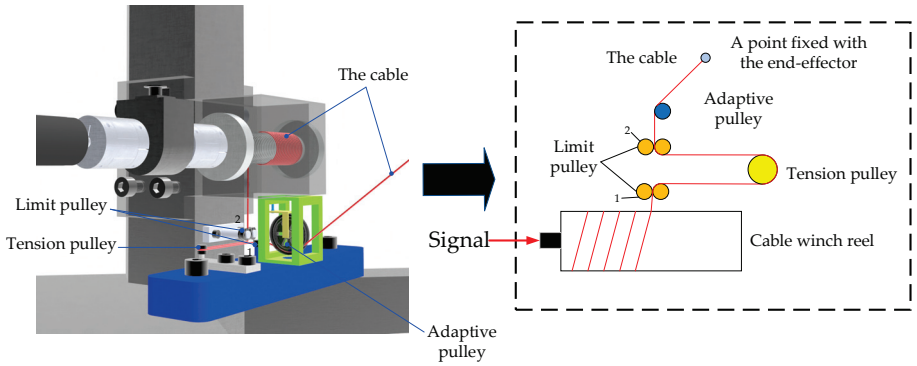


Figure 3. The schematic of the cable path layout.

Simultaneously, the limit pulley 2 further guarantees the cable's orientation. For precisely controlling the cable's direction, an adaptive pulley is equipped behind the limit pulley 2. After passing through the adaptive pulley, the cable is connected to the end-

https://doi.org/10.3390/s26010039

effector of the CDPR. As a result, the desired motion of the end-effector can be achieved by the pulling and releasing cables.

3. Kinematics Modeling of the Robot

The kinematics of the cable-driven parallel robot consists of two parts: the inverse solution of the kinematics and the forward solution of the kinematics. Inverse kinematics means that the position and attitude of the end-effector are known, and then the cable length corresponding to the above pose can be calculated. Forward kinematics involves calculating the position and attitude of the end-effector based on each articulated point position of the robot and its corresponding cable length. In the kinematics analysis of the CDPR, the characteristics of the adaptive pulley model are considered for improving the accuracy of kinematics modeling. In this section, the generalized kinematics model and the improved kinematics model with adaptive pulley considerations are studied. Before the modeling analysis, some assumptions are given as follows.

(a) The cable is handled as a straight line.
(b) The effect of gravity on the cable is neglected.
(c) The cable is in a tensioned state at all times.

3.1. Generalized Kinematics Model

The simplified structure model of the cable-driven parallel robot is shown in Figure 4, which mainly consists of a fixed frame, the end-effector and eight cables. The cube A_1~A_8 represents the fixed frame of the CDPR, and each vertex A_i (i = 1, 2, . . . , 8) is the point of cable exit. The other cube B_1~B_8 is the end-effector of the CDPR and its vertices B_i (i = 1, 2, . . . , 8) are connected with the other end of cables (i.e., the ith cable connects A_i and B_i). The inertial system $\{X_I, Y_I, Z_I\}$ is denoted as Σ_I, which is fixed on the bottom plane center of the fixed frame, and its origin is O_I. $\{x_e, y_e, z_e\}$ (named by Σ_e) is the moving coordinate system, which is fixed in the geometric center of the end-effector. The position and attitude of Σ_e changes with the movement of the end-effector.

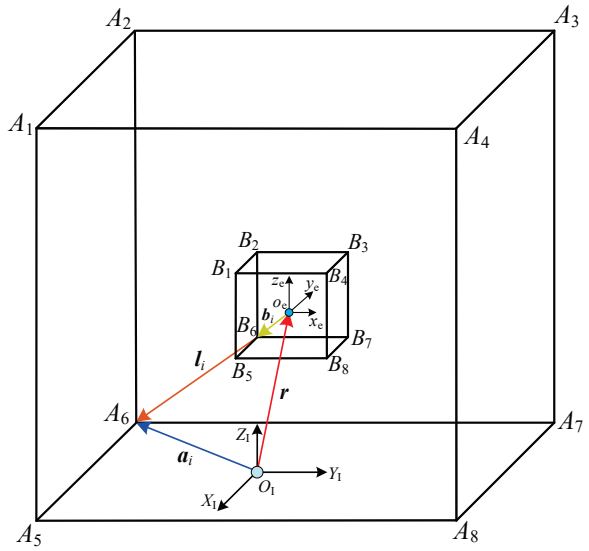


Figure 4. The general model of the cable-driven parallel robot.

As given in Figure 4, based on the closed vector principle, the ith cable vector $\boldsymbol{l}_i$ is obtained as follows.

$$\boldsymbol{l}_i = \boldsymbol{a}_i - \boldsymbol{r} - \boldsymbol{R}\boldsymbol{b}_i \tag{1}$$

where $\boldsymbol{a}_i$ is the position vector of point A_i in the inertial system. $\boldsymbol{b}_i$ is the position vector of point B_i which is connected with the other end of the ith cable and $\boldsymbol{b}_i$ is expressed in Σ_e. $\boldsymbol{r}$ is

https://doi.org/10.3390/s26010039

the position vector from the origin of Σ_I to the origin of Σ_e. $\boldsymbol{R}$ is the attitude transformation matrix from Σ_e to Σ_I.

The end-effector's pose (position and attitude) is defined as $Q = (X_e, Y_e, Z_e, \psi_1, \psi_2, \psi_3)$. X_e, Y_e, and Z_e are the position coordinates in Σ_I. ψ_1, ψ_2, and ψ_3 represent attitude angles (X–Y–Z Euler angles). Hence, $\boldsymbol{r}$ and $\boldsymbol{R}$ can be calculated as follows.

$$\begin{cases} \boldsymbol{r} = \begin{bmatrix} X_e & Y_e & Z_e \end{bmatrix} \\ \boldsymbol{R} = \boldsymbol{R}_{xyz}(\psi_1, \psi_2, \psi_3) = \begin{bmatrix} c_{\psi_2} c_{\psi_3} & -c_{\psi_2} s_{\psi_3} & s_{\psi_2} \\ s_{\psi_1} s_{\psi_2} c_{\psi_3} + c_{\psi_1} s_{\psi_3} & c_{\psi_1} c_{\psi_3} - s_{\psi_1} s_{\psi_2} s_{\psi_3} & -s_{\psi_1} c_{\psi_2} \\ s_{\psi_1} s_{\psi_3} - c_{\psi_1} s_{\psi_2} c_{\psi_3} & c_{\psi_1} s_{\psi_2} s_{\psi_3} + s_{\psi_1} c_{\psi_3} & c_{\psi_1} c_{\psi_2} \end{bmatrix} \end{cases} \tag{2}$$

where $s_{\psi_i} = \sin(\psi_i), c_{\psi_i} = \cos(\psi_i)$, $i = 1, 2, 3$.

According to Equation (2), the length of the ith cable is calculated as follows.

$$l_i = \|\boldsymbol{l}_i\| = \sqrt{(\boldsymbol{a}_i - \boldsymbol{r} - \boldsymbol{R}\boldsymbol{b}_i)^{\mathrm{T}}(\boldsymbol{a}_i - \boldsymbol{r} - \boldsymbol{R}\boldsymbol{b}_i)} \tag{3}$$

Setting $\boldsymbol{l} = [l_1, l_2, l_3, l_4, l_5, l_6, l_7, l_8]$, the mapping relationship between the cable lengths and the end-effector's pose can be expressed as:

$$\begin{cases} [\boldsymbol{r},\ \boldsymbol{R}] = \boldsymbol{FK}^{GK}(\boldsymbol{l}, \boldsymbol{a}_i, \boldsymbol{b}_i), & \text{Forward Kinematics} \\ \boldsymbol{l} = \boldsymbol{IK}^{GK}(\boldsymbol{r}, \boldsymbol{R}, \boldsymbol{a}_i, \boldsymbol{b}_i), & \text{Inverse Kinematics} \end{cases} \tag{4}$$

where $\boldsymbol{FK}^{GK}(\bullet)$ and $\boldsymbol{IK}^{GK}(\bullet)$ are implicit functions of the forward kinematics and the inverse kinematics for the generalized kinematics model, respectively.

3.2. Kinematics Modeling Considering the Adaptive Pulley

Since the eight winch reel modules (four upper modules and four lower modules) have the same mechanical structure, an upper winch reel module and the corresponding lower winch reel module are taken as an example to establish the improved kinematics model of the cable-driven parallel robot.

As shown in Figure 5, the cable in the upper winch reel module can be divided into four sections through limit pulley, tension pulley and adaptive pulley. The cable length of the first section is set as $L_{i,1}$, which is from the cable fixed point C_i to the cable exit point D_i on the winch reel. $L_{i,2}$ is the cable length of the second segment, that is, from point D_i to the tangency point E_i of the tension pulley. The third cable section is from point E_i to the tangency point A_i of the adaptive pulley and its length is denoted by $L_{i,3}$. The length of the last cable section is set as $L_{i,4,}$ that is, from point A_i to the connection point B_i on the end-effector. For the lower winch reel module, the definitions of symbols are the same as in the upper module, except that the subscript i is replaced by the symbol k. It should be noted that $L_{i,2}$ and $L_{i,3}$ are constant, after the structural design of the robot is completed. The values of $L_{i,1}$ and $L_{i,4}$ will change with the movement of the end-effector. However, the primary focus of this study is the corresponding relationship between the end-effector pose and $L_{i,4}$. Even though $L_{i,1}$ is a variable, it does not affect the subsequent kinematics modeling and solution results. As a result, $L_{i,4}$ is served as the main research object.

In order to describe the kinematics model including adaptive pulleys, two coordinate systems $\{x_c, y_c, z_c\}$ and $\{x_p, y_p, z_p\}$ are built, as shown in Figure 6. Their origins (o_c and o_p) are both set at point A_i. In the initial state, the x_c-axis, y_c-axis and z_c-axis are parallel to three axes of $\{x_p, y_p, z_p\}$, respectively. The coordinate system $\{x_p, y_p, z_p\}$ rotates with the movement of the adaptive pulley.

https://doi.org/10.3390/s26010039

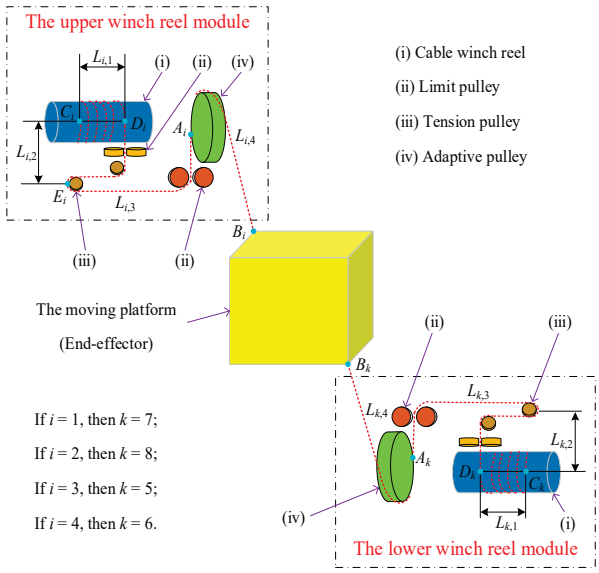


Figure 5. The upper and lower winch modules.

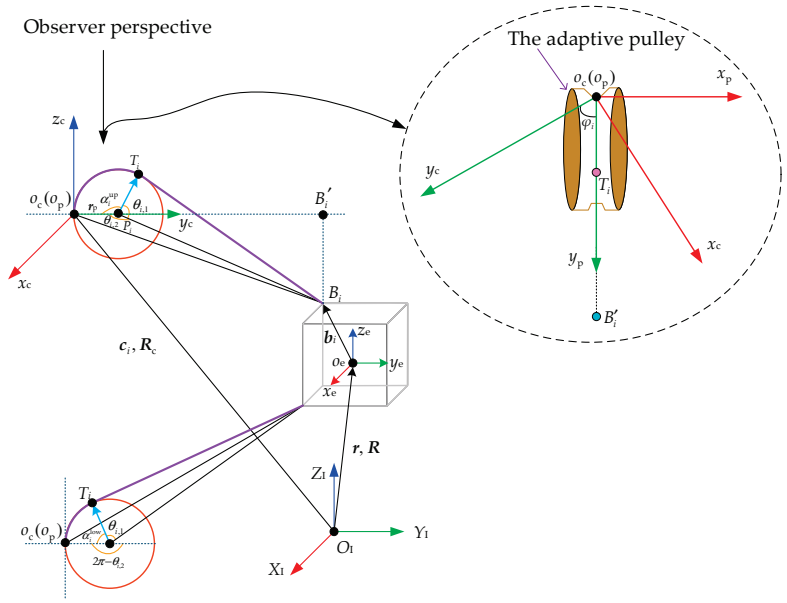


Figure 6. Kinematics model with adaptive pulley.

The rotation angle of the adaptive pulley is denoted by φ_i. T_i is the cable exit point and $\overrightarrow{T_iB_i}$ is the tangent line of the adaptive pulley. α_i is the central angle corresponding to the arc $\overset{\frown}{o_cT_i}$ on the adaptive pulley. In other words, α_i is the wrap angle of the ith cable around the adaptive pulley. The value range of α_i is divided into two categories due to the different installation locations of adaptive pulleys. One is α_i^{up} that corresponds to the adaptive pulley in upper winch modules and $\alpha_i^{\text{up}} \in (90°,\ 180°)$. The other one is α_i^{low} which is associated with the adaptive pulley in lower winch modules and $\alpha_i^{\text{low}} \in (0°,\ 90°)$. The point P_i is the center of the ith adaptive pulley. $\theta_{i,1}$ is the angle between the vector $\overrightarrow{P_iT_i}$ and the vector $\overrightarrow{P_iB_i}$. $\theta_{i,2}$ is the angle between $\overrightarrow{P_iB_i}$ and the vector $\overrightarrow{P_io_c}$.

https://doi.org/10.3390/s26010039

When considering the motion of adaptive pulleys, the actual length of the ith cable is made up of the lengths of arc $\overset{\frown}{o_c T_i}$ and line $\overrightarrow{T_i B_i}$. If the lengths of $\overset{\frown}{o_c T_i}$ and $\overrightarrow{T_i B_i}$ are denoted by L_i^{arc} and L_i^{line}, they can be calculated as follows:

$$\begin{cases} L_i^{\text{arc}} = \alpha_i \left\| \boldsymbol{r}_{\text{p}} \right\| \\ L_i^{\text{line}} = \left\| (\boldsymbol{r} + \boldsymbol{R}\boldsymbol{b}_i) - \left(\boldsymbol{c}_i + \boldsymbol{R}_{\text{c}}\boldsymbol{R}_z(\varphi_i)\boldsymbol{r}_{\text{p}} - \boldsymbol{R}_{\text{c}}\boldsymbol{R}_z(\varphi_i)\boldsymbol{R}_x(-\alpha_i)\boldsymbol{r}_{\text{p}} \right) \right\| \end{cases} \tag{5}$$

where $\boldsymbol{r}_{\text{p}}$ is the vector form o_{c} to P_i at the initial state. Thus, $\boldsymbol{r}_{\text{p}} = [0, d, 0]^{\text{T}}$, the radius vector of the adaptive pulley. Its magnitude is denoted by d. $\boldsymbol{c}_i$ is the position vector in Σ_{I} and it is a known quantity. $\boldsymbol{R}_{\text{c}}$ is the rotation matrix from Σ_{c} to Σ_{I} and $\boldsymbol{R}_z(\varphi_i)$ and $\boldsymbol{R}_x(\alpha_i)$ are rotation matrixes about the z-axis and x-axis, and they are obtained as follows:

$$\boldsymbol{R}_z(\varphi_i) = \begin{bmatrix} \cos\varphi_i & -\sin\varphi_i & 0 \\ \sin\varphi_i & \cos\varphi_i & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ \boldsymbol{R}_x(-\alpha_i) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_i & \sin\alpha_i \\ 0 & -\sin\alpha_i & \cos\alpha_i \end{bmatrix} \tag{6}$$

Unlike the generalized kinematics model that can directly compute the cable lengths if the pose of the end-effector is known, the improved kinematics model requires calculating the angle φ_i and the angle α_i to determine cable lengths. The improved kinematic model refers to the inverse kinematics considering adaptive pulleys movement. The detailed calculation process of the above two angles is given as follows.

When the pose of end-effector and the attitude transformation relationship between Σ_{c} and Σ_{I} is known, the position vector of B_i in the Σ_{c} is determined by:

$${}^{\text{c}}\overrightarrow{o_{\text{c}}B_i} = \boldsymbol{R}_{\text{c}}{}^{\text{T}} \cdot (\boldsymbol{r} + \boldsymbol{R}\boldsymbol{b}_i - \boldsymbol{c}_i) \tag{7}$$

Because the frame $\{x_{\text{p}}, y_{\text{p}}, z_{\text{p}}\}$ moves along with the adaptive pulley and the cable $\overrightarrow{T_i B_i}$ is tangent to the adaptive pulley, $\overrightarrow{T_i B_i}$ is always on the plane $y_{\text{p}} o_{\text{p}} z_{\text{p}}$. As a result, the projection of the point B_i on the plane $x_{\text{p}} o_{\text{p}} y_{\text{p}}$ must be on the y_{p}-axis and it is denoted by $B_i{}'$ (see Figure 6). The position vector of $B_i{}'$ in the Σ_{c} is calculated as follows:

$${}^{\text{c}}\overrightarrow{o_{\text{c}}B_i{}'} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot {}^{\text{c}}\overrightarrow{o_{\text{c}}B_i} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \left(\boldsymbol{R}_{\text{c}}{}^{\text{T}} \cdot (\boldsymbol{r} + \boldsymbol{R}\boldsymbol{b}_i - \boldsymbol{c}_i) \right) \tag{8}$$

According to the definition of dot product, the angle φ_i can be obtained as follows:

$$\varphi_i = \arccos \left(\frac{ {}^{\text{c}}\overrightarrow{o_{\text{c}}B_i{}'} \cdot \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^{\text{T}} }{ \left\| {}^{\text{c}}\overrightarrow{o_{\text{c}}B_i{}'} \right\| } \right) \tag{9}$$

Furthermore, it is the process of calculating the angle α_i. The vector $\overrightarrow{P_i B_i}$ is obtained as follows:

$$\overrightarrow{P_i B_i} = \boldsymbol{r} + \boldsymbol{R}\boldsymbol{b}_i - \boldsymbol{c}_i + \boldsymbol{R}_{\text{c}}\boldsymbol{R}_z(\varphi_i)\boldsymbol{r}_{\text{p}} \tag{10}$$

Based on the previous discussion, $\overrightarrow{T_i B_i}$ is the tangent line of the adaptive pulley. Therefore, the vector $\overrightarrow{P_i T_i}$ is perpendicular to $\overrightarrow{T_i B_i}$ and $\Delta P_i T_i B_i$ is a right triangle. $\theta_{i,1}$ can be obtained by:

https://doi.org/10.3390/s26010039

$$\theta_{i,1} = \arccos\left(\frac{\left\|\overrightarrow{P_iT_i}\right\|}{\left\|\overrightarrow{T_iB_i}\right\|}\right) = \arccos\left(\frac{d}{\left\|\overrightarrow{T_iB_i}\right\|}\right) \tag{11}$$

According to the cosine theorem of a triangle, $\theta_{i,2}$ is calculated as follows:

$$\theta_{i,2} = \arccos\left(\frac{d^2 + \left\|\overrightarrow{P_iB_i}\right\|^2 - \left\|{}^{c}\overrightarrow{o_cB_i}\right\|^2}{2d\left\|\overrightarrow{P_iB_i}\right\|}\right) \tag{12}$$

On basis of the relationship among angles α_i, $\theta_{i,1}$ and $\theta_{i,2}$, α_i can be calculated. However, it should be noted that the angle α_i has two different results, i.e., α_i^{up} and α_i^{low}, which are obtained as follows:

$$\alpha_i = \begin{cases} \alpha_i^{\text{up}} = 2\pi - \theta_{i,1} - \theta_{i,2}, \text{ Corresponding to upper adaptive pulley} \\ \alpha_i^{\text{low}} = \theta_{i,2} - \theta_{i,1}, \quad \text{Corresponding to lower adaptive pulley} \end{cases} \tag{13}$$

Combining Equations (5), (9) and (13), the length of the ith cable is determined as follows:

$$l_i = L_i^{\text{arc}} + L_i^{\text{line}} = d\alpha_i + \left\|(\boldsymbol{r} + \boldsymbol{R}\boldsymbol{b}_i) - \left(\boldsymbol{c}_i + \boldsymbol{R}_c\boldsymbol{R}_z(\varphi_i)\begin{bmatrix}0\\d\\0\end{bmatrix} - \boldsymbol{R}_c\boldsymbol{R}_z(\varphi_i)\boldsymbol{R}_x(-\alpha_i)\begin{bmatrix}0\\d\\0\end{bmatrix}\right)\right\| \tag{14}$$

Here, φ_i and α_i can be obtained based on Equations (9) and (13). Similarly, the pose of the end-effector will be obtained when each cable's length is known. The following two assumptions are given:

$$\boldsymbol{\varphi} = [\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7, \varphi_8]^{\text{T}} \tag{15}$$

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8]^{\text{T}} \tag{16}$$

The models of forward kinematics and inverse kinematics, which both consider adaptive pulleys, can be established as follows:

$$\begin{cases} [\boldsymbol{\varphi}, \boldsymbol{\alpha}, \boldsymbol{r}, \boldsymbol{R}] = \boldsymbol{FK}^{MK}(\boldsymbol{l}, \boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{R}_c), & \text{Forward Kinematics} \\ [\boldsymbol{\varphi}, \boldsymbol{\alpha}, \boldsymbol{l}] = \boldsymbol{IK}^{MK}(\boldsymbol{r}, \boldsymbol{R}, \boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{R}_c), & \text{Inverse Kinematics} \end{cases} \tag{17}$$

where $\boldsymbol{FK}^{MK}(\bullet)$ and $\boldsymbol{IK}^{MK}(\bullet)$ are the forward kinematics and inverse kinematics equations for the improved modeling method, respectively.

4. Kinematics Solution Based on a Hybrid Levenberg–Marquardt and Genetic Algorithm

4.1. The Methodology of the Fusion Algorithm

The cable-driven parallel robot has 6-DOFs, redundantly actuated by eight cables. Due to the existence of redundant constraints, it poses significant challenge for the kinematic solution of the robot, especially solving the forward kinematics.

Although scholars have conducted some research on this issue, some numerical algorithms, such as the Newton–Raphson method and neural network method, have their own limitations in both respects. One is the ability to achieve a kinematic solution on the workspace boundary. The other is the solving accuracy of training samples. In

https://doi.org/10.3390/s26010039

order to address these issues, a hybrid algorithm based on a hybrid Levenberg–Marquardt and Genetic method is proposed to achieve the efficient and high-precision solution of kinematics equations by combining rapid global search and precise local optimization. In the algorithm, the Genetic method is adopted to converge to near the optimal solution rapidly. Based on this, the above results can provide theoretical support for trajectory planning and motion control strategies for cable-driven parallel robots.

According to Equation (17), the specific relationship between the length of each cable, the pose of the end-effector and the angles of adaptive pulleys can be obtained. Therefore, the objective equation of the forward kinematics solution can be established as follows:

$$\begin{cases} \min\|\boldsymbol{G}(\boldsymbol{\varphi},\boldsymbol{\alpha},\boldsymbol{r},\boldsymbol{R})\|,\ \boldsymbol{G}(\varphi,\boldsymbol{\alpha},\boldsymbol{r},\boldsymbol{R}) = \boldsymbol{f}(\boldsymbol{\varphi},\boldsymbol{\alpha},\boldsymbol{r},\boldsymbol{R}) - \boldsymbol{l}_{\mathrm{d}} \\ s.t. \begin{cases} \left(\left\|\boldsymbol{c}_i + \boldsymbol{R}_{\mathrm{c}}\boldsymbol{R}_z(\varphi_i)\boldsymbol{r}_{\mathrm{p}} - (\boldsymbol{r} + \boldsymbol{R}\boldsymbol{b}_i)\right\|^2 - \left\|\boldsymbol{r}_{\mathrm{p}}\right\|^2\right) - \left(\boldsymbol{l}_{\mathrm{d}}(i) - \boldsymbol{\alpha}(i)\left\|\boldsymbol{r}_{\mathrm{p}}\right\|\right)^2 = 0 \\ -\frac{\pi}{2} \le \boldsymbol{\varphi}(i) \le \frac{\pi}{2} \\ \frac{\pi}{2} \le \boldsymbol{\alpha}(i) \le \pi,\ i = 1,2,3,4 \\ 0 \le \boldsymbol{\alpha}(i) \le \frac{\pi}{2},\ i = 5,6,7,8 \end{cases} \end{cases} \tag{18}$$

where $\boldsymbol{l}_{\mathrm{d}}$ is the desired length of all cables and it is a known quantity. $\boldsymbol{f}(\boldsymbol{\varphi},\boldsymbol{\alpha},\boldsymbol{r},\boldsymbol{R})$ represents the function for calculating the length of cables, and it can be obtained using Equation (14). In the same way, the function of the inverse kinematics solution for the robot can also be built by utilizing the approach in Equation (18). Based on the hybrid Levenberg–Marquardt and Genetic algorithm (HLMGA), the process of the forward kinematics solution is listed in Algorithm 1.

Algorithm 1. The process for solving forward kinematics based on HLMGA

Require: Desired cable lengths $\boldsymbol{l}_{\mathrm{d}}$
Initialize the pose $(\boldsymbol{r}_0, \boldsymbol{R}_0)$ and angles $(\boldsymbol{\varphi}_0, \boldsymbol{\alpha}_0)$
for $t = 0$ **to** T **do**
  Genetic algorithm main loop
    Set the size of population N_{pop} and threshold δ_1
    Calculate each individual *fitness*, based on Equation (20)
    Sort population by *fitness* and select top 50% individuals
    Perform single-point crossover with probability P_{rc}
    Apply small perturbations with probability P_{rm}
    Combine selected individuals and mutated individuals
  Levenberg–Marquardt local optimization
    Set scaling coefficient μ, amplification coefficient v and threshold δ_2
    for $k = 1$ **to** $N_{L\text{-}M}$ **do**
    Computing the Jacobian matrix $\boldsymbol{J}_k$and the function value $\boldsymbol{G}_k$
    Update step $\boldsymbol{d}_k$ using Equation (22)
    Recalculate $\hat{\boldsymbol{J}}_k$ and $\hat{\boldsymbol{G}}_k$
    Evaluate step size effectiveness λ, based on Equation (24)
    if λ > 0, reduce the value and adjust the amplification factor
    else if increase μ and increase v
    end if
    if $\|\boldsymbol{d}_k\| < \delta_2$, then, **break**
    else if, reset X_{e}, Y_{e}, Z_{e}, ψ_1, ψ_2, ψ_2, φ_i, α_i, and go to step 13
    end if
  end for
end for
Output the optimal result

https://doi.org/10.3390/s26010039

To facilitate further explanation of the proposed method, the detailed process of the forward kinematics solution is given as follows.

4.2. Parameter Initialization

(a) Set the size of the fixed frame as that described in Section 2. Set the simulation time (T) and the simulation time step (dT). Initialize the pose of end-effector and the angles of adaptive pulleys as follows:

$$\begin{cases} \boldsymbol{r}_0 = \begin{bmatrix} X_{\mathrm{e}}(0) & Y_{\mathrm{e}}(0) & Z_{\mathrm{e}}(0) \end{bmatrix}^{\mathrm{T}},\ \boldsymbol{R}_0 = \boldsymbol{R}_x(\psi_1(0))\boldsymbol{R}_y(\psi_2(0))\boldsymbol{R}_z(\psi_3(0)) \\ \boldsymbol{\varphi}_0 = \begin{bmatrix} \varphi_1(0) & \varphi_2(0) & \varphi_3(0) & \varphi_4(0) & \varphi_5(0) & \varphi_6(0) & \varphi_7(0) & \varphi_8(0) \end{bmatrix}^{\mathrm{T}} \\ \boldsymbol{\alpha}_0 = \begin{bmatrix} \alpha_1(0) & \alpha_2(0) & \alpha_3(0) & \alpha_4(0) & \alpha_5(0) & \alpha_6(0) & \alpha_7(0) & \alpha_8(0) \end{bmatrix}^{\mathrm{T}} \end{cases} \tag{19}$$

where $X_{\mathrm{e}}(0)$, $Y_{\mathrm{e}}(0)$ and $Z_{\mathrm{e}}(0)$ are the position of the end-effector at the T = 0. Its initial attitude angles (X–Y–Z Euler angles) are $X_{\mathrm{e}}(0)$, $Y_{\mathrm{e}}(0)$ and $Z_{\mathrm{e}}(0)$, respectively. $\alpha(i)$ is the initial angle of the adaptive pulley.

(b) Set the parameters of the Genetic algorithm. The population size is set as N_{pop} = 300. The maximum number of iterations $Maxitr_{\mathrm{heredity}}$, the crossover probability P_{rc} and the mutation probability P_{rm} are 50, 0.8, and 0.05, respectively.

(c) Initialize the parameters of the Levenberg–Marquardt algorithm. The initial scaling coefficient μ and amplification coefficient v are set as μ = 0.01 and v = 5. The amplification factor v is used to increase the value of μ when the step size of the Levenberg–Marquardt method is not appropriate. The maximum number of iterations is $Maxitr_{\mathrm{L\text{-}M}}$ = 20. The convergence termination coefficients for the Genetic method and Levenberg–Marquardt are δ_1 and δ_2, respectively.

4.3. Genetic Algorithm Main Loop

(a) Coding and population setup

Through the decimal encoding strategy, the population is randomly generated in the solution space for each time step. Each individual in the population represents a possible variable that includes the pose of the end-effector and the angles of adaptive pulleys. The initial population is generated by adding random noise to ensure diversity and even distribution.

(b) Fitness function calculation

For each individual in the population, its fitness value calculation is the key to solving the forward kinematics rapidly. First, the objective function value is calculated according to the value of each individual, and then the cable length error is obtained. If the above error exceeds the set threshold, a larger penalty term will be introduced and it is denoted by *penalty*, which can let the individual still retain a certain choice possibility but place it at a disadvantage in the choice. Otherwise, the optimized value is obtained. Here, the nonlinear equations representing the solution of the forward kinematics of the cable-driven parallel robot can be obtained using Equation (18). In other words, the pose and angles satisfying $\|\boldsymbol{G}(\boldsymbol{\varphi}, \boldsymbol{\alpha}, \boldsymbol{r}, \boldsymbol{R})\| = 0$ are the optimal solution of the forward kinematics. However, in the actual solution process, it is difficult to obtain a solution that completely satisfies the above equation, so it can only be sorted according to the sum of fitness values, and the smallest individual is selected to inherit to the next generation. The function values of each connection point are added to obtain the fitness value of the individual, i.e.:

$$fitness = \sum_{i=1}^{8} (\boldsymbol{G}(i))^2 \tag{20}$$

https://doi.org/10.3390/s26010039

where $\boldsymbol{G}(i)$ is the ith element of function value $\boldsymbol{G}$.

(c) Select an operation

The absolute fitness value of each individual is sorted, and the half of the individuals with the least fitness are selected as the parents of the next generation.

(d) Cross operation

For each pair of adjacent parent individuals from 1 to $(N_{\mathrm{pop}}/2) - 1$, the crossing operation is performed with the probability P_{rc}.

(e) Mutation operation

For each parent individual from $N_{\mathrm{pop}}/2$ to N_{pop}, the mutation operation is performed with the probability P_{rm}. The variation point is selected randomly, and minimal random disturbance is added to the pose and angles parameters of the point to ensure the diversity of the population and to avoid falling into the local optimal.

(f) Population renewal

For the new individuals, their crossover and mutation are merged with the parent population to form a new population and to continue the next iteration.

4.4. Local Optimization by L-M Algorithm

(a) Local optimization after Genetic algorithm

After each iteration of the Genetic algorithm, the optimal individual with the best fitness value is used as the initial value of the Levenberg-Marquardt algorithm.

(b) Jacobian matrix and function values calculation

The Jacobian matrix and the function value can be calculated based on Equation (18). The Jacobian matrix $\boldsymbol{J}_k$ reflects the sensitivity of cable lengths with respect to the pose of end-effector and angles of adaptive pulleys. The central difference method can also be used for numerical calculation. Hence, it is derived as follows:

$$\boldsymbol{J}_k = \begin{bmatrix} \frac{\partial \boldsymbol{G}(1)}{\partial X_{\mathrm{e}}}, \frac{\partial \boldsymbol{G}(1)}{\partial Y_{\mathrm{e}}}, \frac{\partial \boldsymbol{G}(1)}{\partial Z_{\mathrm{e}}}, \frac{\partial \boldsymbol{G}(1)}{\partial \psi_1}, \frac{\partial \boldsymbol{G}(1)}{\partial \psi_2}, \frac{\partial \boldsymbol{G}(1)}{\partial \psi_3}, \frac{\partial \boldsymbol{G}(1)}{\partial \varphi_1}, \cdots, \frac{\partial \boldsymbol{G}(1)}{\partial \varphi_8}, \frac{\partial \boldsymbol{G}(1)}{\partial \alpha_1} \cdots \frac{\partial \boldsymbol{G}(1)}{\partial \alpha_8} \\ \frac{\partial \boldsymbol{G}(2)}{\partial X_{\mathrm{e}}}, \frac{\partial \boldsymbol{G}(2)}{\partial Y_{\mathrm{e}}}, \frac{\partial \boldsymbol{G}(2)}{\partial Z_{\mathrm{e}}}, \frac{\partial \boldsymbol{G}(2)}{\partial \psi_1}, \frac{\partial \boldsymbol{G}(2)}{\partial \psi_2}, \frac{\partial \boldsymbol{G}(2)}{\partial \psi_3}, \frac{\partial \boldsymbol{G}(2)}{\partial \varphi_1}, \cdots, \frac{\partial \boldsymbol{G}(2)}{\partial \varphi_8}, \frac{\partial \boldsymbol{G}(2)}{\partial \alpha_1} \cdots \frac{\partial \boldsymbol{G}(2)}{\partial \alpha_8} \\ \vdots \\ \frac{\partial \boldsymbol{G}(8)}{\partial X_{\mathrm{e}}}, \frac{\partial \boldsymbol{G}(8)}{\partial Y_{\mathrm{e}}}, \frac{\partial \boldsymbol{G}(8)}{\partial Z_{\mathrm{e}}}, \frac{\partial \boldsymbol{G}(8)}{\partial \psi_1}, \frac{\partial \boldsymbol{G}(8)}{\partial \psi_2}, \frac{\partial \boldsymbol{G}(8)}{\partial \psi_3}, \frac{\partial \boldsymbol{G}(8)}{\partial \varphi_1}, \cdots, \frac{\partial \boldsymbol{G}(8)}{\partial \varphi_8}, \frac{\partial \boldsymbol{G}(8)}{\partial \alpha_1} \cdots \frac{\partial \boldsymbol{G}(8)}{\partial \alpha_8} \end{bmatrix} \tag{21}$$

In addition, the function value $\boldsymbol{G}$ can be obtained using Equation (18) directly. $\boldsymbol{G}(i)$ is the ith element of $\boldsymbol{G}$.

(c) The precise solution calculation

In the iteration process of the *L–M* method, the search step size is calculated first, and the parameters including the pose and angles are updated. Then, the new attitude matrix, Jacobian matrix and function values are obtained. Finally, we determine whether the step size meets the termination condition, and if so, stop the iteration. Otherwise, the scale coefficient is adjusted according to the effect of the step size update. The Jacobian matrix and the function value are updated to continue the next iteration. Here, the step size is calculated as follows:

$$\boldsymbol{d}_k = -\left(\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{J}_k + \mu\boldsymbol{I}\right)^{-1}\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{G}_k \tag{22}$$

where $\boldsymbol{I}$ is the 12th-order identity matrix. The function value $\boldsymbol{G}_k$ represents the deviation between the current cable length and the desired cable length in the kth step calculation.

If the step size fails to meet the termination condition, the values of pose and angles should be updated by the following equation:

https://doi.org/10.3390/s26010039

$$\begin{cases} X_e(k+1) = X_e(k) + \boldsymbol{d}_k(1), Y_e(k+1) = Y_e(k) + d_k(2), Z_P(k+1) = Z_P(k+1) + d_k(3) \\ \psi_1(k+1) = \psi_1(k) + \boldsymbol{d}_k(4), \psi_2(k+1) = \psi_2(k) + \boldsymbol{d}_k(5), \psi_3(k+1) = \psi_3(k) + \boldsymbol{d}_k(6) \\ \varphi_i(k+1) = \varphi_i(k) + \boldsymbol{d}_k(i+6); \alpha_i(k+1) = \alpha_i(k) + \boldsymbol{d}_k(i+14), i = 1, 2, \ldots, 8 \end{cases} \tag{23}$$

Substituting Equation (23) into Equations (18) and (21), the updated Jacobian matrix and the function value can be recalculated.

If the step size $\|\boldsymbol{d}_k\| < \delta_2$, the algorithm converges and the iteration is terminated. If not, we calculate the value of the adaptive damped shadow λ as follows:

$$\lambda = \frac{\boldsymbol{G}_k^{\mathrm{T}} \boldsymbol{G}_k - \hat{\boldsymbol{G}}_k^{\mathrm{T}} \hat{\boldsymbol{G}}_k}{\mu \boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{d}_k - \boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{J}_k^{\mathrm{T}} \boldsymbol{G}_k} \tag{24}$$

where $\hat{\boldsymbol{G}}_k$ is the deviation of cable length in the next moment.

Adjust the proportional coefficient μ according to the value λ obtained by Equation (24). If $\lambda > 0$, reduce the value μ and adjust the amplification factor v. Otherwise, increase the μ value and increase the value v until $\|\boldsymbol{d}_k\| < \delta_2$ is satisfied. Finally, the optimal result is obtained by utilizing the hybrid Levenberg–Marquardt and Genetic algorithm.

5. Simulation Study

To validate the proposed method, the simulation model of the cable-driven parallel robot is established by Matlab2020 software. The straight and elliptical paths are used as the desired end-effector trajectories for verifying the forward and inverse kinematic solution. In the simulation, the parameters of the robot are listed in Table 1.

Table 1. The parameters of the cable-driven parallel robot.

Item	Value	Unit
The size of the frame	$1000 \times 1000 \times 1000$	mm
The size of the end-effector	$100 \times 100 \times 100$	mm
The radius of the adaptive pulley	0.015	m
The number of driven cables	8	/
Degrees of freedom of the end-effector	6	DOF

5.1. The Simulation Results for Inverse Kinematics

5.1.1. The Straight Path Simulation

First, the space straight trajectory is taken as the desired path of the end-effector for the simulation, as shown in Figure 7. The total simulation time is T = 12 s. The start point and the end point of the straight trajectory are set as follows:

$$\begin{cases} \boldsymbol{P}_{\mathrm{ls}} = \begin{bmatrix} -0.35 & 0.35 & 0.2 \end{bmatrix}^{\mathrm{T}} \mathrm{m} \\ \boldsymbol{P}_{\mathrm{lf}} = \begin{bmatrix} 0.35 & -0.35 & 0.8 \end{bmatrix}^{\mathrm{T}} \mathrm{m} \end{cases} \tag{25}$$

Meanwhile, we assume that the end-effector moves at constant speed. Therefore, the real-time position of the end-effector is obtained as follows:

$$\boldsymbol{P}_1(t) = \boldsymbol{P}_{\mathrm{ls}} + \frac{t}{T}(\boldsymbol{P}_{\mathrm{lf}} - \boldsymbol{P}_{\mathrm{ls}}),\ t \in [0, 12], T = 12 \tag{26}$$

https://doi.org/10.3390/s26010039

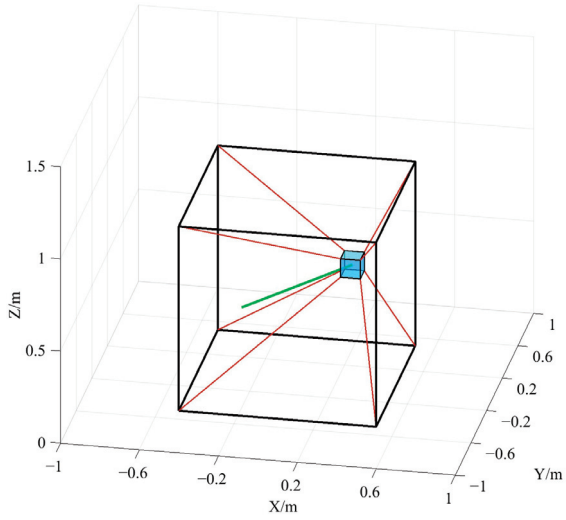


Figure 7. Spatial linear trajectory.

The desired attitude angles (X–Y–Z Euler angles) of the end-effector are set as [0°, 0°, 0°], which means the end-effector maintains its orientation aligned with the inertial frame.

Based on Equations (4) and (17), the lengths of eight cables utilizing the generalized kinematics model or the improved kinematics model that accounts for the adaptive pulley are shown in Figure 8. It can be seen that the variations in cable lengths from both methods exhibit a largely consistent trend, with certain deviations in the specific values.

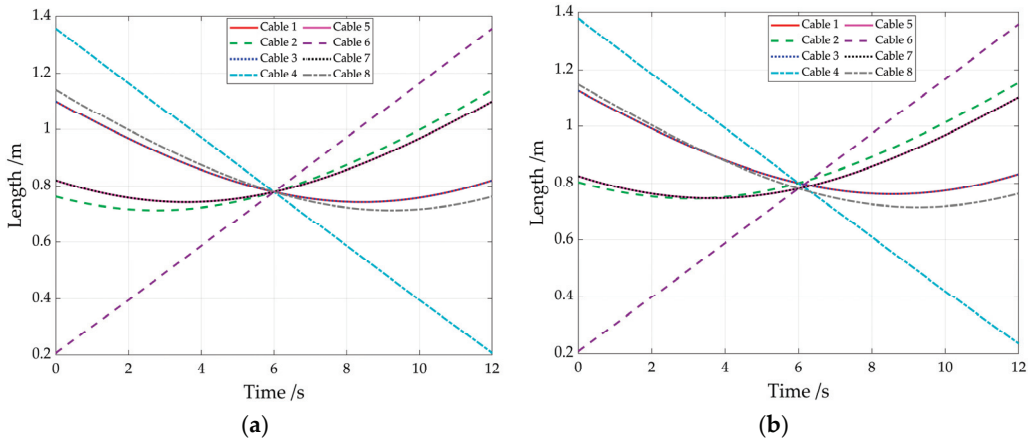


Figure 8. The variations in cable lengths. (**a**) For the generalized kinematics model; (**b**) for the improved kinematics model.

The position deviations of the cable release points for both methods are shown in Figure 9. Figure 9a is the position deviation heatmap along the X_I axis of the inertial system. Cable 1 and Cable 2 exhibit positive deviations, and Cable 3 and Cable 4 display negative deviations. The position deviation heatmaps along the Y_I axis and Z_I axis are Figures 9b and 9c, respectively. All cables are towards positive deviations, as shown in Figure 9c.

Additionally, it can be concluded that the actual cable departure point on the pulley is affected continuously during the movement of adaptive pulleys, and the deviation is roughly 0.02 m. The comparison of cable release point position deviations presented in Figure 9 aims to demonstrate that the traditional method, by neglecting the geometric variations introduced by the motion of the adaptive pulley, leads to a significant deviation (approximately 0.02 m) between the calculated cable release point positions and the actual cable release point positions. This deviation is not random noise but a systematic error

https://doi.org/10.3390/s26010039

introduced by model simplification. If the traditional method is used for the actual control directly, a non-negligible error between the theoretical trajectory and the actual trajectory of the end-effector would arise, severely limiting the robot's performance in high-precision tasks such as precise trajectory tracking.

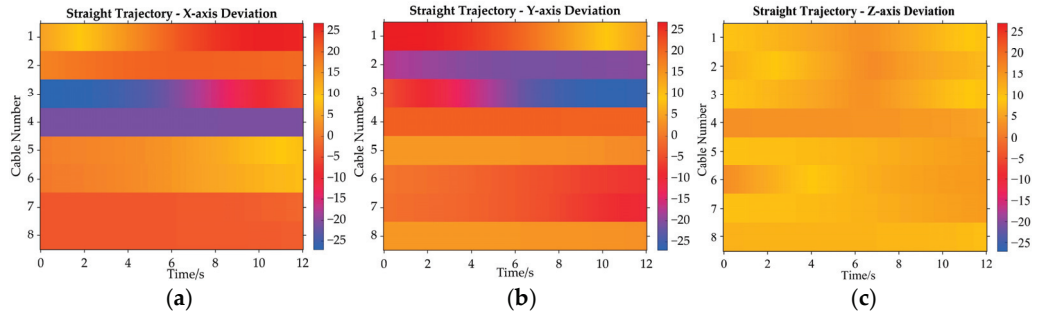


Figure 9. The position deviations of the cable release points for both methods. (**a**) Along the X_I axis; (**b**) along the Y_I axis; (**c**) along the Z_I axis.

The 3D surface diagram for the rotation angles and wrap angles of cables is shown in Figure 10. From Figure 10a, one can see that the variations in rotation angles for all adaptive pulleys are bounded between $-60°$ and $80°$ at each time step of the simulation. Moreover, the wrap angle of each cable around the corresponding adaptive pulley is shown in Figure 10b. It can be observed that the wrap angles of Cables 1–4 range from $90°$ to $180°$. They fall within the range of $0°$ to $90°$ for Cables 5–8. The correctness of the improved kinematics model is further validated by these results.

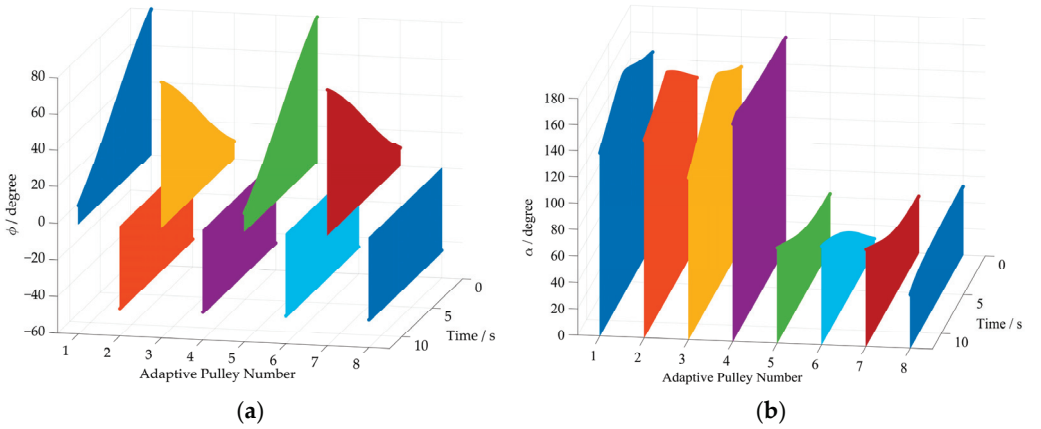


Figure 10. The rotation angles and the wrap angles of cables in straight path simulation. (**a**) The rotation angle of each adaptive pulley; (**b**) the wrap angles of cables around adaptive pulleys.

5.1.2. The Ellipse Path Simulation

Here, an elliptic trajectory is adopted to verify the inverse kinematics. The position of the ellipse's center is set as $O_{\text{ell}} = [0.1,\ 0.15,\ 0.6]$. The normal vector of the plane which the ellipse is on is $\boldsymbol{n}_{\text{ell}} = [0,\ 0,\ 1]$. The major semi-axis of the ellipse is along the X_I axis and its length is 0.3 m. The length of the minor semi-axis is 0.25 m and it is along the Y_I axis. Therefore, the desired elliptic trajectory at every moment is obtained as follows.

$$\boldsymbol{P}_{\text{ell}}(t) = \begin{bmatrix} P_{\text{ell_}x} & P_{\text{ell_}y} & P_{\text{ell_}z} \end{bmatrix}^{\text{T}} \tag{27}$$

where:

$$\begin{cases} P_{\text{ell_}x} = 0.1 + 0.3\cos(2\pi t/T) \\ P_{\text{ell_}y} = 0.15 + 0.25\sin(2\pi t/T) \quad t \in [0, 12], T = 12 \\ P_{\text{ell_}z} = 0.6 \end{cases} \tag{28}$$

https://doi.org/10.3390/s26010039

In addition, the desired attitude angles (X–Y–Z Euler angles) of the end-effector are also set as [0°, 0°, 0°]. The space ellipse trajectory is as shown in Figure 11.

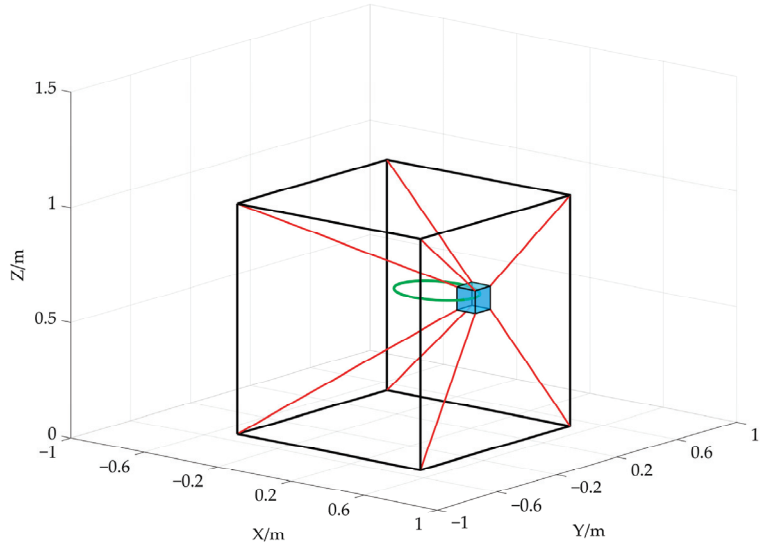


Figure 11. Spatial elliptic trajectory.

For the elliptic path simulation, the cable length variations are given in Figure 12. They exhibit clear periodic characteristics for the generalized kinematics model and the improved kinematics model. This is because the end-effector moves along the elliptical trajectory, enabling its pose to change periodically along the major and minor axes of the ellipse.

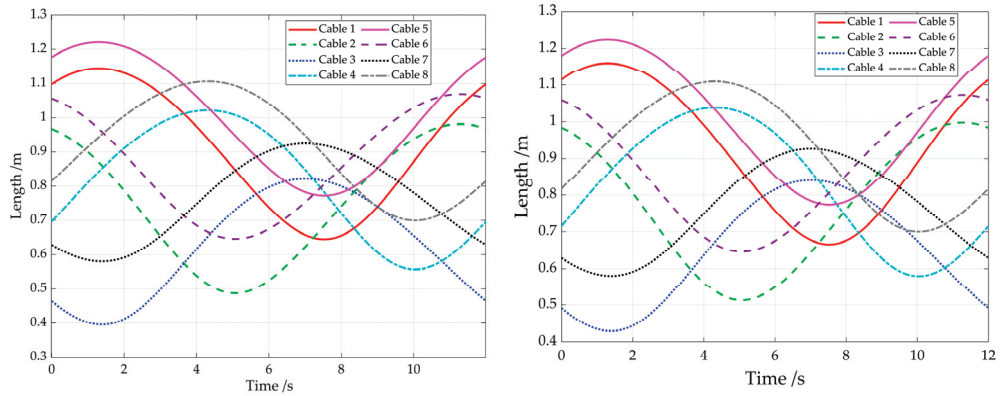


Figure 12. Cable length variations. (**a**) For the generalized kinematics model; (**b**) for the improved kinematics model.

As shown in Figure 13, through the heatmap analysis, the position deviation of each cable's departure point is visually characterized and compared. Here, Figure 13a shows the comparison of position deviations along the X_I axis for the traditional method and the improved method, considering the adaptive pulley effect. Figure 13b,c represents the position deviations along the Y_I axis and Z_I axis, respectively. Similarly to the straight path simulation, the position deviations are all approximately 0.02 m in the ellipse path simulation. As a result, these results demonstrate that, if the traditional method is employed in practical computations, it fails to calculate the cable length variations of the actual model. This finding further validates the correctness of the improved method.

https://doi.org/10.3390/s26010039

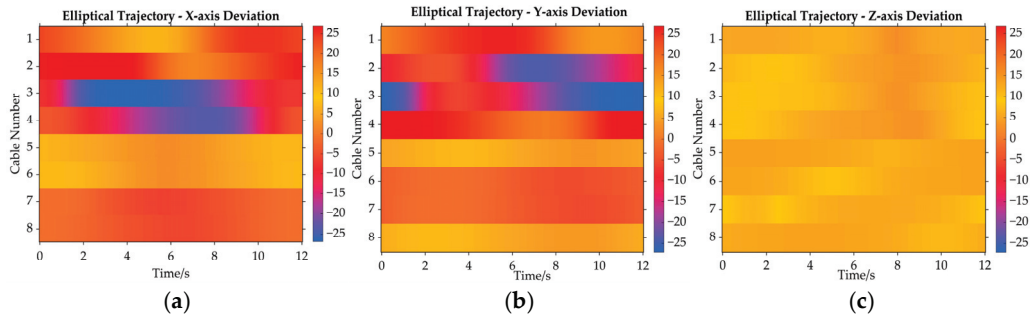


Figure 13. Cable release position deviations in the ellipse path simulation. (**a**) Along the X_I axis; (**b**) along the Y_I axis; (**c**) along the Z_I axis.

The rotation angles of all adaptive pulleys are shown in Figure 14a. Figure 14b shows the variation in the cable wrap angle around each adaptive pulley. Like the linear trajectory simulation, the values of the rotation angles and cable wrap angles of the adaptive pulleys are bounded within their defined intervals, which verifies the effectiveness of the proposed method.

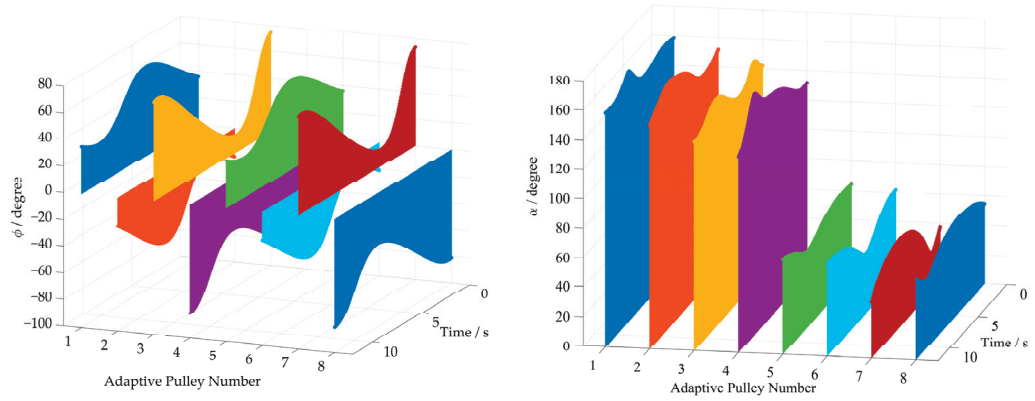


Figure 14. The rotation angles and the wrap angles of cables. (**a**) The rotation angle of each adaptive pulley; (**b**) the wrap angles of cables around the adaptive pulleys.

5.2. Simulation Results for Forward Kinematics

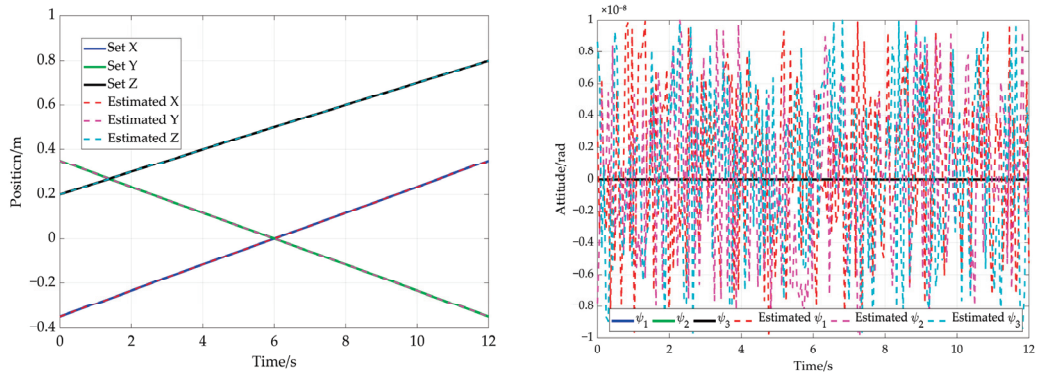
For further verification, the simulation model of the cable-driven parallel robot is also established. The model parameters are the same as those used in the inverse kinematics simulation. In the forward kinematics simulation, the cable lengths are taken as known inputs to calculate the position and attitude of the end-effector. Here, the cable length data are generated by the straight and elliptical trajectories used in Section 5.1. In this section, we first use HLMGA to solve the forward kinematics of the straight-line trajectory and the elliptical trajectory, and then we use the particle swarm optimization algorithm to solve the forward kinematics of the straight-line trajectory.

5.2.1. Forward Kinematics Simulation Based on the Straight Path via the HLMGA Algorithm

Based on the hybrid Levenberg–Marquardt and Genetic algorithm, the pose of the end-effector is obtained based on the cable length data. The comparison between the preset value of end-effector's pose and the solved value produced using the proposed method is shown in Figure 15. In this figure, the solid line represents the preset value, while the dotted line denotes the solved value. Further comparison is given in Figure 16, which shows the end-effector movement trajectories in the inertial frame. One is the preset straight

https://doi.org/10.3390/s26010039

trajectory (denoted by the solid line) obtained using the preset data. The other one is the estimated trajectory (represented by dotted line), generated via the solved data. According to Figures 15 and 16, the curve corresponding to the solved values almost coincides with that of the preset values.



(**a**) (**b**)

Figure 15. Comparison between preset values and solved values for the end-effector pose. (**a**) End-effector position comparison; (**b**) end-effector attitude comparison.

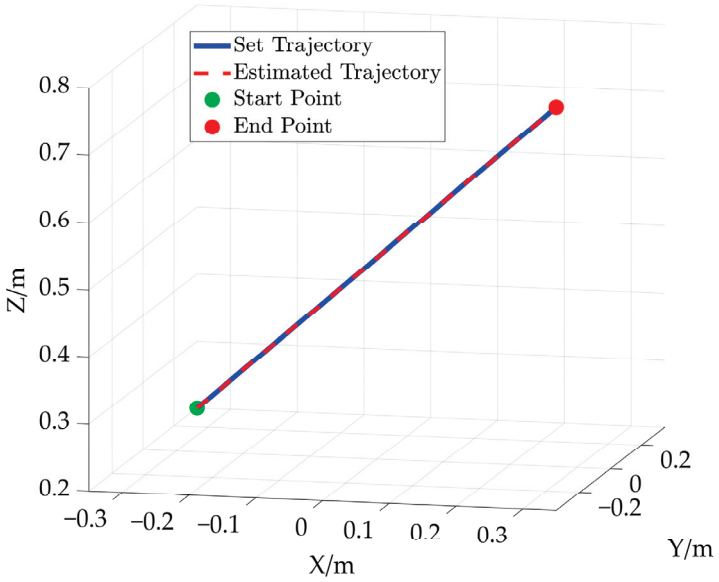


Figure 16. Comparison between the preset straight trajectory and estimated straight trajectory.

The error curves of end-effector pose are shown in Figure 17. It can be seen from Figure 17a that the maximum position error remains below 4×10^{-8} mm. The maximum attitude error is less than 1×10^{-8} radians (see Figure 17b). Moreover, the rotation angle error of adaptive pulley and the wrap angle error of the cable around adaptive pulley are given in Figure 18. The maximum errors for the two parameters do not exceed 1.5×10^{-7} milliradians (see Figure 18a) and 2×10^{-7} milliradians (see Figure 18b). These simulation results demonstrate the high solution accuracy of the proposed method, thereby validating its correctness.

The cable length errors in the simulation are shown in Figure 19; it can be seen that that the error between $\boldsymbol{l}_{\mathrm{d}}$ and $\boldsymbol{l}$ is below 6×10^{-8} mm. It demonstrates the high precision and effectiveness of the proposed solution approach.

https://doi.org/10.3390/s26010039

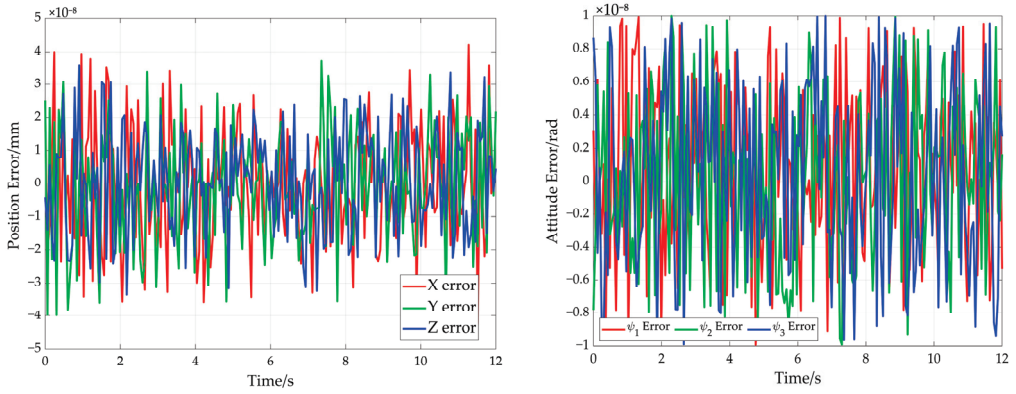


Figure 17. Pose errors of end-effector for the straight trajectory. (**a**) Position error; (**b**) attitude error.

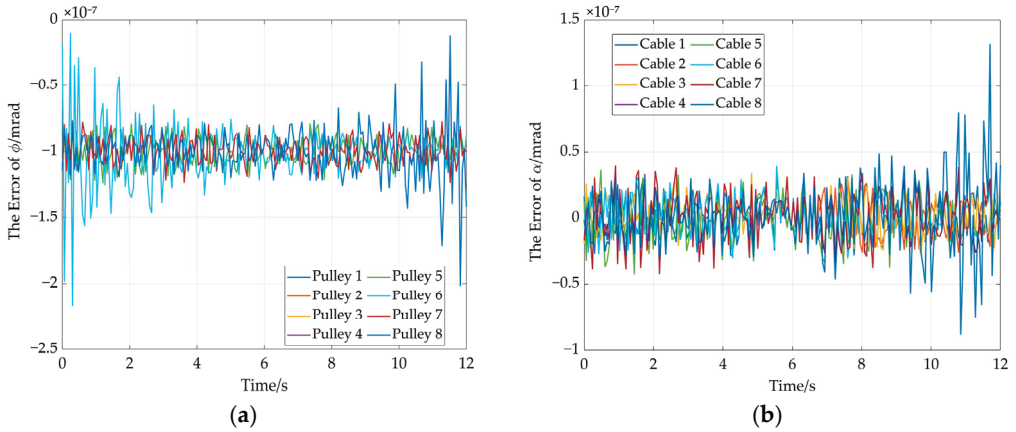


Figure 18. The rotation angle error and wrap angle error in the straight trajectory. (**a**) Rotation angle errors of adaptive pulleys; (**b**) cable wrap angle errors around adaptive pulleys.

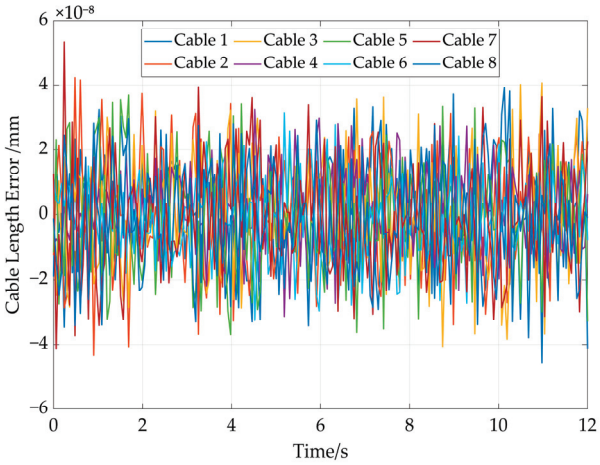


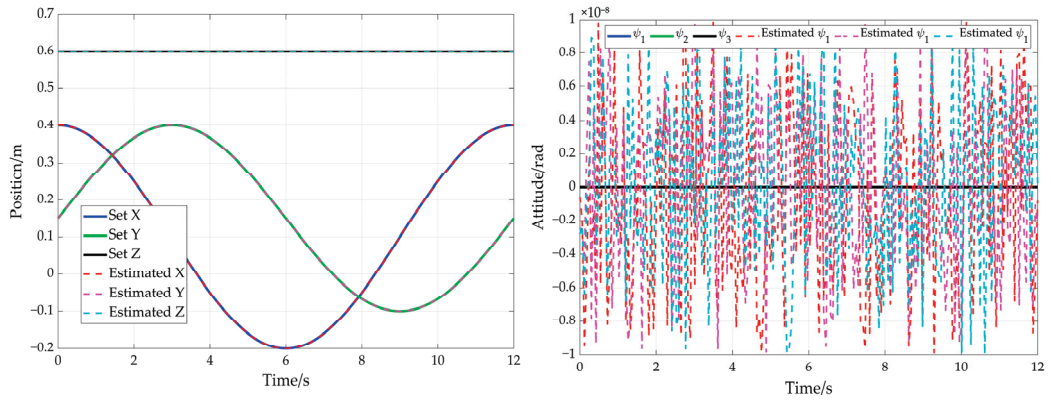
Figure 19. The cable length errors in the straight trajectory simulation.

5.2.2. Forward Kinematics Simulation Based on the Elliptical Path via the HLMGA Algorithm

In order to further verify the proposed method, an additional simulation based on an elliptical trajectory is conducted. In this simulation, the cable length data corresponding to the ellipse are utilized to compute the end-effector's pose. The comparison between the preset path of the end-effector and the solved path is given in Figure 20. Figure 20a represents the end-effector position comparison and Figure 20b shows the end-effector

https://doi.org/10.3390/s26010039

attitude comparison, where the solid line and the dotted line are preset values and solved values for the elliptical path, respectively. Furthermore, the preset trajectory and the solved trajectory are visualized in three-dimensional space, as shown in Figure 21. Based on Figures 20 and 21, the curve corresponding to the solved values almost coincides with that of the preset values.



(**a**) (**b**)

Figure 20. The comparison between preset values and solved values for the elliptical path. (**a**) End-effector position comparison; (**b**) end-effector attitude comparison.

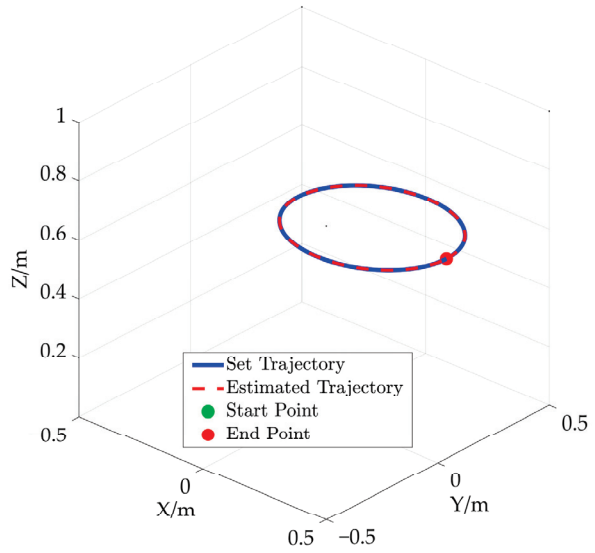


Figure 21. The comparison between the preset elliptical path and estimated elliptical path.

The pose error curves of the end-effector are given in Figure 22. It can be concluded that the maximum position error is less than 5×10^{-8} mm (see Figure 22a). The maximum attitude error is less than 1×10^{-8} radians (see Figure 22b). In addition, the rotation angle error of the adaptive pulley and wrap angle error of the cable around the adaptive pulley are given in Figure 23. The maximum error of rotation angle remains below 8×10^{-8} milliradians and it is 2×10^{-7} milliradians for the cable wrap angle error. They are both less than 2.5×10^{-7} milliradians. These simulation results further demonstrate the high solution accuracy of the proposed method, thereby validating its effectiveness.

The cable length errors in the simulation are shown in Figure 24, which shows that the error between $\boldsymbol{l}_{\mathrm{d}}$ and $\boldsymbol{l}$ is below 5×10^{-8} mm. It further demonstrates the high precision and effectiveness of the proposed solution approach.

https://doi.org/10.3390/s26010039

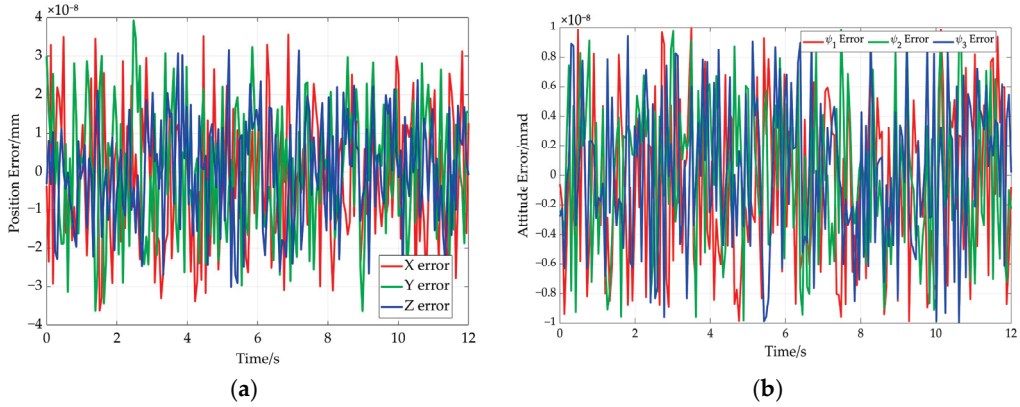


Figure 22. The pose errors of the end-effector for the elliptical path. (**a**) Position error; (**b**) attitude error.

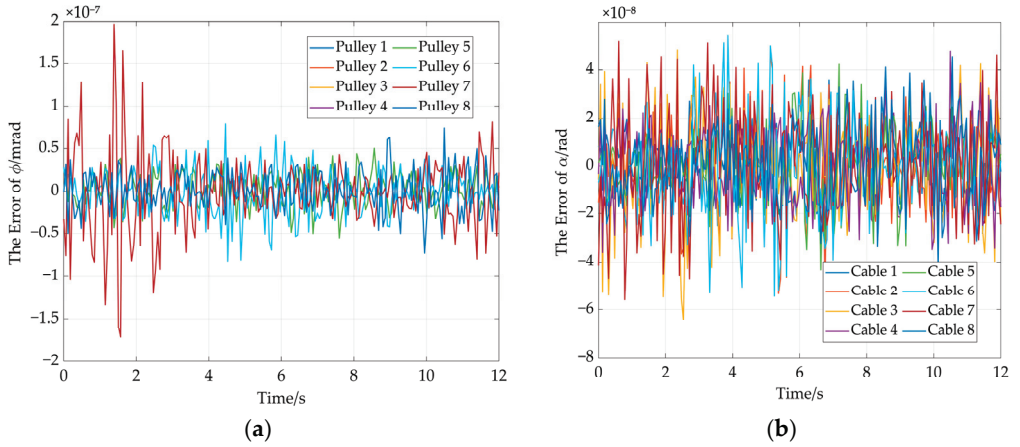


Figure 23. The rotation angle error and wrap angle error. (**a**) Rotation angle error of each adaptive pulley; (**b**) cable wrap angle error around the adaptive pulley.

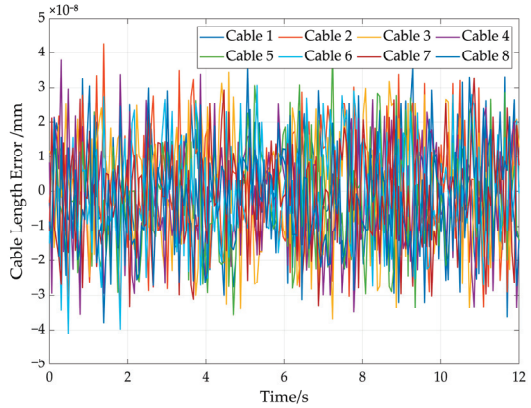
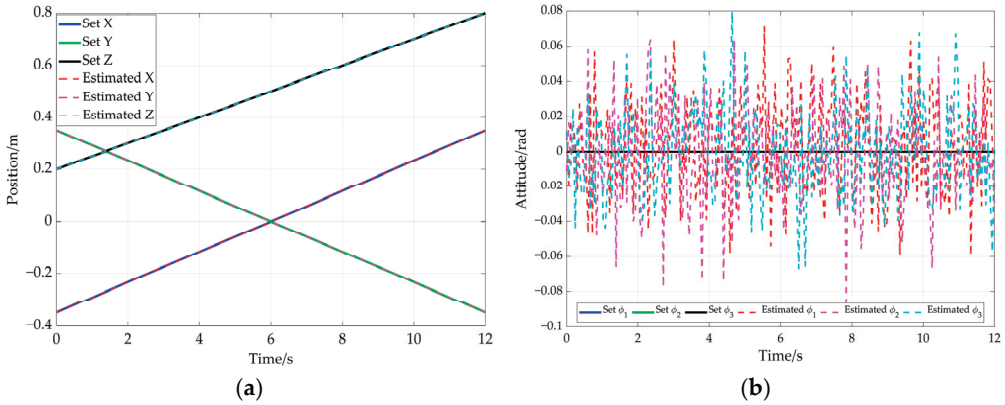


Figure 24. The cable length errors in the elliptic trajectory simulation.

5.2.3. Forward Kinematics Simulation Based on the Straight Path via the PSO Algorithm

Based on the Particle Swarm Optimization (PSO) algorithm, the comparison results between the set pose quantities and the solution values obtained using the PSO method are presented. Figure 25a shows the comparison of the end-effector position, and Figure 25b shows the comparison of the end-effector attitude. In these figures, the solid lines and dashed lines represent the preset values and the solution values of the straight path, respectively.

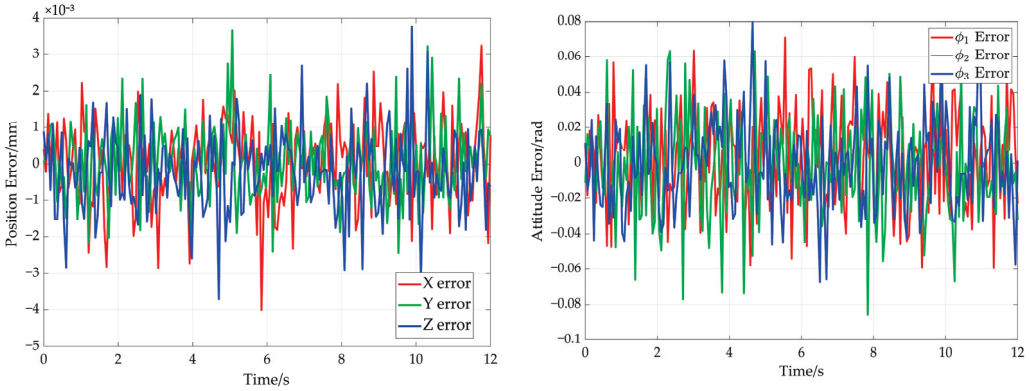
https://doi.org/10.3390/s26010039



(**a**) (**b**)

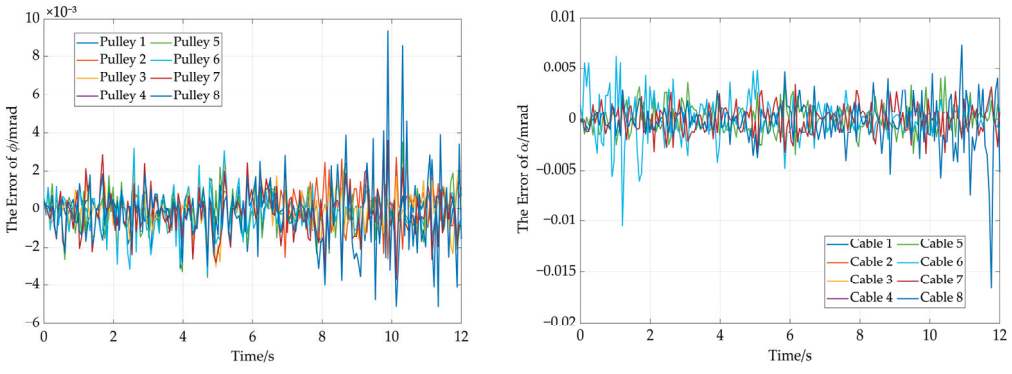
Figure 25. The comparison between preset values and solved values for the straight path. (**a**) End-effector position comparison; (**b**) end-effector attitude comparison.

The pose error curve of the end effector is shown in Figure 26. It can be concluded that the maximum position error is less than 4×10^{-3} mm (as shown in Figure 26a). The maximum attitude error is less than 6×10^{-2} radians (as shown in Figure 26b). Additionally, the rotation angle error of the adaptive pulley and the cable wrap angle error around the adaptive pulley are as shown in Figure 27. The maximum error in the angles of both is kept below 1×10^{-2} milliradians.



(**a**) (**b**)

Figure 26. The pose errors of the end-effector for the straight path. (**a**) Position error; (**b**) attitude error.



(**a**) (**b**)

Figure 27. The rotation angle error and wrap angle error. (**a**) Rotation angle error of each adaptive pulley; (**b**) cable wrap angle error around the adaptive pulley.

https://doi.org/10.3390/s26010039

5.3. Simulation Results Analysis

To evaluate the performance of the proposed method, the results of the two aforementioned algorithms were compared, and linear and elliptical trajectories were simulated. Based on the simulation results, the following conclusions can be drawn:

(a) By performing inverse kinematics simulations on the generalized kinematic model and the improved kinematic model considering the adaptive pulley effect, we concluded that the general method would result in significant errors in the calculated cable values (more than 0.02 m). This means that the above calculated values cannot be used as cable length control data in the actual operation of the CDPR. Compared with the generalized kinematic model, the improved method can effectively solve this problem.

(b) Although the forward kinematics problem is more challenging than the inverse kinematics problem, the proposed method can also effectively handle this issue. According to the simulation results, for both linear and elliptical paths, the position error of the end effector is below 5×10^{-8} m. For the attitude error of the end effector, it is still less than 1×10^{-8} radians. However, when using the Particle Swarm Optimization algorithm to solve the forward kinematics of the linear trajectory, the result is much lower in accuracy than the HLMGA method. The simulation results demonstrate the high accuracy of these solutions, thereby verifying the overall effectiveness of the proposed method.

6. Conclusions

To improve the kinematic accuracy of the cable-driven parallel robot (CDPRs), a kinematic modeling and solution method based on the hybrid Levenberg–Marquardt algorithm and genetic algorithm is proposed. The structural design is analyzed to clarify the working principle of the CDPR. A generalized kinematic model and an improved kinematic model considering adaptive pulleys are established. Then, a kinematic solution method based on the hybrid Levenberg–Marquardt algorithm and genetic algorithm is developed to achieve high-precision solutions. Finally, to evaluate the effectiveness of the proposed method, a simulation system including a forward kinematic simulation model and an inverse kinematic simulation model is constructed using Matlab2020 software. It should be noted that the effectiveness of the method is fundamentally rooted in its rigorous mathematical derivation and algorithmic design, rather than in any particular software. Here, Matlab2020 is employed merely as an efficient and general-purpose numerical computing and simulation tool to execute the algorithm, solve the equations, and visualize the results. Straight line paths and elliptical paths are introduced as the expected trajectories of the end effector in the two simulation models. To demonstrate the performance of the HLMGA method, the particle swarm optimization algorithm was used to solve the forward kinematic solution of the linear trajectory for comparison. The simulation results show that, for the generalized kinematics, the error of the inverse kinematic solution is greater than 0.02 m. This indicates that the generalized kinematic model is not suitable for the CDPR with adaptive pulleys. Moreover, based on the hybrid Levenberg–Marquardt algorithm and genetic algorithm, the position and posture accuracy of the end effector are, respectively, lower than 5×10^{-8} mm and 1×10^{-7} radians. By comparing these results with the results of the particle swarm optimization algorithm, it is shown that the accuracy of the HLMGA method is relatively high. The simulation results demonstrate the high accuracy and overall effectiveness of the proposed method.

The current research only focuses on the kinematic modeling and solution of the CDPR. The dynamic characteristics and control are not investigated for the CDPR. Moreover, the effects of the elastic and nonlinear cable deformation on the kinematic accuracy are not considered in this study. Future research will address this area.

https://doi.org/10.3390/s26010039

Author Contributions: Conceptualization, Z.H. and J.P.; methodology, C.D., K.W. and Z.H.; software, K.W. and Z.H.; validation, C.D. and K.W.; formal analysis, Z.H. and J.P.; investigation, C.D., and Z.H.; resources, K.W. and J.P.; data curation, C.D. and K.W.; writing—original draft preparation, Z.H. and C.D.; writing—review and editing, K.W. and J.P.; visualization, K.W.; supervision, Z.H. and J.P.; project administration, J.P., C.D. and Z.H.; funding acquisition, Z.H. and J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (62303205), the National Key R&D Program of China under Grant 2022YFB4703100, project ZR2023QF137 sup-ported by Shandong Provincial Natural Science Foundation, the Guangdong Natural Science Foundation under Grant 2024A1515010540, the Shenzhen Science and Technology Program under Grant JCYJ20220530150006014, the University-Industry Collaborative Education Program under Grant 240506429202244, and Doctoral Startup Fund of Liaocheng University (318052231).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zarebidoki, M.; Dhupia, J.S.; Xu, W. A review of cable-driven parallel robots: Typical configurations, analysis techniques, and control methods. *IEEE Robot. Autom. Mag.* **2022**, *29*, 89–106. [CrossRef]
2. Wang, R.; Li, J.; Li, Y. A review on design, modeling and control technology of cable-driven parallel robots. *Robotics* **2025**, *14*, 116. [CrossRef]
3. Piao, J.; Kim, E.; Choi, H.; Moon, C.; Choi, E.; Park, J.; Kim, C. Indirect force control of a cable-driven parallel robot: Tension estimation using artificial neural network trained by force sensor measurements. *Sensors* **2019**, *19*, 2520. [CrossRef]
4. Peng, J.; Guo, Y.; Meng, D.; Han, Y. Kinematics, statics modeling and workspace analysis of a cable-driven hybrid robot. *Multibody Syst. Dyn.* **2024**, *61*, 163–193.
5. Orbea, D.; Cruz Ulloa, C.; Del Cerro, J.; Barrientos, A. RUDE-AL: Roped UGV deployment algorithm of an MCDPR for sinkhole exploration. *Sensors* **2023**, *23*, 6487. [CrossRef]
6. Ha, S.; Woo, S.; Kim, M.; Piao, J.; Kim, G.; Sarker, M.S.; Park, B.; Kim, S.; Jung, M.; Hong, K.; et al. A Cable-driven Parallel Robot for High-rack Logistics Automation. *Int. J. Control Autom. Syst.* **2024**, *22*, 3329–3340. [CrossRef]
7. Sun, G.; Liu, Z.; Gao, H.; Deng, Z. FFS-CDPR: A free-floating simulator based on cable-driven parallel robots coordinating vibration from discontinuous external wrench. *Acta Astronaut.* **2025**, *232*, 80–93. [CrossRef]
8. Shah, M.F.; Hussain, S.; Goecke, R.; Jamwal, P.K. Mechanism design and control of shoulder rehabilitation robots: A review. *IEEE Trans. Med. Robot. Bionics* **2023**, *5*, 780–792. [CrossRef]
9. Lee, C.; Gwak, K. Design of a novel cable-driven parallel robot for 3D printing building construction. *Int. J. Adv. Manuf. Technol.* **2022**, *123*, 4353–4366. [CrossRef]
10. Qian, S.; Bao, K.; Zi, B.; Wang, N. Kinematic calibration of a cable-driven parallel robot for 3D printing. *Sensors* **2018**, *18*, 2898. [CrossRef]
11. Sancak, C.; Itik, M. Out-of-plane vibration suppression and position control of a planar cable-driven Robot. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 1311–1320. [CrossRef]
12. Mattioni, V.; Ida, E.; Carricato, M. Design of a planar cable-driven parallel robot for non-contact tasks. *Appl. Sci.* **2021**, *11*, 9491. [CrossRef]
13. Job, J.; Ponnan, R.J.; Chandran, V.J.; Biju, N. Workspace analysis of an underconstrained cable-driven parallel robot with four cables. *Constr. Robot.* **2022**, *6*, 109–119. [CrossRef]
14. Hwang, S.W.; Kim, D.H.; Park, J.; Park, J.H. Equilibrium configuration analysis and equilibrium-based trajectory generation method for under-constrained cable-driven parallel robot. *IEEE Access* **2022**, *10*, 112134–112149.
15. Gagliardini, L.; Caro, S.; Gouttefarde, M.; Wenger, P.; Girin, A. Optimal design of cable-driven parallel robots for large industrial structures. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 5744–5749.
16. Lin, J.; Wu, C.Y.; Chang, J. Design and implementation of a multi-degrees-of-freedom cable-driven parallel robot with gripper. *Int. J. Adv. Robot. Syst.* **2018**, *15*, 1729881418803845. [CrossRef]

https://doi.org/10.3390/s26010039

17. Wu, Q.; Zi, B.; Hu, B.; Li, Y. Design, Kinematics and Stiffness Analysis of a Reconfigurable Cable-Driven Parallel Robot. *Chin. J. Mech. Eng.* **2025**, *38*, 98. [CrossRef]
18. Zhang, Z.; Xie, G.; Shao, Z.; Gosselin, C. Kinematic calibration of cable-driven parallel robots considering the pulley kinematics. *Mech. Mach. Theory* **2022**, *169*, 104648. [CrossRef]
19. Martin-Parra, A.; Juarez-Perez, S.; Gonzalez-Rodriguez, A.; Gonzalez-Rodriguez, A.G.; Lopez-Diaz, A.I.; Rubio-Gomez, G. A novel design for fully constrained planar Cable-Driven Parallel Robots to increase their wrench-feasible workspace. *Mech. Mach. Theory* **2023**, *180*, 105159.
20. Li, J.; Zhao, Y.; Tang, Q.; Sun, W.; Yuan, F.; Lu, X. Conceptual design and error analysis of a cable-driven parallel robot. *Robotica* **2022**, *40*, 2152–2167. [CrossRef]
21. Merlet, J. The kinematics of cable-driven parallel robots with sagging cables: Preliminary results. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 1593–1598.
22. Merlet, J. On the inverse kinematics of cable-driven parallel robots with up to 6 sagging cables. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 4356–4361.
23. Zi, B.; Yin, G.; Zhang, D. Design and optimization of a hybrid-driven waist rehabilitation robot. *Sensors* **2016**, *16*, 2121. [CrossRef]
24. Wang, W.; Zhao, J.; Li, Y.; Zi, B. Kinematic modeling and workspace analysis of a cable-driven parallel robot considering environmental collisions. *J. Mech. Robot.* **2025**, *17*, 91011. [CrossRef]
25. Li, J.; Li, Y. The structural design, kinematics, and workspace analysis of a novel rod–cable hybrid cable-driven parallel robot. *Biomimetics* **2025**, *10*, 4.
26. Pott, A.; Schmidt, V. On the forward kinematics of cable-driven parallel robots. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 3182–3187.
27. Merlet, J. A generic numerical continuation scheme for solving the direct kinematics of cable-driven parallel robot with deformable cables. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Republic of Korea, 9–14 October 2016; pp. 4337–4343.
28. Briot, S.; Merlet, J. Direct kinematic singularities and stability analysis of sagging cable-driven parallel robots. *IEEE Trans. Robot.* **2023**, *39*, 2240–2254. [CrossRef]
29. Baskar, A.; Plecnik, M.; Hauenstein, J.D.; Wampler, C.W. A numerical continuation approach using monodromy to solve the forward kinematics of cable-driven parallel robots with sagging cables. *Mech. Mach. Theory* **2024**, *195*, 105609. [CrossRef]
30. Piao, J.; Kim, M.; Kim, E.; Kim, C. A self-adaptive inertia hybrid control of a fully constrained cable-driven parallel robot. *IEEE Trans. Autom. Sci. Eng.* **2024**, *21*, 2063–2073. [CrossRef]
31. Marufkhani, H.; Khosravi, M.A. Robust optimal constrained control of fully-constrained cable-driven parallel robots based on GSDRE. *Nonlinear Dyn.* **2023**, *111*, 16159–16174. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

https://doi.org/10.3390/s26010039

sensors

MDPI

Article

Real-Time 6D Pose Estimation and Multi-Target Tracking for Low-Cost Multi-Robot System

Bo Shan [1], Donghui Zhao [1,2,*], Ruijin Zhao [1] and Yokoi Hiroshi [3]

[1] School of Electrical Engineering, Shenyang University of Technology, Shenyang 110178, China
[2] Jianghuai Advanced Technology Center, Hefei 230088, China
[3] Department of Mechanical Engineering and Intelligent Systems, University of Electro-Communications, Tokyo 182-8585, Japan
* Correspondence: putongdeyu@126.com

Abstract

In the research field of multi-robot cooperation, reliable and low-cost motion capture is crucial for system development and validation. To address the high costs of traditional motion capture systems, this study proposes a real-time 6D pose estimation and tracking method for multi-robot systems based on YolPnP-FT. Using only an Intel RealSense D435i depth camera, the system achieves simultaneous robot classification, 6D pose estimation, and multi-target tracking in real-world environments. The YolPnP-FT pipeline introduces a keypoint confidence filtering strategy (PnP-FT) at the output of the YOLOv8 detection head and employs Gaussian-penalized Soft-NMS to enhance robustness under partial occlusion. Based on these detection results, a linearly weighted combination of Mahalanobis distance and cosine distance enables stable ID assignment in visually similar multi-robot scenarios. Experimental results show that, at a camera height below 2.5 m, the system achieves an average position error of less than 0.009 m and an average angular error of less than 4.2°, with a stable tracking frame rate of 19.8 FPS at 1920 × 1080 resolution. Furthermore, the perception outputs are validated in a CoppeliaSim-based simulation environment, confirming their utility for downstream coordination tasks. These results demonstrate that the proposed method provides a low-cost, real-time, and deployable perception solution for multi-robot systems.

Keywords: multi-robot system; RGB sensing; 6D pose estimation; multi-target tracking; real-time perception; low-cost perception

1. Introduction

With the ongoing advancement of automation and intelligence, multi-robot systems are being increasingly deployed in military, industrial, and agricultural applications. These systems critically rely on real-time and accurate state perception to provide essential 6D pose (position and orientation) and identity information for each robot. However, current mainstream motion capture (mocap) systems face significant limitations. Commercial optical mocap systems, such as Vicon [1], offer sub-millimeter accuracy but suffer from high cost, strong dependence on controlled environments (e.g., fixed cameras, no occlusion, stable lighting), and the requirement for specialized markers—severely restricting their scalability in open, dynamic scenarios. Meanwhile, markerless deep learning-based pose estimation methods, while more flexible, often exhibit sensitivity to partial occlusion and frequent identity switches (ID switches) in multi-target settings. Consequently, there is

a pressing need for an integrated perception solution that balances accuracy, robustness, scalability, and cost-effectiveness.

In this context, robot localization methods can be broadly categorized into two paradigms: absolute and relative localization, each with distinct trade-offs. Relative localization relies on internal sensors [2] to estimate displacement and heading based on an initial pose. For instance, odometry using encoders and inertial sensors [3,4] is simple and low-cost, suitable for local navigation and short-term tasks; however, it is prone to cumulative errors and depends heavily on the initial pose. In contrast, absolute localization determines robot pose by recognizing environmental landmarks, making it suitable when the initial pose is unknown. LiDAR-based localization [5] provides accurate 3D positioning and is robust to lighting and electromagnetic interference [6], yet it incurs high computational overhead, lacks semantic information, requires global reference points for high accuracy, and entails substantial hardware costs. In practice, different localization strategies are selected based on application requirements. While multi-sensor fusion can combine the strengths of relative and absolute approaches to enhance information richness and estimation accuracy [7,8], it introduces new challenges, including strong dependency on model design, high computational demands for data alignment, increased system cost, and spatial deployment constraints.

Table 1 summarizes the key characteristics of representative visual localization methods. Current approaches either suffer from high cost and stringent environmental constraints (e.g., Vicon) or lack multi-target 6D tracking capability (e.g., GDR-Net, FairMOT). Particularly in visually homogeneous and resource-constrained multi-robot scenarios, there remains a critical gap for an integrated perception solution that simultaneously ensures 6D accuracy, ID stability, and deployment affordability—this constitutes the core motivation of our work.

Table 1. Comparison of the advantages of Different Visual Positioning Methods.

Method	6D Pose Est.	6D Pose Track.	Multi Target	Real Time	Hardware Cost	Camera Setup	Env. Constraints
Vicon [1]	Yes	Yes	Yes	Yes	Very High	Multi-camera, fixed	Controlled (no occlusion, stable light)
LOAM [5]	No	No	Limited	Yes	High	LiDAR only	Outdoor/indoor, but needs geometry
GDR-Net [9]	Yes	No	No	No	Medium	Single RGB	Controlled lab
PVN3D [10]	Yes	No	No	Yes	Medium	Single RGB-D	Moderate occlusion OK
FairMOT [11]	No	No	Yes	Yes	Medium	Single RGB	General
DeepSORT [12]	No	No	Yes	Yes	Low	Single RGB	General
Ours	Yes	Yes	Yes	Yes	Low	Single RGB	Indoor, partial occlusion OK

To address these challenges, this study proposes YolPnP-FT, a low-cost, and deployable integrated perception system tailored for top-down multi-robot scenarios. We do not introduce a novel 6D pose estimation algorithm or a new tracking architecture. Instead, we integrate and optimize existing components—YOLOv8, a keypoint confidence filtering strategy (PnP-FT), and DeepSORT—to construct an end-to-end pipeline for multi-target 6D pose estimation, tracking, and task-level validation. The main contributions are summarized as follows:

- We present a low-cost semi-physical verification platform for multi-robot systems, integrating an affordable RGB-D camera (Intel D435i), a standard GPU workstation, and multiple Mecanum-wheeled robots. This setup supports closed-loop validation from real-world perception to simulation-based control, lowering the barrier for algorithm development and system testing.

- We propose the YolPnP-FT perception pipeline, which introduces a keypoint confidence filtering strategy (PnP-FT) after the YOLOv8 detection head. This strategy effectively enhances the robustness of PnP-based pose estimation under partial occlusion while reducing unnecessary computation.
- We adapt the DeepSORT tracker by employing a linearly weighted combination of Mahalanobis distance and cosine distance with cascaded matching. This strategy enables the system to achieve stable ID assignment in visually similar multi-robot scenarios, effectively mitigating tracking instability and identity confusion caused by occlusion or appearance homogeneity.
- We establish a perception-driven validation paradigm: real-time perception outputs (6D pose + ID) are directly injected into a CoppeliaSim simulation environment to verify their direct utility in downstream tasks such as formation control and path planning, offering a reproducible deployment-validation pathway for low-cost systems.

In summary, this paper presents a low-cost, deployable 6D perception system for specific multi-robot applications. By integrating and optimizing existing techniques, we evaluate pose accuracy and tracking stability in real-world environments and validate task-level performance through simulation-based closed-loop testing. Experimental results demonstrate that the proposed solution achieves an effective balance between affordability and practicality.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 details the proposed methodology; Section 4 presents experimental results and analysis; and Section 5 concludes the study.

2. Related Works

Achieving robust and efficient perception in complex environments is essential for enhancing the adaptability and collaborative performance of multi-robot systems. In this context, 6D pose estimation and multi-object tracking (MOT) serve as two foundational pillars. While significant progress has been made in each subfield independently, there remains a critical gap in integrated solutions tailored for visually homogeneous, resource-constrained multi-robot scenarios.

2.1. Six-Dimensional Pose Estimation

Six-dimensional pose estimation is a key enabler for precise navigation and interaction in multi-robot systems. Recent advances in deep learning have substantially advanced this field. PoseNet [13] was among the first to employ convolutional neural networks (CNNs) for end-to-end 6D pose regression directly from RGB images, demonstrating the feasibility of learning-based approaches. PoseCNN [14] introduced direct pose regression but suffered from limited generalization and accuracy. Deep-6Dpose [15], built upon Mask R-CNN, added a dedicated pose prediction branch to streamline the end-to-end pipeline. YOLO-6D [16] reformulated pose estimation as a keypoint regression task within the real-time YOLOv2 framework, achieving high inference speed but exhibiting sensitivity to symmetric objects and partial occlusion. To better handle occlusion, ZebraPose [17] proposed a dense surface representation using discrete descriptors, showing improved robustness under controlled benchmarks. EfficientPose [18] reduced computational overhead by predicting translation and rotation through auxiliary sub-networks, though its performance degraded under severe occlusion. GDR-Net [10] introduced a geometry-guided approach that leveraged dense correspondences as an intermediate representation, achieving state-of-the-art results in the BOP2022 challenge; however, it required extensive post-processing and was not designed for real-time deployment. DPOD [19] combined detection with dense matching for single-RGB pose estimation and demonstrated robustness

to natural occlusions. RNNPose [20] employed recurrent networks for iterative refinement, while PointNet++ [21] learned geometric features directly from point clouds but incurred high computational cost. PVN3D [11] and DenseFusion [22] adopted two-stage pipelines—using least-squares fitting or voting mechanisms—and achieved strong performance on benchmarks like YCB-Video.

However, these methods share a common limitation: they primarily focus on single-object pose estimation [23]. None have explicitly addressed the challenges of multi-target association or maintaining temporal identity consistency—capabilities that are crucial for algorithm-driven multi-robot cooperation in intelligent task allocation and formation planning.

2.2. Multi-Object Tracking (MOT)

In multi-robot systems, MOT ensures consistent cross-frame identity assignment and real-time localization, which are critical for collaborative tasks. JDE [24] proposed an end-to-end framework that jointly performs detection and feature embedding, reducing system complexity. FairMOT [12] adopted a two-stage architecture with a "fair" matching strategy to handle occlusion and appearance changes. CenterTrack [25] relied on center-point association for fast tracking, while ChainedTracker [26] used graph-based modeling to resolve target associations in highly interactive scenarios. SORT [27] combined Kalman filtering with IoU-based matching, offering simplicity and efficiency but suffering from frequent ID switches under occlusion. DeepSORT [13] extended SORT by incorporating deep appearance descriptors and a cascaded matching strategy, significantly improving ID stability in pedestrian tracking benchmarks [28].

Nevertheless, existing MOT frameworks exhibit two major limitations in the context of multi-robot perception: They typically output only 2D bounding boxes or centroids, without providing 6D pose information; they heavily rely on appearance diversity for ReID feature learning. In multi-robot systems, robots often share visually homogeneous designs (e.g., similar chassis, color schemes), causing ReID features to lack discriminative power and leading to severe tracking failures in practice [29].

2.3. Research Gap

In summary, although both 6D pose estimation and MOT have advanced considerably, existing works largely optimize single-object pose accuracy or ID continuity in isolation. This fragmented approach fails to meet the integrated demands of resource-constrained, visually homogeneous multi-robot scenarios.

To address this gap, this study proposes a pragmatic integration of YOLOv8, a keypoint-based PnP solver with confidence filtering (PnP-FT), and an adapted DeepSORT tracker. Our system is designed for single top-down RGB-D camera setups, prioritizing deployability and reliability in real-world conditions over benchmark-leading performance. The primary goal is not to surpass SOTA methods on generic datasets, but to deliver a coherent, low-cost perception pipeline that enables robust 6D tracking for visually similar robots—a capability currently missing in the literature.

3. Method

3.1. System Overview

The proposed low-cost perception system for multi-robot formation validation comprises two complementary components:

(1) A physical hardware platform for real-time 6D pose estimation and tracking, consisting of an Intel RealSense D435i RGB-D camera (Santa Clara, CA, USA), a commercial GPU workstation, and multiple heterogeneous Mecanum-wheeled robots;

(2) A simulation validation environment based on CoppeliaSim, which receives the 6D pose and identity (ID) outputs from the physical perception system to drive virtual robot models.

This design establishes a perception-driven validation mechanism: the perception system operates independently in the real world, and its outputs are fed into the simulation environment to enable safe and reproducible verification of downstream coordination algorithms (e.g., formation control, path planning). The overall system architecture is illustrated in Figure 1, which clearly depicts the inter-component relationships and data flow.

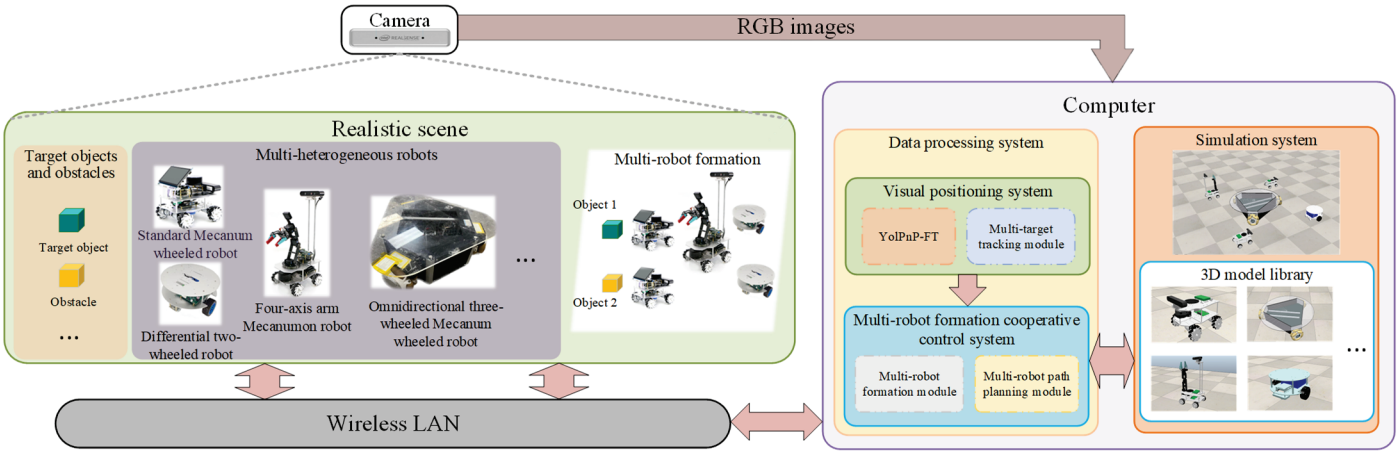


Figure 1. System Architecture for Perception-Driven Multi-Robot Validation: Real-world 6D pose and ID estimates from the YolPnP-FT pipeline (D435i camera) are injected into CoppeliaSim to drive virtual robots for downstream task validation.

3.2. A 6D Pose Estimation Method Based on the YolPnP-FT

This section introduces the proposed YolPnP-FT perception pipeline, with heterogeneous Mecanum-wheeled robots selected as targets due to their low cost, relatively complex kinematics, and visual homogeneity in multi-robot scenarios. The pipeline implementation consists of three sequential stages: (1) dataset construction and network training; (2) real-time 6D pose estimation using the trained YolPnP-FT network; (3) multi-object tracking to maintain cross-frame ID consistency.

3.2.1. Dataset Construction

This subsection details the dataset construction process for Mecanum-wheeled robots. The dataset focuses on typical indoor collaborative scenarios rather than extreme lighting or background clutter. Specifically, 700 images covering all semantic keypoints were collected under typical office lighting conditions (300–800 lux) using an Intel D435i camera mounted at heights of 0.5–2.5 m to simulate practical deployment scenarios.

For each robot type, a 3D CAD model was reconstructed in a virtual environment, and 22 semantic keypoints were defined to capture structural features (e.g., wheel hub centers, robotic arm joints). These 2D keypoints were manually annotated on extracted keyframes using Labelme, ensuring sufficient visibility under typical top-down viewpoints (see Figure 2 for an annotation example). Approximately 30% of the frames exhibit partial occlusion due to inter-robot occlusion, realistically reflecting multi-robot operational conditions.

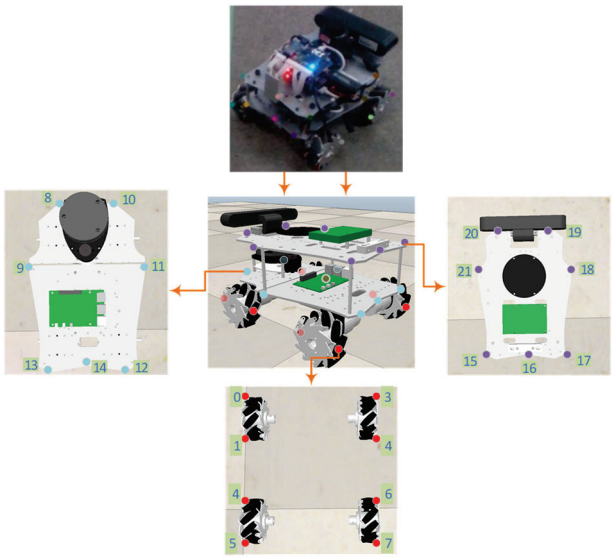


Figure 2. Example of 2D-3D Keypoint Correspondence: Illustration of the 22 semantic keypoints (e.g., wheel hub centers, arm joints) annotated on a Mecanum-wheeled robot under a top-down view, derived from its 3D CAD model.

3.2.2. Six-Dimensional Pose Estimation Based on the YolPnP-FT Perception Pipeline

This subsection details the proposed method for multi-robot 6D pose estimation. We formulate the YolPnP-FT perception pipeline as an integrated system for multi-object detection and 6D pose estimation, comprising three core components: a YOLOv8 detector, Gaussian-penalized Soft-NMS, and a non-learnable keypoint confidence filtering strategy (Perspective-n-Point with Filtered Threshold, PnP-FT). The overall architecture is illustrated in Figure 3.

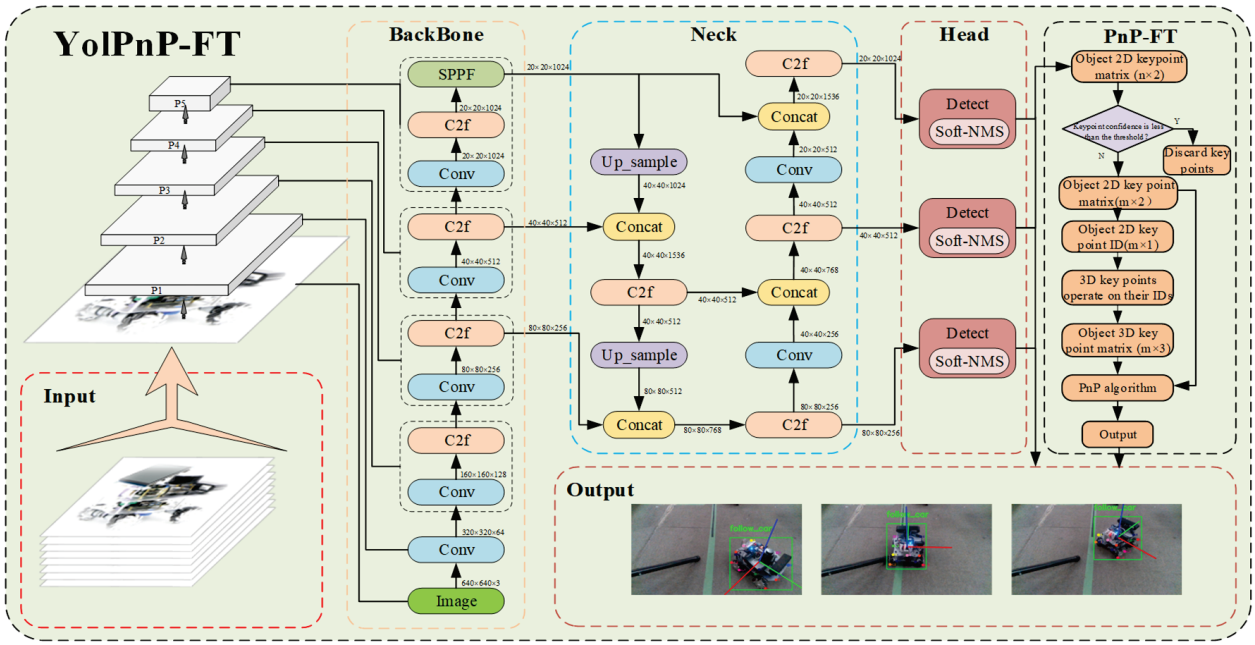


Figure 3. Network Architecture for 6D Pose Analysis Based on YolPnP-FT.

As shown in Figure 3, the input consists of video frame sequences. The Backbone employs an enhanced C2F module for feature extraction, the Neck performs multi-scale feature fusion, and the Head adopts a decoupled design to separately handle localization and classification tasks. In contrast to the standard YOLOv8 network—which is specialized

for object detection and outputs class labels, keypoints, and segmentation masks—we introduce a lightweight, non-learnable PnP-FT strategy module immediately after the detection head. This module filters out low-confidence keypoints prior to PnP-based pose estimation, thereby enhancing robustness under partial occlusion. The keypoint matching workflow of the PnP-FT strategy is illustrated in Figure 4.

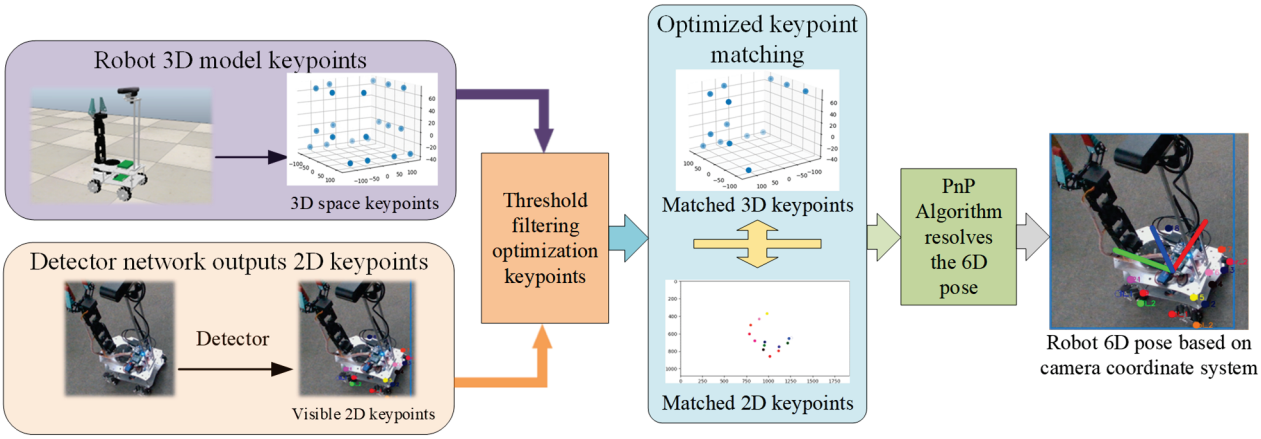


Figure 4. Schematic Diagram of the Keypoint Matching Algorithm for PnP-FT 6D Pose Estimation.

To address the frequent keypoints occlusion caused by complex robot configurations and top-down camera viewpoints—particularly in dense multi-robot formations where bounding box overlap (IoU > 0.5) is common—we embed Gaussian-penalized Soft-NMS [30] into the detection head, replacing conventional hard NMS or linear Soft-NMS. In such scenarios, hard NMS tends to over-suppress overlapping detections, while linear Soft-NMS (Equation (1)) causes abrupt score decay that destabilizes downstream keypoint matching. In contrast, Gaussian-penalized Soft-NMS (Equation (2)) provides smooth confidence attenuation, preserving recall while maintaining score discriminability—critical for the stability of PnP-FT inputs. This design aligns with practices in recall-sensitive pose estimation systems (e.g., EfficientPose [14]).

Specifically, candidate detection boxes are first sorted in descending order of confidence. The box with the highest confidence serves as the reference. The IoU between this reference box and all other candidates is then computed. When the IoU exceeds a preset threshold, the confidence of the overlapping box is decayed using the Gaussian Soft-NMS formulation. If the decayed confidence falls below a secondary threshold, the candidate is discarded. This process iterates until all boxes are processed.

$$s_j' = \begin{cases} s_j & \text{if } \mathrm{IoU}(b_i, b_j) < \mathrm{IoU}_{\text{threshold}} \\ s_j \times (1 - \mathrm{IoU}(b_i, b_j)) & \text{otherwise} \end{cases} \tag{1}$$

$$s'_j = s_j \exp\left(-\frac{\mathrm{IOU}^2(b_j, b_i)}{\sigma}\right) \tag{2}$$

$s_j^{'}$ represents the adjusted confidence score, s_j is the original confidence score, b_j denotes overlapping candidate boxes, b_i is the current box, and σ is a hyperparameter controlling the rate of score decay. Linear penalty is suitable for simple scenarios with high real-time performance and limited resources, whereas Gaussian penalty is better for complex scenes requiring high-precision detection. Given the complexity of multi-robot systems and precision needs, we choose Gaussian penalty to maintain detection continuity, avoid sudden score drops, improve recall rates, and enhance system robustness.

Due to occlusion induced by camera angles and robot formations, traditional PnP algorithms often suffer from degraded performance. To mitigate this, we propose a geometry-aware keypoint filtering strategy (formulated in Equations (5) and (6)) that evaluates whether the relative positions among neighboring keypoints conform to the robot's known 3D structural prior. Invalid or structurally implausible keypoints are discarded before PnP pose estimation, thereby improving both accuracy and stability.

Figure 5 illustrates the computational workflow of the PnP-FT 6D pose estimation algorithm, with its corresponding pseudocode provided in Algorithm 1. Beginning with the image data captured by camera, each frame of image data, denoted as I, is processed by a trained keypoint detector to extract 2D keypoints $p_i(X_c, Y_c, Z_c)$ from the current video frame. These keypoints form a keypoint matrix, as shown in Equation (3).

$$K_{2D} = D(I) \tag{3}$$

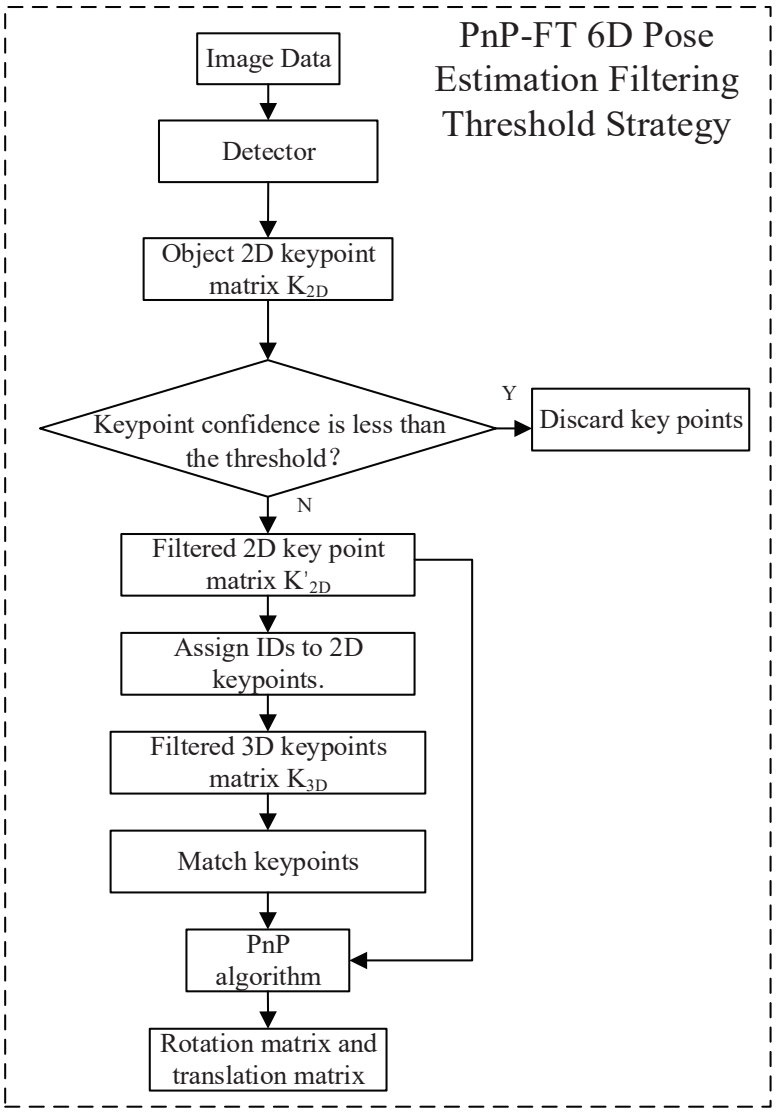


Figure 5. Keypoint Confidence Filtering Strategy (PnP-FT) in the YolPnP-FT Pipeline.

Subsequently, to ensure the reliability of the 2D keypoints, the confidence of each keypoint is evaluated. A confidence threshold, denoted as τ, is applied, and only keypoints

with a confidence higher than the threshold are retained, forming a filtered keypoint matrix $K_{2D}^{'}$ (Equation (4)) as follows.

$$K^{'}{}_{2D} = \left\{ k_{2D}^{(i)} \mid C\left(k_{2D}^{(i)}\right) \geq \tau \right\}, i = 1, 2, \ldots, n \tag{4}$$

The confidence function $C\left(k_{2D}^{(i)}\right)$ is defined to assess and filter the confidence of the 2D keypoints. This function takes into account the response strength of keypoint detection as well as the geometric consistency with adjacent keypoints. The specific definition is as shown in Equation (5):

$$C(k_{2D}^{(i)}) = \alpha \cdot S(k_{2D}^{(i)}) + \beta \cdot G(k_{2D}^{(i)}) \tag{5}$$

Here, α and β are weighting coefficients. The parameter α controls the tolerance to inter-keypoint deviations: higher values impose stricter geometric consistency constraints by penalizing larger displacements, whereas lower values allow more lenient matching. α should be adaptively tuned according to the actual keypoint distribution in the scene; in cases with significant appearance variations (e.g., due to lighting changes or texture ambiguity), increasing α can help prioritize reliable detection confidence over geometric fit. The parameter β governs the weight of geometric consistency in the overall association score. A higher β emphasizes the spatial configuration of keypoints, thereby enhancing robustness to challenging conditions such as occlusion or viewpoint changes—common in multi-robot interaction scenarios.

Algorithm 1: PnP-FT 6D Pose Estimation Filtering Threshold Strategy

Input: *Image data I, 3D keypoint model K_{3D} confidence threshold θ*
Output: *Rotation matrix R, translation vector t*

```
K_2D ← Detector(I);
// Detect 2D keypoints from image
if confidence (k) < for any k ∈ K_2D then
    Discard low-confidence keypoints;
K'_2D ← {k∈K_2D | confidence(k)≥θ};
Assign IDs to keypoints in K'_2D
K_3D^filtered ← GetMatching 3D Keypoints (K'_2D, K_3D);
// Select corresponding 3D keypoints
Match keypoints between K'_2D and K_3D^filtered;
[R,t] ← PnP(K'_2D, K_3D^filtered);
// Solve for pose using PnP algorithm
return R,t;
```

$S\left(k_{2D}^{(i)}\right)$ represents the keypoint detection response strength—i.e., the raw confidence score. However, relying solely on the detection response strength $S\left(k_{2D}^{(i)}\right)$ is often insufficient to reliably assess keypoint quality. For example, under partial occlusion, the network may output keypoints with high confidence but significantly inaccurate locations. To address this limitation, we introduce a geometric consistency constraint $G\left(k_{2D}^{(i)}\right)$, which evaluates whether the relative positions between a keypoint and its neighboring keypoints conform to the robot's 3D structural prior, thereby filtering out structurally implausible keypoints. Considering that keypoint errors in real-world scenarios typically follow a distribution where "small deviations are common, while large structural distortions are rare," we model $G\left(k_{2D}^{(i)}\right)$ using an exponential decay function. This formulation remains tolerant to minor positional deviations while being highly sensitive to significant geometric

distortions, effectively suppressing the influence of outlier keypoints on subsequent 6D pose estimation.

The geometric consistency function is designed as shown in Equation (6):

$$G(k_{2D}^{(i)}) = \exp\left(-\frac{1}{|\mathcal{N}|}\sum_{\mathbf{k}_{2D}^{(j)}\in\mathcal{N}}\frac{\left\|k_{2D}^{(i)} - k_{2D}^{(j)}\right\| - d_{ij}}{\sigma}\right) \tag{6}$$

This approach ensures that only keypoints with high confidence are utilized in subsequent pose estimation and tracking processes, enhancing the robustness and accuracy of the system. Where $\mathcal{N}$ represents the neighborhood set of keypoint $k_{2D}^{(i)}$, d_{ij} is the desired distance between keypoints, and σ is a scale parameter that controls the sensitivity of geometric consistency. $\left\|k_{2D}^{(i)} - k_{2D}^{(j)}\right\|$ is the Euclidean distance between keypoint $k_{2D}^{(i)}$ and its adjacent keypoint $k_{2D}^{(j)}$. By adjusting the weight coefficients α and β, as well as the parameters d_{ij} and σ in geometric consistency, the method can be flexibly adapted to different application scenarios.

This approach filters out low-confidence keypoints, thereby improving the accuracy and robustness of pose estimation. Following confidence filtering, the retained 2D keypoints are assigned IDs, and the corresponding 3D keypoints $P_i(X_w, Y_w, Z_w)$ are matched to form a matched 3D keypoint matrix K_{3D}.

In summary, the threshold-based filtering and matching yield 2D keypoint matrix $K_{2D}^{'}$ composed of 2D keypoints $p_i(X_c, Y_c, Z_c)$ and 3D keypoint matrix K_{3D} composed of 3D keypoints $P_i(X_w, Y_w, Z_w)$. Subsequently, the PnP algorithm employing the DLT (Direct Linear Transform) method is utilized. As illustrated in Figure 6, the re-projection equation is derived based on the camera projection model and the pose relationship between the camera coordinate system and the world coordinate system, which can be expressed as shown in Equation (7):

$$\begin{aligned} p_i &= K[R \mid t]P_i \\ i &= 1, \ldots, n \end{aligned} \tag{7}$$

where K is the intrinsic matrix of the camera obtained through Zhang's calibration method. The coordinates $P_i(X_w, Y_w, Z_w)$ in the world coordinate system can be represented as homogeneous coordinates $[X_w \quad Y_w Z_w \quad 1]^T$ and the 2D keypoint $p_i(X_c, Y_c, Z_c)$ in the image coordinate system can be represented as homogeneous coordinates $[u \quad v \quad 1]^T$. The equation can be rearranged as shown in Equation (8):

$$\lambda \begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} = M \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{8}$$

where λ is the scaling factor, and M is the projection matrix, obtained by the product of the intrinsic matrix k and the extrinsic matrix $[R|t]$. The extrinsic parameters can be separated through inversion, as shown in Equation (9):

$$[R \mid t] = K^{-1} \cdot M \tag{9}$$

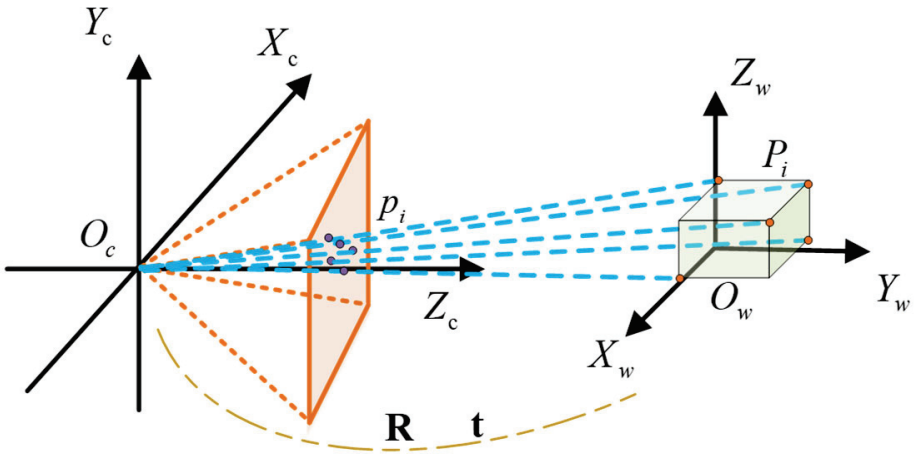


Figure 6. Schematic Diagram of the Camera Coordinate System and World Coordinate System Model.

To ensure that R is a valid rotation matrix, perform a singular value decomposition (SVD) on the 3 × 3 matrix R, as shown in Equation (10).

$$R = U \cdot \sum \cdot V^T \tag{10}$$

Further, to extract the translation vector t, as shown in Equation (11), multiply the inverse of the intrinsic matrix K^{-1} by the fourth column of M:

$$t = K^{-1} \cdot M[:, 4] \tag{11}$$

This method effectively addresses the limitations of traditional pose estimation techniques in handling partially occluded or invisible key points. By eliminating redundant or unreliable keypoints, it reduces computational load and enhances robustness in complex real-world environments.

3.2.3. Adapted DeepSORT for Multi-Object Tracking in Visually Homogeneous Multi-Robot Scenarios

The classic multi-object tracking algorithm SORT combines Kalman filtering with the Hungarian algorithm, using Intersection over Union (IoU) as the sole association metric. While computationally efficient, this approach is prone to frequent identity switches (ID switches) when targets undergo occlusion or temporarily leave and re-enter the field of view—common occurrences in complex multi-robot scenarios.

To address this limitation, we adapt the DeepSORT tracking framework, using the YolPnP-FT pipeline as the detector. Specifically, we employ a linearly weighted combination of Mahalanobis distance and cosine distance within a cascaded matching strategy to enhance ID stability in visually homogeneous multi-robot environments and provide reliable state estimates for downstream coordination tasks.

Although single-stage trackers such as FairMOT [23] offer end-to-end efficiency, they heavily rely on appearance diversity for ReID feature learning. In multi-robot systems, where robots often exhibit high visual similarity (e.g., identical chassis, color schemes, and structural symmetry), ReID features lack discriminative power, leading to degraded tracking performance. In contrast, DeepSORT's cascaded matching mechanism explicitly fuses motion and appearance cues, making it better suited for scenarios with low appearance variance—a characteristic inherent to visually similar robot fleets.

(1) Motion Model Information Feature Matching: Mahalanobis Distance

Using the multi-object bounding boxes from the YolPnP-FT detected, Mahalanobis distance enhances tracking stability by considering the covariance between data points, reducing scale variation effects. Therefore, the matching degree Equation (12) between Trajectory and Detection is as follows:

$$d^{(1)}(i,j) = (d_j - y_i)^T S_i^{-1}(d_j - y_i) \tag{12}$$

$$b_{i,j}^{(1)} = 1[d^{(1)}(i,j) \leq t^{(1)}] \tag{13}$$

The Mahalanobis distance $d^{(1)}(i,j)$ represents the motion matching degree between the state d_j of the j-th detection and the predicted observation y_i of the i-th trajectory at the current moment. S_i^{-1} represents the inverse covariance matrix predicted by the Kalman filter. The association between the tracking box and the detection box in terms of motion information is determined by Equation (13).

Considering the continuous motion of dynamic targets, the distance between consecutive frames remains small, allowing similar coordinate boxes to be identified as the same target. A threshold t = 0.95 is used to filter detections. If the distance between the detection box and the tracking box is less than t, they are considered associated, setting b = 1; otherwise, b = 0.

(2) Appearance Model Information Feature Matching: Cosine Distance

While the Mahalanobis distance is effective for well-defined target motions, real-world conditions such as camera movement and target uncertainty can degrade Kalman filter performance. When targets are close or occluded, bounding box distance and Mahalanobis distance become limitation. To alleviate this problem, the YolPnP-FT outputs is used to extract feature vectors, and cosine similarity is employed to quantify the feature differences between targets in adjacent frames. The calculation equation is as follows:

$$d^{(2)}(i,j) = \min\{1 - r_j^T r_k^{(i)} | r_k^{(i)} \in R_i\} \tag{14}$$

$$R_i = \{r_k^{(i)}\}_{k=1}^{L_k} \tag{15}$$

$d^{(2)}(i,j)$ (Equation (14)) represents the minimum cosine distance between the feature vector of the tracked target and the feature vector of the detected target in the current frame within set R_i. r_j denotes the appearance feature vector of the j-th detected result, and R_i. (Equation (15)) is the set of the most recent L_k feature vectors of the i-th tracked target. The cosine similarity $r_j^T r_k^{(i)}$, measures the appearance information, improving target ID prediction accuracy.

$$b_{i,j}^{(2)}(i,j) = 1[d^{(2)}(i,j) \leq t^{(2)}] \tag{16}$$

The threshold in the cosine metric is set according to Equation (16). When the metric value falls below $t^{(2)}$, we consider the detection box and the tracking box content to be successfully matched. Using cosine distance for correlation analysis can effectively address uncertainties in motion and camera vibrations, thus reducing the interference of external environments on system stability.

In summary, appearance information feature solves some limitations of Mahalanobis distance in target tracking, which makes the appearance information feature better at long-term tracking.

(3) Comprehensive Matching Degree

Combining the motion model and appearance model through linear weighting complements their strengths and weaknesses to form the final matching score. The combined score equation is as follows:

$$c(i,j) = \lambda d^{(1)}(i,j) + (1-\lambda) d^{(2)}(i,j) \tag{17}$$

The combined matching score $c(i,j)$ uses a linear weighting coefficient λ to bring both distance and feature metrics close to their minimum values. When motion and appearance information correlate, a link is established between tracking and detection boxes. The final association is defined by Equation (18).

$$b_{i,j} = \prod_{m=1}^{2} b_{i,j}^{(m)} \tag{18}$$

When the indicator b equals 1, the initial match is considered successful. The combined measurement leverages the strengths of both methods, considering both the distance similarity between bounding boxes and the content similarity within them.

(4) Cascade Matching Strategy

When targets are occluded for long periods, the performance of the Kalman filter is limited, leading to discontinuities in the probability of consecutive predictions. To improve the accuracy of the matches, we introduce a cascaded matching strategy. The cascaded matching strategy consists of two parts: upper and lower, as illustrated in Figure 7. In the upper part, using the motion model to calculate the distance between the bounding boxes of tracks and observations, construct the cost matrix. By linearly weighting $d^{(1)}(i,j)$ and $d^{(2)}(i,j)$, obtain the combined matching score $c(i,j)$. A threshold matrix is used to limit the higher values in the cost matrix. In the lower part, starting with tracks that have not been lost, prioritize these tracks and match tracks lost for the longest last. Unconfirmed tracks and those consistently tracked are matched with detection results based on Intersection over Union (IOU) to reduce matching errors caused by sudden appearance changes or partial occlusions. Throughout the process, matching priority is dynamically adjusted, with more frequently appearing targets receiving higher priority. In summary, unlike SORT, DeepSORT combines motion and appearance information for association, using cascade matching + IOU matching to enhance tracking accuracy and robustness. The algorithm steps are shown in the algorithm flowchart in Algorithm 2.

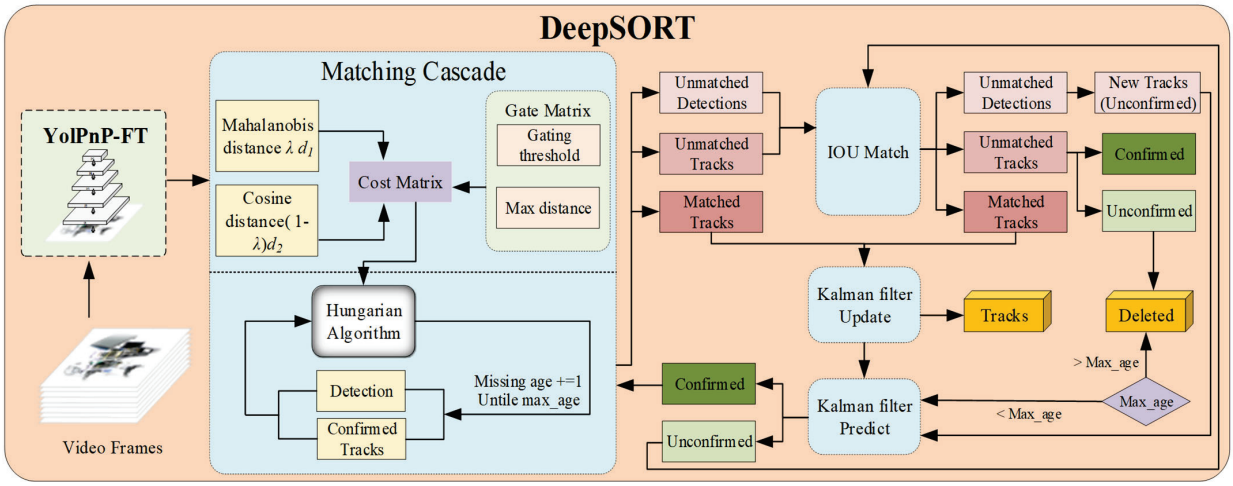


Figure 7. Architecture Diagram of the DeepSort Algorithm.

Algorithm 2: DeepSORT Algorithm

Input: *Frame* F_t, *detections* $D = (d_1, ..., d_M)$, *track set T, max age* A_{max}, *IOU threshold* λ
Output: *Updated track set T*

for *each track* $\tau_i \in$ T **do**
  τ_i . predict() *// Kalman prediction*
// Stage 1: Cascade Matching (appearance +motion)
$T_{\text{conf}} \leftarrow \{\tau_i \in T \mid \tau_i$.is_confirmed()};
$M_1, U_T, U_D \leftarrow \emptyset, T_{\text{conf}}$, D;
for n = 1 **to** $A_{\max}$ **do**
  $T_{\text{n}} \leftarrow \{\tau_i \in U_T | \tau_i$. age = n};
  if $T_{\text{n}} \neq \emptyset$ **then**
  Compute cost matrix C using appearance + Mahalanobis;
  $[x_{ij}] \leftarrow$ Hungarian(*C*);
  $M_n \leftarrow \{(i,j) | x_{ij} = 1$ and $c_{ij} <$ gate};
  $M_1 \leftarrow M_1 \cup M_n$;
  $U_T \leftarrow U_T \backslash \{i | (i,j) \in M_n\}$;
  $U_D \leftarrow U_D \backslash \{j | (i,j) \in M_n\}$;
// Stage2: IOU Matching for remaining confirmed tracks
$M_2, U_T, U_D \leftarrow$ IOUMatch(U_T, U_D, λ);
// Update matched tracks
for each $(i,j) \in \text{M}_1 \cup \text{M}_2$ **do**
  τ_i. update(d_j);
  τ_i. age ← 0;
// Handle unmatched confirmed tracks
for each $\tau_i \in U_T$where τ_i.is_confirmed() **do**
  τ_i . age ← τ_i. age + 1;
  if τ_i.age > A_{max} **then**
  T ← T\{τ_i};
// Create new unconfirmed tracks
for each $d_j \in U_D$ **do**
  $\tau_{new} \leftarrow$ newTrack(d_j);
  τ_{new} . confirmed ← false;
  τ_{new} . age ← 1;
  T ← T ∪ {τ_{new}};
// Manage unconfirmed tracks
for each $\tau_i \in$ T where $\neg\tau_i$.is_confirmed() **do**
  if τ_i.age > 1 **then**
  T ← T\{τ_i} *// Delete if not confirmed in 2 frames*
return T;

3.3. Downstream Task Validation of Perception Outputs

To validate the practical utility of the YolPnP-FT perception pipeline outputs, we inject the real-time 6D pose and identity ID information into a CoppeliaSim-based simulation environment, where they serve as state inputs for downstream coordination tasks such as formation control and path planning. This setup provides an intuitive and user-friendly interface for closed-loop algorithm validation without requiring physical robot deployment.

It is emphasized that the control and planning algorithms employed in simulation—namely, the leader–follower formation controller and the RRT path planner—are standard methods and do not constitute contributions of this work. Their sole purpose is to act as consumers of

the perception outputs, thereby verifying whether YolPnP-FT provides sufficiently accurate and stable state estimates to support real-world multi-robot coordination.

As illustrated in Figure 1, the RGB-D data stream from the Intel D435i camera is fed into the YolPnP-FT perception pipeline. The resulting 6D pose and ID estimates are transmitted to CoppeliaSim via a TCP/IP interface to drive virtual robot models. Consequently, the observed performance in simulation—such as formation stability and path feasibility—directly depends on the pose accuracy and ID consistency delivered by YolPnP-FT, thereby enabling task-level validation of the perception system.

4. Experiment and Analysis

4.1. Real-Time 6D Pose Estimation and Tracking Experiments for Multi-Robot Formations Based on the YolPnP-FT Perception Pipeline

This section presents a systematic evaluation of the proposed YolPnP-FT perception pipeline in real-world multi-robot scenarios. The experimental procedure consists of three stages: (1) construction of a robot-specific dataset and training of the detection network; (2) 6D pose estimation using the trained YolPnP-FT, with validation through comparison between ground-truth and estimated poses; (3) real-time multi-robot formation tracking experiments.

4.1.1. Experimental Environment

All experiments are implemented using the PyTorch deep learning framework on a Windows 10 system. The hardware configuration includes an NVIDIA GeForce RTX™ 3060 GPU (12 GB VRAM), an Intel® Core™ i7-12700 CPU (up to 4.90 GHz), and 32 GB RAM. Visual data are captured using an Intel® RealSense™ D435i depth camera, with only the RGB stream serving as input to the YolPnP-FT pipeline.

4.1.2. 6D Pose Estimation Experiments Based on the YolPnP-FT Perception Pipeline

Following the dataset construction procedure described in Section 3.2.1, the network is trained with a learning rate of 0.001, batch size of 16, IoU threshold of 0.5, and 300 training epochs. The robot keypoint dataset is split into training and validation sets at a 5:1 ratio. The training loss curve is shown in Figure 8. The loss decreases rapidly in early epochs and stabilizes after approximately 200 iterations, though minor fluctuations are observed. These fluctuations are typical in keypoint detection tasks and primarily stem from partial keypoint occlusion in certain frames, which leads to discontinuous supervision signals [Ultralytics, YOLOv8 Docs] [31]. Nonetheless, the overall convergence behavior confirms the effectiveness of the optimization process.

The best-performing weights from training are deployed in the YolPnP-FT pipeline. To evaluate model performance, a video sequence of a four-axis Mecanum-wheeled robot is processed, and intermediate outputs—including object detection, keypoint localization, and 6D pose estimation—are extracted programmatically (Figure 9). Results demonstrate that the pipeline effectively achieves simultaneous object classification, keypoint detection, and full 6D pose estimation.

To investigate the impact of camera height on visual localization accuracy—a critical factor in practical deployments—we conduct comparative experiments using a four-axis Mecanum-wheeled robot at three camera heights: 0.5 m, 1.5 m, and 2.5 m. In each setting, the robot executes a standard circular trajectory. The actual and estimated trajectories are compared (Figure 10), with the camera positioned directly above the world coordinate origin and aligned parallel to the ground plane to eliminate coordinate misalignment.

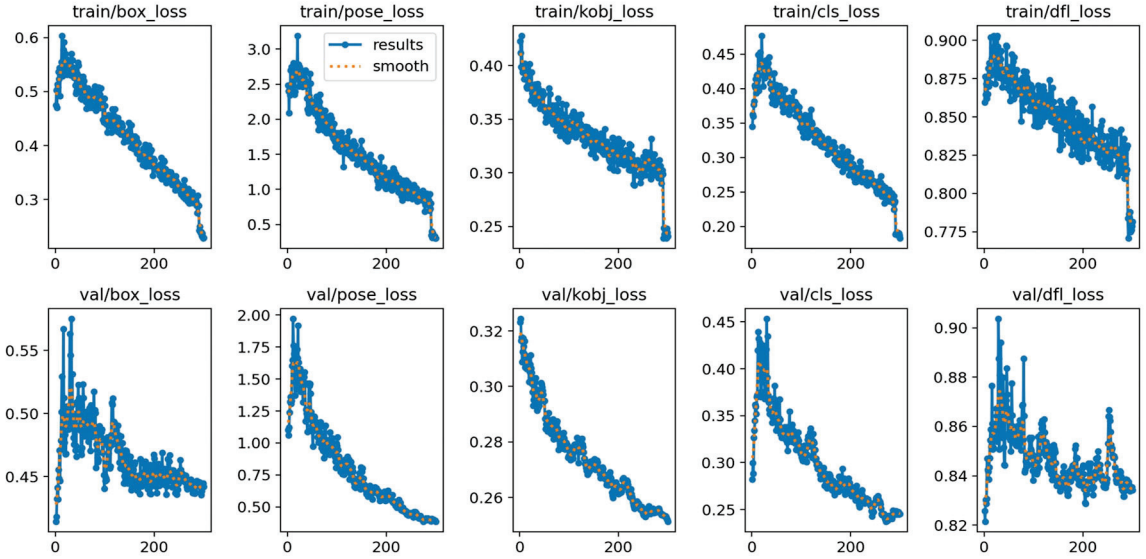


Figure 8. The Loss Function Curve of Network Training.

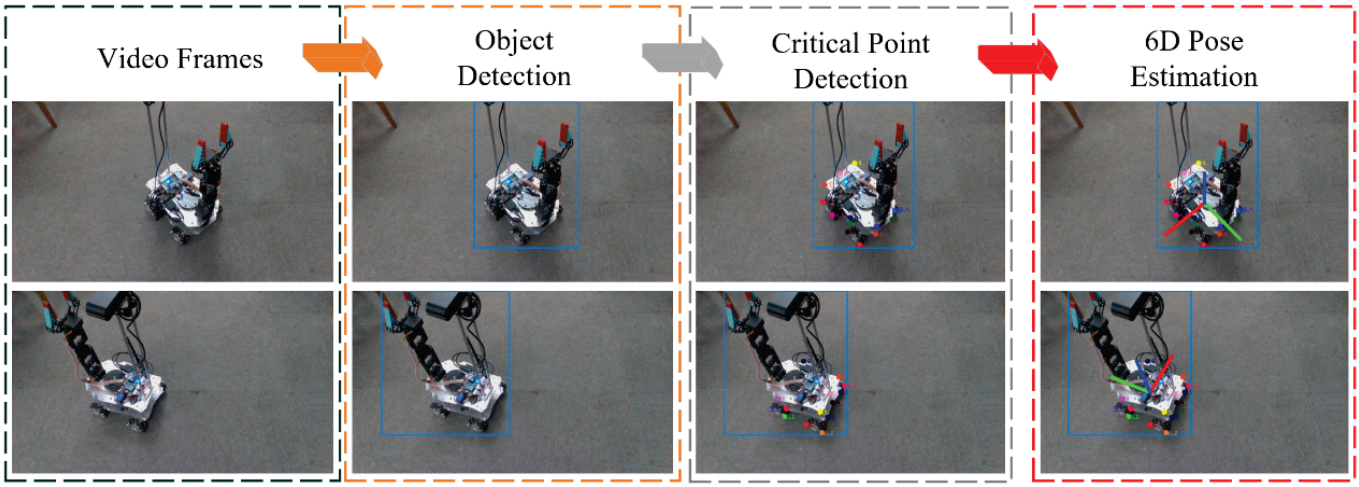


Figure 9. Illustrative Diagram of Staged Outputs from the YolPnP-FT Network.

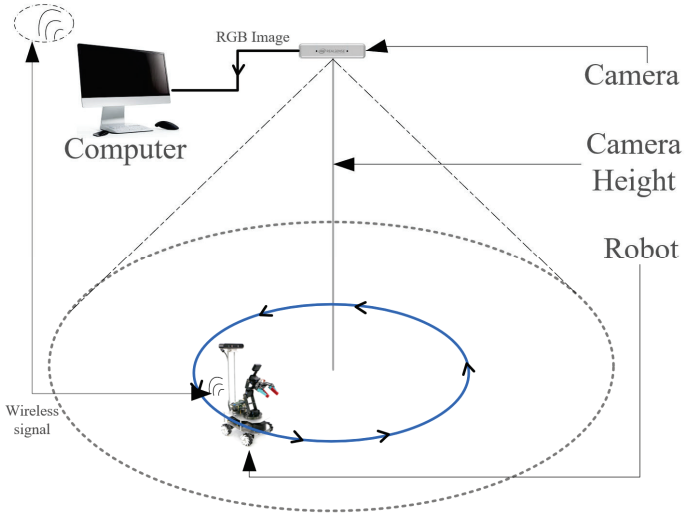


Figure 10. Scenario Diagram of 6D DoF Estimation Experiment Based on the YolPnP-FT Network.

As shown in Figure 11, the detected trajectories closely match the ground-truth circular paths across all heights, with only minor deviations. This indicates that the model provides reliable pose estimates from varying viewpoints. To quantify performance, we compute positional and angular errors between ground-truth and estimated poses (Figure 12).

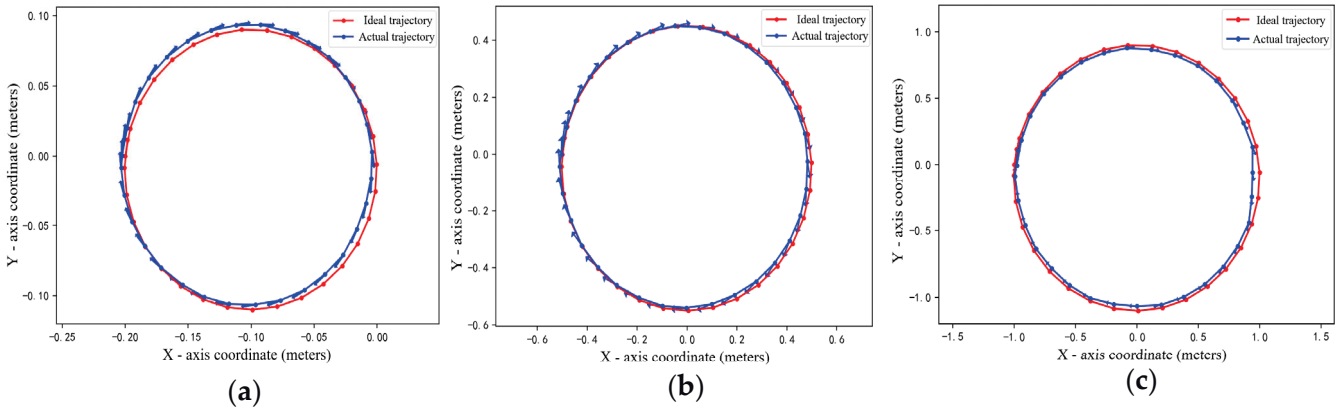


Figure 11. The Difference Between the Actual Circular Motion Trajectories and the Detected Trajectories Under Different Camera Height Settings: (**a**) 0.5 m, (**b**) 1.5 m, and (**c**) 2.5 m.

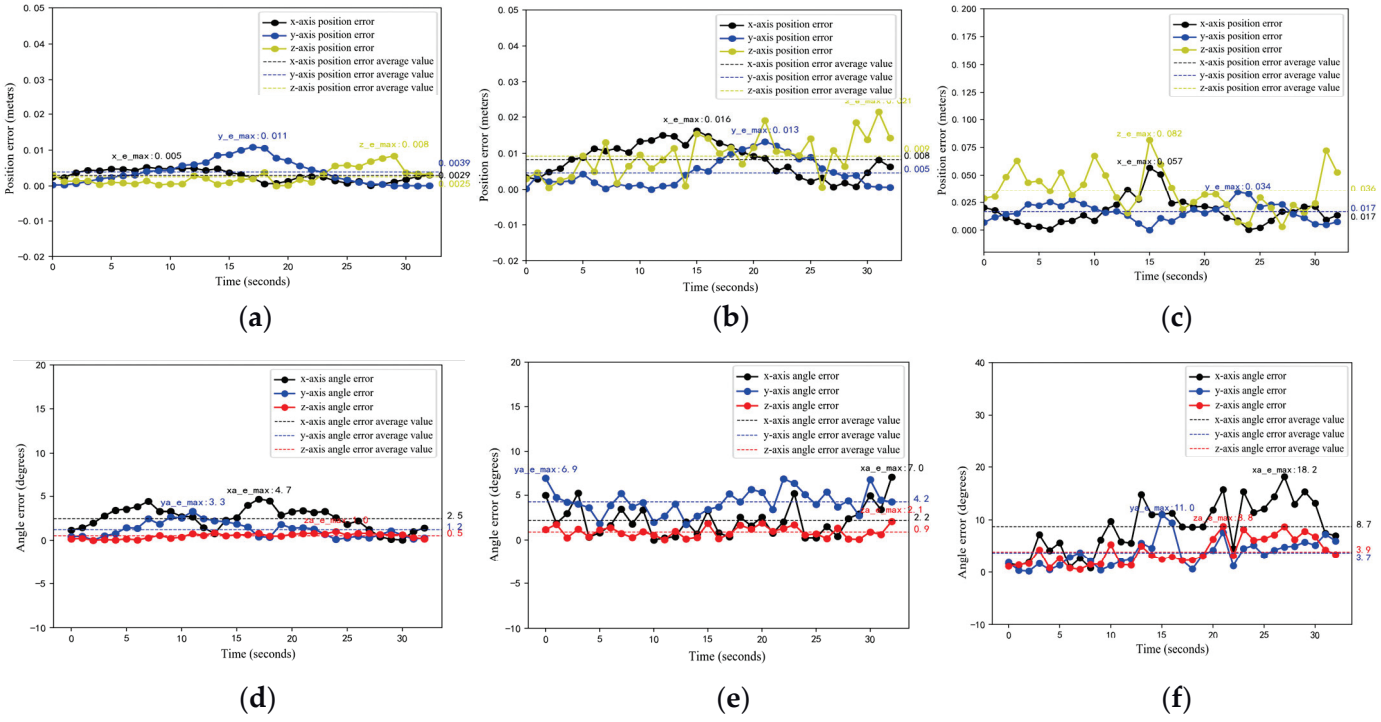


Figure 12. Position (meters) and Angle (degrees) Error Curves for the Robot's Motion Under Different Camera Heights. (**a**) Positional errors at a height of 0.5 m. (**b**) Positional errors at a height of 1.5 m. (**c**) Positional errors at a height of 2.5 m. (**d**) Angle errors at a height of 0.5 m. (**e**) Angle errors at a height of 1.5 m. (**f**) Angle errors at a height of 2.5 m.

Tables 2 and 3 summarize the position and orientation errors along the x, y, and z axes at different camera heights. A clear inverse relationship between camera height and localization accuracy is observed: as height increases from 0.5 m to 2.5 m, both positional and angular errors grow significantly. Specifically, the average *z*-axis position error rises from 0.0025 m to 0.043 m, the average pitch (*y*-axis) angle error increases from 1.2° to 6.5°, and the maximum roll (*x*-axis) angle error surges from 4.7° to 18.2°—consistent with the

well-established principle that visual pose estimation accuracy degrades with increasing observation distance.

Table 2. Position Error Between the Robot's Actual and Detected Positions Under Different Camera Heights.

Camera Height (m)	Mean X Position Error (m)	Maximum X Position Error (m)	Mean Y Position Error (m)	Maximum Y Position Error (m)	Mean Z Position Error (m)	Maximum Z Position Error (m)
0.5	0.0039	0.005	0.0025	0.011	0.0025	0.008
1.5	0.008	0.016	0.005	0.013	0.009	0.021
2.5	0.017	0.057	0.017	0.034	0.043	0.082

Table 3. Angle Error Between the Robot's Actual and Detected Positions Under Different Camera Heights.

Camera Height (m)	Mean X Angle Error (°)	Maximum X Angle Error (°)	Mean Y Angle Error (°)	Maximum Y Angle Error (°)	Mean Z Angle Error (°)	Maximum Z Angle Error (°)
0.5	2.5	4.7	1.2	3.3	0.5	1.0
1.5	2.2	7.0	4.2	6.9	0.9	2.1
2.5	6.6	18.2	6.5	12.6	3.9	8.8

We attribute this error growth to two primary mechanisms:

(1) Small-object effect: As camera height increases, the robot occupies fewer pixels in the image, causing minor 2D keypoint detection errors to be significantly amplified through PnP triangulation;
(2) Feature degradation: At high (near-top-down) viewpoints, the robot's upper surface often lacks texture or exhibits symmetry (e.g., planar surfaces), leading to ambiguous keypoint localization. In contrast, lower viewpoints, though more prone to occlusion, provide richer lateral geometric features that enhance matching reliability.

Notably, angular errors exhibit strong directional dependency: yaw (rotation about the z-axis) remains relatively stable across heights, as robot orientation can still be inferred from the planar layout of keypoints in the top-down view. In contrast, pitch and roll—dependent on side-view structures—suffer severe degradation at high viewpoints due to feature scarcity.

Considering that a 0.5 m height limits the field of view and hinders multi-robot coverage, we recommend a camera height of 1.5 m, which achieves an optimal trade-off between localization accuracy (average position error: 0.008 m; average angular error: 4.2°) and field-of-view coverage. This height is adopted in all subsequent multi-object tracking experiments.

4.1.3. Comprehensive Real-Time Multi-Robot Formation Tracking Experiments

To further evaluate real-time tracking performance, we conduct a multi-robot formation experiment using one four-axis Mecanum-wheeled robot as the leader and two standard Mecanum-wheeled robots as followers. The leader follows a pre-programmed trajectory, while followers track it using a standard formation control algorithm. The D435i camera captures the formation in real time, and the YolPnP-FT pipeline performs multi-robot recognition, 6D pose estimation, and tracking.

First, we compare tracking performance with and without the multi-object tracking module (Figure 13). Without tracking, the detector assigns IDs randomly in each frame,

resulting in severe ID instability. This is attributed to the visual homogeneity among robots, which violates the continuity requirement for reliable multi-robot localization. In contrast, the proposed system, which integrates the adapted DeepSORT tracker, enables consistent ID assignment and significantly improves tracking robustness.

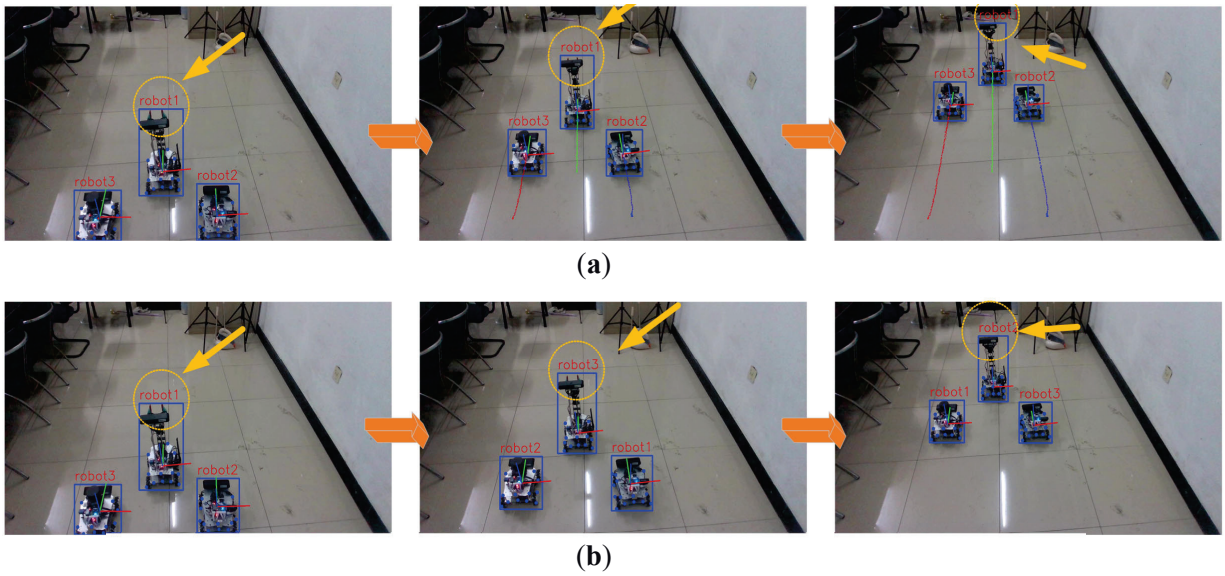


Figure 13. Comparison of Effects with and without the Multi-Target Algorithm. (**a**) The Effect of the Visual Positioning System after Incorporating the Multi-Target Real-Time Tracking Algorithm. (**b**) The Effect of the Visual Positioning System without the Multi-Target Real-Time Tracking Algorithm.

Next, we evaluate the impact of input resolution on system throughput. As shown in Figure 14, at 640 × 480 resolution, the system achieves over 30 FPS, satisfying real-time requirements. At 1920 × 1080 resolution, the frame rate reaches 38.2 FPS without tracking and drops to 19.8 FPS with tracking—still sufficient for real-time multi-robot applications.

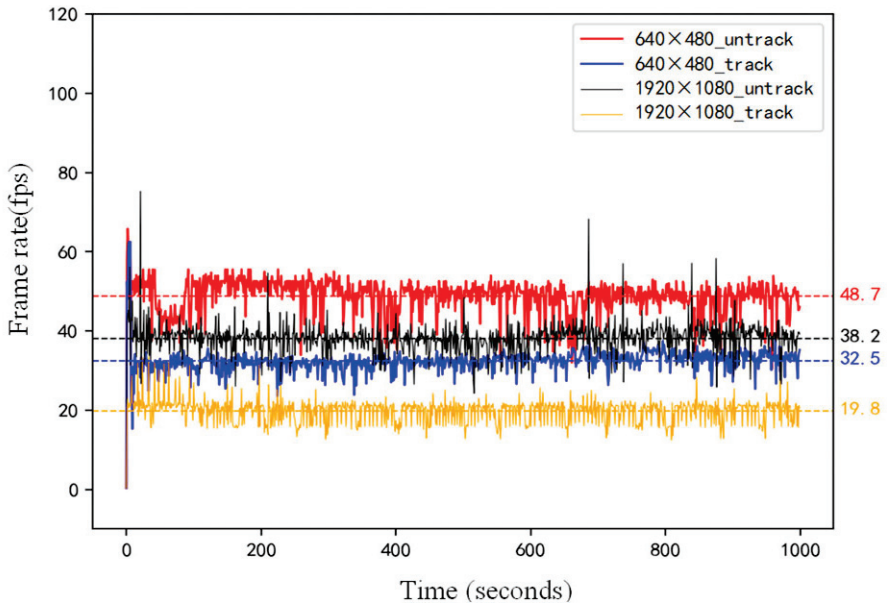


Figure 14. Frame Rate Plots of Multi-Target Real-Time Tracking Based on the YolPnP-FT Under Different Conditions.

Finally, we perform a comprehensive tracking experiment with heterogeneous robots (Figure 15). Figure 16 compares the actual and estimated trajectories of a follower robot

in the x–y plane. Results show that initial errors (up to 0.07 m) gradually converge to below 0.05 m. Notable error spikes at 18 s and 34 s (Figure 16b) are attributed to keypoint occlusion during rotational maneuvers and inertial effects from abrupt stops, respectively. These errors subsequently decay, demonstrating the system's self-correcting capability.

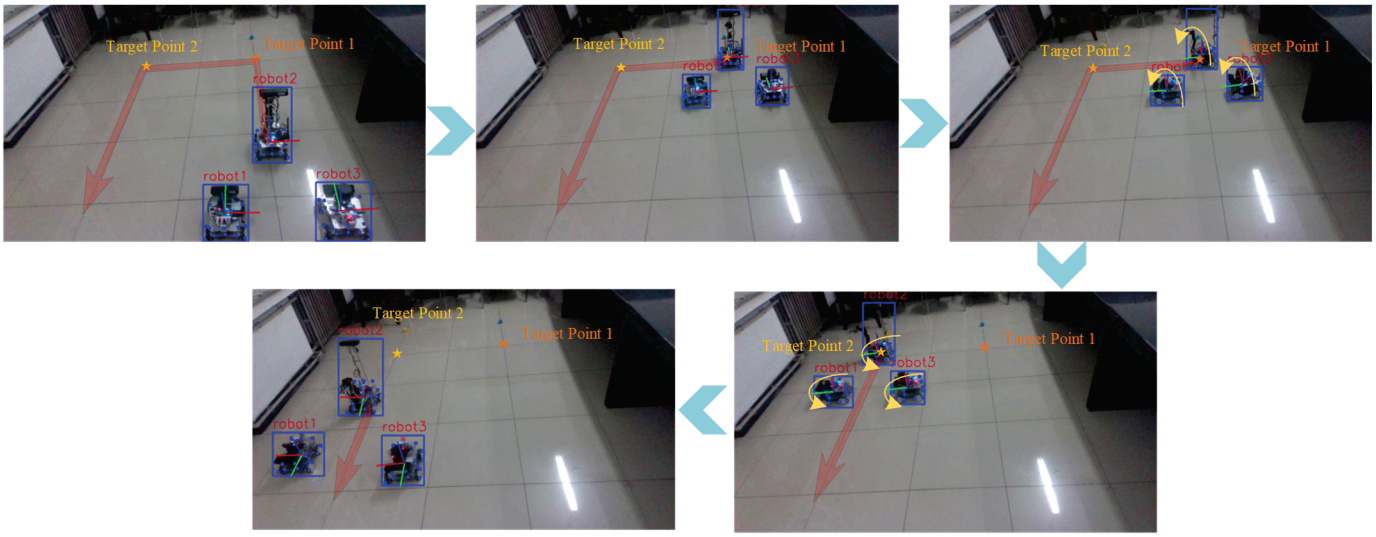
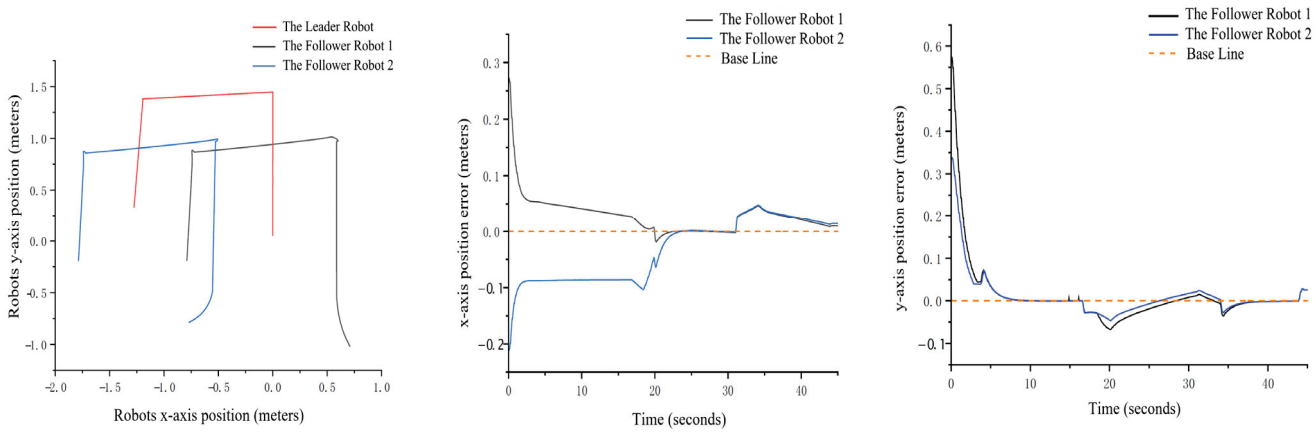


Figure 15. Comprehensive Experimental Process Diagram for Multi-Robot Formation Real-Time Tracking and Positioning.



(a) **(b)** **(c)**

Figure 16. Comparison of Actual and Positioning Trajectories for Multi-Robot Formation Movement. (**a**) Multi-Robot Formation Trajectories; (**b**) Baseline-Relative X-Position Error; (**c**) Baseline-Relative Y-Position Error.

In conclusion, high-precision 6D pose estimation is necessary but insufficient; stable ID tracking is the prerequisite for multi-robot coordination. Our experiments show that 640 × 480 resolution offers the best balance between computational load and perception quality. For high-precision tasks (e.g., fine manipulation), 1080p resolution can be used at the cost of frame rate. At 19.8 FPS, the system exhibits strong error convergence and robustness to transient disturbances, confirming its readiness for real-world deployment.

4.2. Perception-Driven Simulation Validation for Multi-Robot Systems

Section 4.1 validates the accuracy and robustness of the proposed perception system in real-world scenarios. To further demonstrate its practical utility, we inject the real-time 6D pose and identity (ID) outputs from YolPnP-FT into a CoppeliaSim-based simulation

environment to drive downstream coordination tasks, including formation control and path planning. It is emphasized that the leader–follower control law and the RRT path planner employed in this section are standard methods and do not constitute contributions of this work. They serve solely as consumers of the perception outputs, thereby verifying whether YolPnP-FT provides sufficiently accurate and stable state estimates to support closed-loop multi-robot coordination.

4.2.1. Multi-Robot Formation Validation

This experiment evaluates whether the ID stability and 6D pose accuracy delivered by YolPnP-FT are sufficient to support formation control in simulation. We deploy five heterogeneous robots in the CoppeliaSim environment and apply a standard leader–follower algorithm. Followers dynamically adjust their relative errors with respect to the leader to maintain formation.

As shown in Figure 17a, the formation is successfully maintained: followers adjust their velocities to preserve inter-robot spacing. The velocity profile in Figure 17b exhibits a change at 30 s, which corresponds to the leader executing a new trajectory command (e.g., a turn or direction change)—not a perception anomaly. These results confirm that the stable IDs and precise 6D poses provided by YolPnP-FT effectively support the development and validation of formation control algorithms.

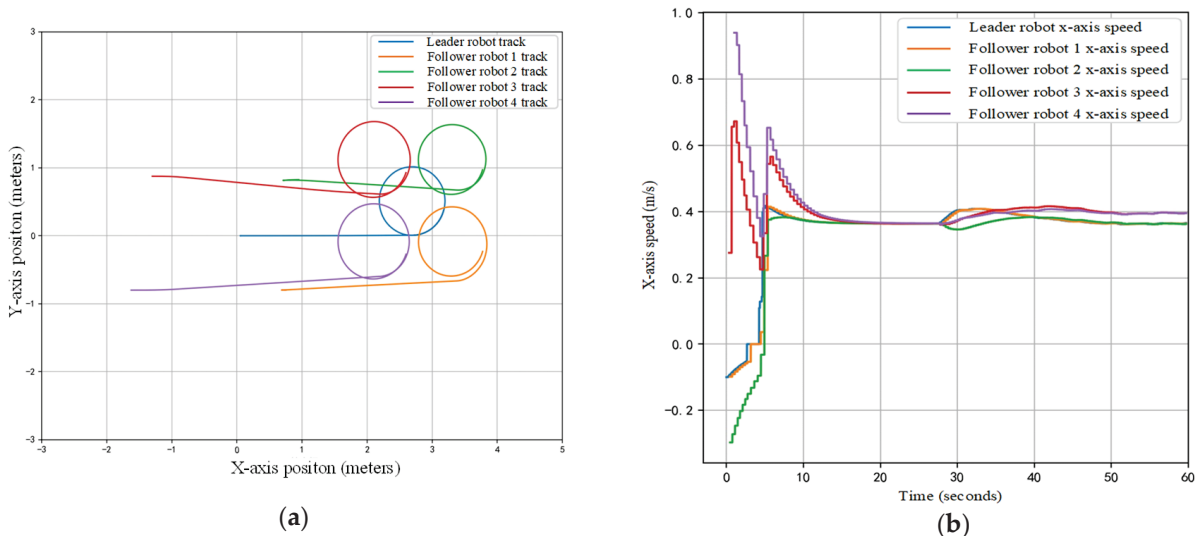


Figure 17. (**a**) Trajectory Plots for Multi-Robots Based on Leader–Follower Algorithm. (**b**) X-Axis Velocity Plot for Multi-Robots Based on Leader–Follower Algorithm.

4.2.2. Path Planning Validation for Multi-Robot Formations

This experiment assesses whether the 6D pose accuracy from YolPnP-FT is adequate for path planning in simulation. We employ a standard RRT algorithm to evaluate navigation performance.

Figure 18a demonstrates successful navigation in an obstacle-free environment. Figure 18b further validates the system's capability to avoid static obstacles. The results indicate that the 6D pose information from YolPnP-FT effectively enables path planning for multi-robot formations. Notably, dynamic obstacles are not considered in the current simulation—a limitation to be addressed in future work.

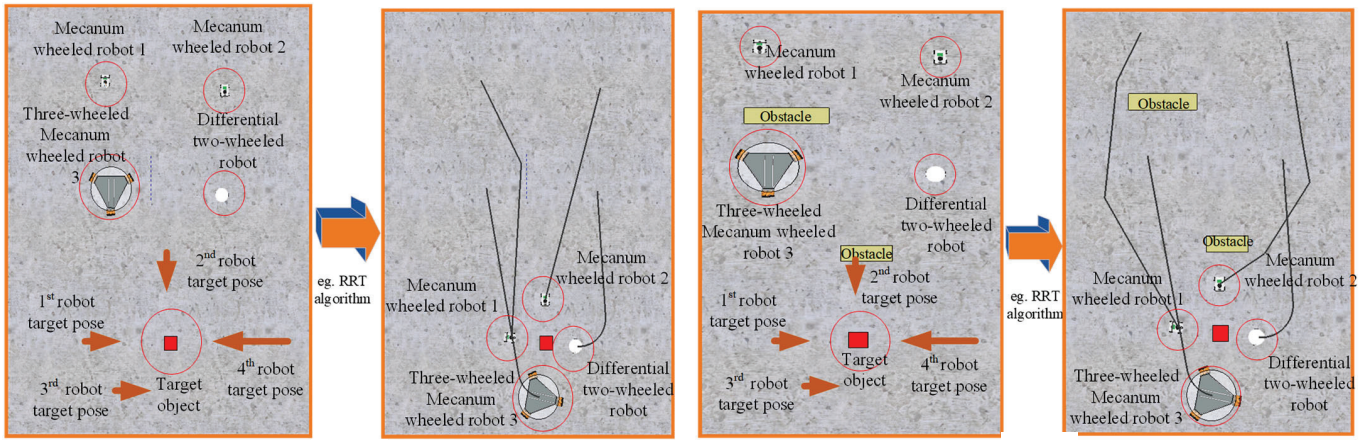


Figure 18. Path Planning for Multi Heterogeneous Robots Based on RRT Algorithm. (**a**) Obstacle-Free Environment; (**b**) Environment with Obstacles.

4.2.3. Collaborative Task Validation for Multi-Robot Formations

This experiment verifies the utility of YolPnP-FT outputs in complex collaborative tasks within simulation. As illustrated in Figure 19, we construct a simple room environment with an overhead RGB camera. The formation consists of one omnidirectional three-wheeled Mecanum robot (leader), two differential-drive robots, and two standard Mecanum-wheeled robots (followers). The objective is to navigate the formation to a target endpoint using RRT, after which followers autonomously plan paths around the target and orient toward it.

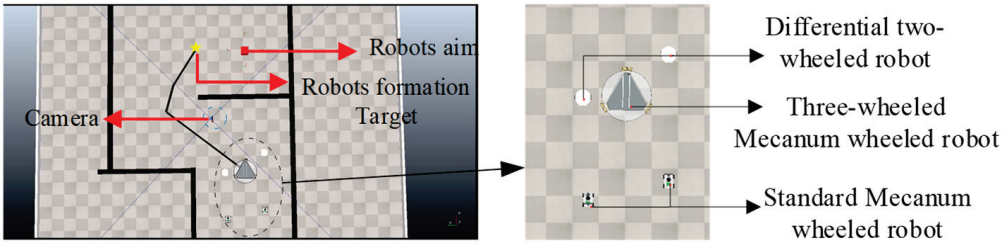


Figure 19. Schematic Diagram for Setting Up a Simulation Scenario of Multi-Robot Formation.

During initial simulation runs, the formation collided with room boundaries. Analysis revealed that the RRT planner considered only individual robot poses and ignored the spatial occupancy of the entire formation. Crucially, YolPnP-FT's complete state output (6D pose + ID for all robots) enabled this diagnostic insight. By redefining the planning reference point as the center of the minimum enclosing circle of the formation and incorporating formation footprint constraints, collisions were eliminated, and the task was successfully completed (Figure 20).

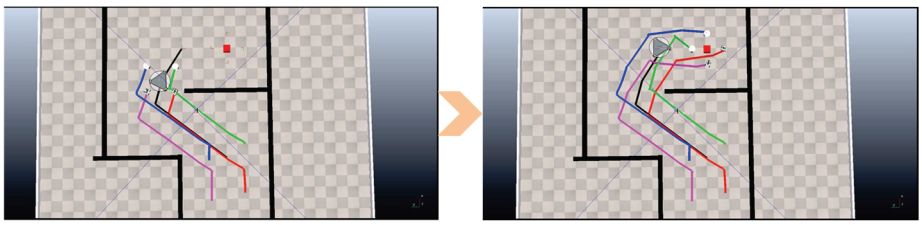


Figure 20. Simulation Verification Process for Multi-Robot Formation Comprehensive Experiment.

This demonstrates that YolPnP-FT not only provides per-robot states but also supports formation-level algorithm debugging and optimization, significantly reducing trial-and-error costs in real-world deployment.

Limitations and Future Work

The current validation assumes static obstacles and a rigid formation model, without considering robot dynamics (e.g., acceleration limits) or interactions with moving obstacles. Additionally, scalability to larger robot fleets (e.g., >10 agents) has not been evaluated. Future work will (1) integrate dynamic obstacle detection, (2) develop real-time replanning frameworks driven by YolPnP-FT outputs, and (3) systematically analyze system throughput and tracking stability as the number of robots increases.

5. Conclusions

This study presents a low-cost, real-time 6D pose estimation and tracking method based on YolPnP-FT perception pipeline for motion capture in multi-robot systems. Using only an Intel RealSense D435i depth camera, several Mecanum-wheeled robots, and a standard host computer, the system achieves simultaneous robot classification, 6D pose estimation, and multi-object tracking in real-world environments. By introducing a keypoint confidence filtering strategy (PnP-FT) and an adapted DeepSORT tracking with cascaded matching mechanism that combines geometric and appearance cues, it effectively mitigates ID switches caused by visual similarity and occlusion. Experimental results demonstrate that, at the recommended camera height of 1.5 m, the system achieves sub-centimeter positional accuracy (average error < 0.009 m) and angular precision (<4.2°), while maintaining robust identity consistency at 19.8 fps with 1080p resolution. Furthermore, the perception outputs are validated in a CoppeliaSim-based simulation environment, confirming their utility for downstream coordination tasks. In summary, this work proposes a low-cost, real-time perception system that provides a practical foundation for affordable and deployable multi-robot algorithms, particularly in visually homogeneous and resource-constrained scenarios.

Author Contributions: Conceptualization, B.S. and Y.H.; Methodology, B.S.; Software, B.S. and R.Z.; Validation, B.S. and R.Z.; Formal analysis, B.S. and R.Z.; Investigation, B.S. and R.Z.; Resources, B.S. and D.Z.; Data curation, B.S., D.Z. and Y.H.; Writing—original draft, B.S.; Writing—review & editing, B.S., D.Z. and Y.H.; Visualization, B.S. and D.Z.; Supervision, D.Z. and Y.H.; Project administration, D.Z.; Funding acquisition, D.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Dreams Foundation of Jianghuai Advance Technology Center (No.2023-ZM01Z016), in part by National Natural Science Foundation of China (No.92248304), in part by the Key research and development projects of artificial intelligence in Liaoning Province (No.2023JH26/10200018), in part by Basic scientific research project of Liaoning Provincial Department of Education (No.LJ212410142073).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy or ethical restrictions.

Conflicts of Interest: The authors declare that there are no conflicts of interest.

References

1. Yu, Y.G.; Miao, Z.Q.; Wang, X.K.; Shen, L.C. Cooperative circumnavigation control of multiple unicycle-type robots with non-identical input constraints. *IET Control. Theory A* **2022**, *16*, 889–901. [CrossRef]
2. Huang, G. Visual-inertial navigation: A concise review. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; IEEE: New York, NY, USA, 2019; pp. 9572–9582.
3. Wu, L.; Guo, S.L.; Han, L.; Baris, C.A. Indoor positioning method for pedestrian dead reckoning based on multi-source sensors. *Measurement* **2024**, *229*, 114416. [CrossRef]
4. Herath, S.; Yan, H.; Furukawa, Y. Ronin: Robust neural inertial navigation in the wild: Benchmark, evaluations, & new methods. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; IEEE: New York, NY, USA, 2020; pp. 3146–3152.
5. Yu, D.; Li, C.G. An Accurate WiFi Indoor Positioning Algorithm for Complex Pedestrian Environments. *IEEE Sens. J.* **2021**, *21*, 24440–24452. [CrossRef]
6. Cui, Y.E.; Chen, X.Y.L.; Zhang, Y.L.; Dong, J.H.; Wu, Q.X.; Zhu, F. BoW3D: Bag of Words for Real-Time Loop Closing in 3D LiDAR SLAM. *IEEE Robot. Autom. Lett.* **2023**, *8*, 2828–2835. [CrossRef]
7. Zhang, T.; Zhao, D.; Yang, J.; Wang, S.; Liu, H. A smart home based on multi-heterogeneous robots and sensor networks for elderly care. In *International Conference on Intelligent Robotics and Applications, Harbin, China, 1–3 August 2022*; Springer: Cham, Switzerland, 2022; pp. 98–104.
8. Xia, X.; Bhatt, N.P.; Khajepour, A.; Hashemi, E. Integrated Inertial-LiDAR-Based Map Matching Localization for Varying Environments. *IEEE Trans. Intell. Veh.* **2023**, *8*, 4307–4318. [CrossRef]
9. Wang, G.; Manhardt, F.; Tombari, F.; Ji, X. Gdr-net: Geometry-guided direct regression network for monocular 6d object pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 16611–16621.
10. Wang, C.; Xu, D.; Zhu, Y.; Martín-Martín, R.; Lu, C.; Fei-Fei, L.; Savarese, S. Densefusion: 6d object pose estimation by iterative dense fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3343–3352.
11. Zhang, Y.F.; Wang, C.Y.; Wang, X.G.; Zeng, W.J.; Liu, W.Y. FairMOT: On the Fairness of Detection and Re-identification in Multiple Object Tracking. *Int. J. Comput. Vis.* **2021**, *129*, 3069–3087. [CrossRef]
12. Wojke, N.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; IEEE: New York, NY, USA, 2018; pp. 3645–3649.
13. Altillawi, M.; Li, S.L.; Prakhya, S.M.; Liu, Z.Y.; Serrat, J. Implicit Learning of Scene Geometry From Poses for Global Localization. *IEEE Robot. Autom. Lett.* **2024**, *9*, 955–962. [CrossRef]
14. Nguyen, A.; Do, T.-T.; Caldwell, D.G.; Tsagarakis, N.G. Real-time 6DOF pose relocalization for event cameras with stacked spatial LSTM networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–20 June 2019.
15. Hoque, S.; Xu, S.X.; Maiti, A.; Wei, Y.C.; Arafat, M.Y. Deep learning for 6D pose estimation of objects-A case study for autonomous driving. *Expert Syst. Appl.* **2023**, *223*, 119838. [CrossRef]
16. Tekin, B.; Sinha, S.N.; Fua, P. Real-time seamless single shot 6d object pose prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–31 June 2018; pp. 292–301.
17. Su, Y.; Saleh, M.; Fetzer, T.; Rambach, J.; Navab, N.; Busam, B.; Stricker, D.; Tombari, F. Zebrapose: Coarse to fine surface encoding for 6dof object pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 6738–6748.
18. Bukschat, Y.; Vetter, M. EfficientPose: An efficient, accurate and scalable end-to-end 6D multi object pose estimation approach. *arXiv* **2020**, arXiv:2011.04307.
19. Zakharov, S.; Shugurov, I.; Ilic, S. Dpod: 6d pose object detector and refiner. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1941–1950.
20. Xu, Y.; Lin, K.-Y.; Zhang, G.; Wang, X.; Li, H. Rnnpose: Recurrent 6-dof object pose refinement with robust correspondence field estimation and pose optimization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–22 June 2022; pp. 14880–14890.
21. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*; Curran Associates Inc.: New York, NY, USA, 2017; Volume 30.
22. Zhao, D.H.; Yang, C.H.; Zhang, T.Q.; Yang, J.Y.; Hiroshi, Y. A Task Allocation Approach of Multi-Heterogeneous Robot System for Elderly Care. *Machines* **2022**, *10*, 622. [CrossRef]
23. Xin, P.; Dames, P. Comparing stochastic optimization methods for multi-robot, multi-target tracking. In *International Symposium on Distributed Autonomous Robotic Systems*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 378–393.

24. Wang, Z.; Zheng, L.; Liu, Y.; Li, Y.; Wang, S. Towards real-time multi-object tracking. In *European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 107–122.
25. Zhou, X.; Koltun, V.; Krähenbühl, P. Tracking objects as points. In *European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 474–490.
26. Peng, J.; Wang, C.; Wan, F.; Wu, Y.; Wang, Y.; Tai, Y.; Wang, C.; Li, J.; Huang, F.; Fu, Y. Chained-tracker: Chaining paired attentive regression results for end-to-end joint multiple-object detection and tracking. In *European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 145–161.
27. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple online and realtime tracking. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; IEEE: New York, NY, USA, 2016; pp. 3464–3468. [CrossRef]
28. Jawad Alzubairi, S.M.; Petunin, A.; Humaidi, A.J. Multi-robot task allocation based on an automatic clustering strategy employing an enhanced dynamic distributed PSO. *Int. Rev. Appl. Sci. Engineering* **2025**, *16*, 347–359. [CrossRef]
29. Nie, Z.; Zhang, Q.; Wang, X.; Wang, F.; Hu, T. Triangular lattice formation in robot swarms with minimal local sensing. *IET Cyber-Syst. Robot.* **2023**, *5*, e12087. [CrossRef]
30. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS--improving object detection with one line of code. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5561–5569.
31. Jocher, G.; Qiu, J.; Chaurasia, A. Ultralytics YOLO, version 8.0.0. (2023). Available online: https://github.com/ultralytics/ultralytics (accessed on 16 November 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

sensors

MDPI

Article

A Universal Tool Interaction Force Estimation Approach for Robotic Tool Manipulation

Diyun Wen, Jiangtao Xiao, Yu Xie *, Tao Luo, Jinhui Zhang and Wei Zhou

Pen-Tung Sah Institute of Micro-Nano Science and Technology, Xiamen University, Xiamen 361005, China; wendiyun@stu.xmu.edu.cn (D.W.); 19920211151542@stu.xmu.edu.cn (J.X.); luotao@xmu.edu.cn (T.L.); jinhuizhang@xmu.edu.cn (J.Z.); weizhou@xmu.edu.cn (W.Z.)
* Correspondence: xieyu@xmu.edu.cn

Abstract

The six-degree-of-freedom (6-DoF) interaction forces/torque of the tool-end play an important role in the robotic tool manipulation using a gripper, which are usually indirectly measured by a robot wrist force/torque sensor. However, the real-time decoupling of the tool's inertial force remains a challenge when different tools and grasping postures are involved. This paper presents a universal tool-end interaction forces estimation approach, which is capable of handling diverse grippers and tools. Firstly, to address uncertainties from varying tools and grasping postures, an online-identifiable tool dynamics model was built based on the Newton–Euler approach for the integrated gripper–tool system. Sensor zero-drift caused by factors such as the tool weight and prolonged operation is incorporated into the dynamic model and identified online in real time, enabling a coarse estimation of the interaction forces. Secondly, a spiking neural network (SNN) is specially employed to compensate for uncertainties caused by the wrist sensor creep effect, since its temporal processing and event-driven characteristics match the time-varying creep effects introduced by tool changes. The proposed method is experimentally validated on a robotic arm with a gripper, and the results show that the root mean square errors of the estimated tool-end interaction forces are below 0.5 N with x, y, and z axes and 0.03 Nm with τ_x, τ_y, and τ_z axes, which has a comparable precision with the in situ measurement of the interaction forces at the tool-end. The proposed method is further applied to robotic scraper manipulation with impedance control, achieving the interaction forces feedback during compliant operation precisely and rapidly.

Keywords: interaction forces estimation; tool dynamics identification; wrist force/torque sensor; creep compensation; sensor zero-drift

1. Introduction

Robotic tool manipulation capability is increasingly demanded in various application domains such as home assistance, industrial manufacturing, and minimally invasive surgery [1,2]. Particularly in robotic-assisted minimally invasive surgery, force-feedback-based manipulation control can prevent unintended tissue damage [3]. Dexterous tool manipulation requires the robot to flexibly grasp, use, and switch between different tools in-hand, similar to the way humans use pliers, screwdrivers, surgical scalpel, or scrapers [4,5].

The accurate measurement of interaction forces at the tool end is a critical prerequisite for robotic tool manipulation control [6]. Due to size constraints in practical applications,

a common approach is to use a wrist-mounted six-axis force/torque sensor to indirectly measure external forces acting on the robot [7–9]. However, because of its mounting position, the sensor cannot directly measure the interaction forces at the tool end. Instead, it provides a coupled output that includes the inertial forces of both the gripper and the tool [10,11]. Therefore, it is essential to rapidly identify the inertial parameters (such as the mass, center of mass, and inertia tensor) of the tool and thereby accurately decouple the tool-end interaction force from the wrist force/torque sensor when the robot changes tools and their in-hand grasping postures.

Newton–Euler-based dynamic modeling combined with least squares parameter identification is widely applied to identify the tool inertial parameters and decouple the gravitational force, allowing for the accurate estimation of the net external force [12–14]. Enhancing the accuracy of tool inertial parameter identification has been the focus of extensive research efforts. Xie et al. [15] accounted for the effects of the sensor weight and the robot installation orientation in parameter identification and employed the least-squares method to estimate these parameters, thereby enhancing the identification accuracy. Zhang et al. [16] compensated for the sensor mounting offsets and installation angle deviations by leveraging the force sensor data from multiple robotic poses to identify the parameters, thereby enhancing the accuracy of static external force estimation. Farsoni et al. [17] incorporated physical consistency constraints into parameter identification to enhance the reliability of the identification parameters, employing nonlinear optimization to ensure compliance with rigid-body dynamics. Nadeau et al. [18] proposed a tool-mass discretization method to enable tool identification under trajectory constraints. This approach incorporates tool shape information to assist in parameter identification, thereby simplifying the model. Duan et al. [19] employed a self-developed wrist six-axis force/torque sensor with embedded linear and angular acceleration measurements, together with known payload mass and inertia, to compute and compensate inertial forces online. These studies demonstrate the effectiveness of Newton–Euler formulations in improving identification accuracy under static or structured trajectories. Beyond Newton–Euler-based dynamic modeling, several studies have also explored model-free approaches for estimating tool inertial forces during operation. Su et al. [20] used an artificial neural network (ANN) to identify the tool gravity by regressing the tool orientation against the sensor-projected force, then applied gravity decoupling and calibration to obtain the interaction force. Furthermore, Su in [21] proposed a model-free deep convolutional neural network (CNN) to identify the tool's dynamic model and compensate the sensor measurements.

Another concern is that sensor-intrinsic errors—such as zero-drift, creep, and other uncertainties—can adversely affect accuracy during tool parameter identification and tool-end interaction force estimation [22–25]. Xue et al. [26] calculated and compensated the sensor's zero-drift and gravitational components based on measurements under no external contact conditions. Yu et al. [27] utilized multi-pose sensor data and applied least-squares estimation combined with singular value decomposition (SVD) to identify the sensor's zero-drift under constrained working conditions. Shin et al. [28] addressed sensor crosstalk (channel coupling in a six-axis force/torque sensor) and used gradient-based nonlinear optimization to simultaneously identify the sensor bias, tool gravity (center-of-mass torque), crosstalk parameters, and base inclination from a limited set of poses. Although these model-based methods can effectively remove steady-state biases, they are less capable of representing dynamic drift and coupling effects that evolve over time. Therefore, learning-based methods have been introduced as a complementary strategy, capable of capturing complex temporal correlations and achieving a higher precision in sensor error compensation. Yao et al. [29] applied a deep neural network (DNN) to predict the sensor zero-point—encompassing zero-drift, systematic errors, and self-weight effects—and com-

pensated accordingly, reducing the tool-end interaction force error to below 1.5 N and the torque error to below 4.3 Nm. Liu et al. [30] proposed a two-stage identification framework for high-load, large-component collaboration scenarios, jointly identifying the static errors (base tilt and sensor zero-drift) and dynamic inertial parameters, and achieving accurate collision force perception via secondary zero-drift compensation during low-speed operation. Wu et al. [31] mitigated the force-signal creep induced by temperature drift by developing a time–temperature joint compensation model, with model parameters identified via particle swarm optimization (PSO), thereby enabling real-time temperature-drift compensation. Sun et al. [32] modeled the sensor output as a linear function of temperature, combined least-squares compensation with RBF-NN (Radial Basis Function Neural Network) nonlinear fitting, and employed a PSO-optimized LSSVM (Least Square Support vector machine) to achieve high-precision real-time temperature compensation with markedly reduced residuals. Gao et al. [33] addressed sensor creep by combining historical loading data with state classification and using an LSTM (Long Short-Term Memory) network for real-time creep prediction and compensation, significantly enhancing the measurement accuracy and stability.

In previous research, the tool was typically fixed at the end of the robotic arm and maintaining a constant relative pose. However, changing the tool types and grasping postures can induce varying levels of zero-drift and creep effects in wrist force/torque sensors. Previous methods did not account for these factors and were not validated under such conditions.

In this work, in addition to tool inertial force identification, we implemented a two-stage strategy for wrist sensor compensation. First, as the zero-drift errors caused by tool gravity and grasping variations change with the tool, we treated them as tool-induced characteristics and identified them during tool parameter identification. For minor uncertain errors with temporal characteristics, such as creep and temperature drift, we employed a Spiking Neural Network (SNN) for compensation, leveraging its advantages in processing time-series signals and event-driven capabilities. Noting that creep error is also tool-related, we created a training dataset incorporating payload information, enabling it to accurately identify creep errors that exhibit both temporal effects and tool-specific characteristics.

The contributions of this paper are as follows:

1 A high-precision and rapid tool-end force estimation approach is proposed for robotic tool manipulation, especially when the robot changes tools or their grasping postures during a manipulation task.
2 A two-stage wrist sensor compensation strategy is presented to achieve high accuracy in tool-end force estimation with a root-mean-square error below 0.5 N, through zero-drift error identification and creep error SNN-based compensation.
3 The proposed tool-end interaction force estimation method is implemented in a robotic manipulation controller and verified to be precise and rapid in a robotic adhesive scraping manipulation task.

The remainder of this paper is as follows: Section 2 provides a detailed description of the proposed method for identifying tool inertial parameters and presents the SNN-based compensation of sensor uncertainty errors. Section 3 shows the experiments and verifies the proposed method. In Section 4, we have conducted an analysis and discussion. Finally, Section 5 concludes this paper and suggests the directions for future work.

2. Methodology

2.1. Force Decoupling

In robotic tool manipulation, the wrist force/torque sensor is installed between the robot's end effector and gripper, show in Figure 1. The gripper grip different tools

to perform manipulation tasks. The output of the wrist force/torque sensor includes the following:

$$F_s = {}^sF_g + F_0 + F_u + F_e, \tag{1}$$

where $F_s = [f_{sx}, f_{sy}, f_{sz}, \tau_{sx}, \tau_{sy}, \tau_{sz}]^T$ is the raw measurement values from the wrist force/torque sensor, ${}^sF_g = [f_{gx}, f_{gy}, f_{gz}, \tau_{gx}, \tau_{gy}, \tau_{gz}]$ is the inertial force and torque caused by the gripper-tool, $F_0 = [f_{0x}, f_{0y}, f_{0z}, \tau_{0x}, \tau_{0y}, \tau_{0z}]^T$ is the zero-drift of the sensor, $F_u = [f_{ux}, f_{uy}, f_{uz}, \tau_{ux}, \tau_{uy}, \tau_{uz}]^T$ is the uncertainties caused by wrist sensor creep effect, and $F_e = [f_{ex}, f_{ey}, f_{ez}, \tau_{ex}, \tau_{ey}, \tau_{ez}]^T$ is the tool-end interaction force between the tool and the environment during manipulation.

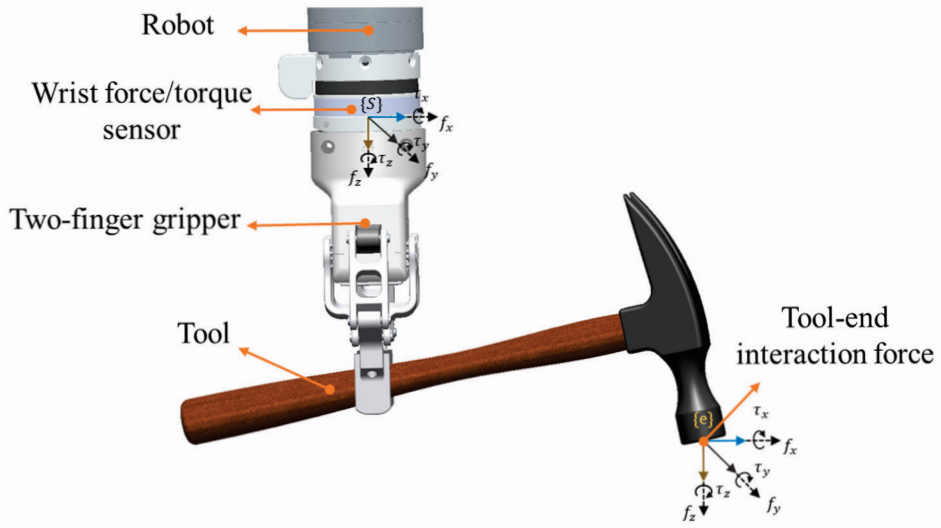


Figure 1. The installation of the wrist force/torque sensor and tool-end interaction forces in robotic tool manipulation.

2.2. Gripper–Tool Dynamic Parameter Identification

Based on Newton–Euler equations, the dynamic model of the gripper–tool assembly is established with respect to the robot end-effector coordinate frame $\{S\}$.

$$\begin{cases} {}^sf_g = m^sa - m^sg + {}^s\alpha \times m^sc + {}^s\omega \times ({}^s\omega \times m^sc) \\ {}^s\tau_g = {}^sI^s\alpha + {}^s\omega \times ({}^sI^s\omega) + m^sc \times {}^sa - m^sc \times {}^sa \end{cases} \tag{2}$$

where ${}^sf_g$ and ${}^s\tau_g$ are the inertial force and torque caused by the gripper–tool on the sensor, m denotes the mass of gripper–tool, ${}^sa$ denotes the linear acceleration vector, ${}^s\alpha$ denotes the angular acceleration vector, ${}^s\omega$ represents the angular velocity vector, ${}^sg$ represents the gravity vector, ${}^sc$ denotes the position of the gripper–tool's center of mass, and ${}^sI$ represents the inertia tensor. All these parameters are described relative to the sensor coordinate system, denoted as $\{S\}$, specifically,

$${}^sI = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{xy} & I_{yy} & I_{yz} \\ I_{xz} & I_{yz} & I_{zz} \end{bmatrix} \tag{3}$$

Given with Equation (2), the observation matrix A is then separated as follows:

$${}^sF_g = \begin{bmatrix} f_g \\ \tau_g \end{bmatrix} = \mathrm{A}({}^sa, {}^s\alpha, {}^s\omega, {}^sg)^s\varphi, \tag{4}$$

where $\mathrm{A}({}^sa, {}^s\alpha, {}^s\omega, {}^sg) \in R^{6\times10}$ and the expressions of $\mathrm{A}({}^sa, {}^s\alpha, {}^s\omega, {}^sg)$ is

$$
\mathrm{A}({}^{s}a, {}^{s}\alpha, {}^{s}\omega, {}^{s}g) = \begin{bmatrix}
a_x - g_x & -\omega_y^2 - \omega_z^2 & \omega_x\omega_y - \alpha_z & \omega_x\omega_z + \alpha_y & 0 & 0 & 0 & 0 & 0 & 0 \\
a_y - g_y & \omega_x\omega_y + \alpha_z & -\omega_x^2 - \omega_z^2 & \omega_y\omega_z - \alpha_x & 0 & 0 & 0 & 0 & 0 & 0 \\
a_z - g_z & \omega_x\omega_z - \alpha_y & \omega_y\omega_z + \alpha_x & -\omega_y^2 - \omega_x^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & a_z - g_z & -a_y + g_y & \alpha_x & -\omega_x\omega_z + \alpha_y & \omega_x\omega_y + \alpha_z & -\omega_y\omega_z & \omega_y^2 - \omega_z^2 & \omega_y\omega_z \\
0 & -a_z + g_z & 0 & a_x - g_x & \omega_x\omega_z & \omega_y\omega_z + \alpha_x & \omega_z^2 - \omega_x^2 & \alpha_y & -\omega_x\omega_y + \alpha_z & -\omega_x\omega_z \\
0 & a_y - g_y & -a_x + g_x & 0 & -\omega_x\omega_y & \omega_x^2 - \omega_y^2 & -\omega_y\omega_z + \alpha_x & \omega_x\omega_y & \omega_x\omega_z + \alpha_y & \alpha_z
\end{bmatrix}, \tag{5}
$$

${}^{s}\varphi$ is the parameters' vector to be identified,

$$
{}^{s}\varphi = \left[m, m^{s}c_x, m^{s}c_y, m^{s}c_z, I_{xx}, I_{xy}, I_{xz}, I_{yy}, I_{yz}, I_{zz}\right]^{T}. \tag{6}
$$

Inspired by reference [33], we recognized that varying tool qualities and distinct force application modes can induce differing initial zero-drift effects in sensors. Consequently, in this section, the zero-drift of the sensor F_0 is treated as an additional parameter and then added to the inertial parameters of the tool for unified identification. The expanded equation is now represented as follows:

$$
F = A_{ext}{}^{s}\varphi_{ext}, \tag{7}
$$

where $F = F_0 + F_g$ is the non-interactive force portion of the sensor measurement, $A_{ext} = [E_{6\times6}A]$ denotes extended regression matrix, and ${}^{s}\varphi_{ext} = \left[f_{0x}, f_{0y}, f_{0z}, \tau_{0x}, \tau_{0y}, \tau_{0z}, {}^{s}\varphi\right]^{T}$ is the expanded parameter vector.

The least-squares method (LSM) is characterized by high computational efficiency, accuracy, and real-time capability, making it suitable for real-time tool dynamic parameter identification. Therefore, the identified parameters are as follows:

$$
{}^{s}\varphi_{ext} = \left(A_{ext}{}^{T}A_{ext}\right)^{-1}A_{ext}{}^{T}F. \tag{8}
$$

To optimize the tool's dynamic behavior and minimize identification errors, the formulation of an appropriate excitation trajectory is crucial. Trajectories generated using Fourier series and polynomials are periodic, ensuring smooth initiation and termination, minimal jitter, and reduced data noise [34]. The trajectory expression is articulated as follows:

$$
\theta_i(t) = \sum_{k=1}^{N}\left[\frac{a_{i,k}}{\omega_f k}\sin\left(\omega_f k t\right) - \frac{b_{i,k}}{\omega_f k}\cos\left(\omega_f k t\right)\right] + \sum_{i=0}^{5} c_{i,m}\left(t - \left\lfloor\frac{t}{T}\right\rfloor T\right)^{m}, \tag{9}
$$

where $\theta_i(t)$ represents the angle of joint i of the robotic arm, N is the number of harmonics in the Fourier series trajectory, k is the harmonic index, ω_f is the fundamental frequency, $a_{i,k}$ and $b_{i,k}$ are the coefficients of the sine and cosine components of the Fourier series, $c_{i,m}$ represents the coefficient of the m-th term in the fifth-degree polynomial, t denotes the trajectory running time, $T = 2\pi/\omega_f$ denotes trajectory period, and $\lfloor\ \rfloor$ denotes the floor function. The parameters $a_{i,k}$, $b_{i,k}$, and $c_{i,k}$ are the optimization variables.

To minimize the impact of data errors on identification accuracy, it is essential to reduce the condition number $k(\tilde{A})$ of the regression matrix and optimize the coefficients in the excitation trajectory within the specified constraints. The constraints are as follows:

$$\begin{cases} \min(k(\tilde{A})) \\ \theta_{i,min} \le |\theta_i(t)| \le \theta_{i,max} \\ \left|\dot{\theta}_i(t)\right| \le \dot{\theta}_{i,max} \\ \left|\ddot{\theta}_i(t)\right| \le \ddot{\theta}_{i,max} \\ \theta_i(t_0) = \theta_i\left(t_f\right) \\ \dot{\theta}_i(t_0) = \dot{\theta}_i\left(t_f\right) = 0 \\ \ddot{\theta}_i(t_0) = \ddot{\theta}_i\left(t_f\right) = 0 \\ w(\theta(t)) \subset W_0 \end{cases}, \tag{10}$$

where $\theta_{i,min}$ and $\theta_{i,max}$ are the angle limits of joint i, where $\dot{\theta}_{i,max}$ and $\ddot{\theta}_{i,max}$ represent the preset maximum joint angular velocity and angular acceleration, respectively; t_0 and t_f are the initial and ending moments of the excitation trajectory, respectively; $w(\theta(t))$ represents the Cartesian space position of the robotic arm; and W_0 represents the movable space of the robotic arm.

The optimal excitation trajectory is then obtained by iterative optimization of the trajectory using MATLAB Optimization Toolbox (MATLAB R2023a, Optimization Toolbox Version 9.4), where $\omega_f = 0.2\pi$ Hz, $N = 5$, and i takes value from 3 to 6 to decrease the required working space and ensure the condition number meets identification requirements.

Control the robotic arm to follow the excitation trajectory without contact, that is, $F_e = 0$. In this scenario, the measured values from the sensor are denoted as $F_s = F + F_u$, and n sets of sensor data and motion data ${}^s a$, ${}^s\alpha$, ${}^s\omega$, and ${}^s g$ are calculated with Jacobi. Substitute them into Equation (8); then, the extended model parameters can be obtained as follows:

$${}^s\varphi_{ext} = (\tilde{A}_{ext}^T \tilde{A}_{ext})^{-1} \tilde{A}_{ext}^T \tilde{F}, \tag{11}$$

where $\tilde{F} = \left[F_1,\ F_2, \ldots, F_n\right]^T$ represents the 6n dimensional column vector composed of n groups of sensor data, and $\tilde{A}_{ext} = \left[A_{ext_1},\ A_{ext_2}, \ldots, A_{ext_n}\right]^T$ is the $6n \times 16$ dimensional observation matrix composed of n groups of motion data.

By substituting the identified parameters into Equation (11) and using Equations (4) and (7) to decouple the sensor measurements, the interactive force at the tool-end of the robot arm leads to

$$F_e + F_u = F_s - F = F_s - \tilde{A}_{ext}{}^s\varphi_{ext} \tag{12}$$

2.3. Sensor Uncertain Error Compensation

After accounting for the inertial forces of the gripper–tool assembly and the sensor's zero-drift error, the readings from the wrist force/torque sensor still include uncertain errors F_u due to sensor creep effects, as expressed in Equations (1) and (12). A novel deep neural network, a spiking neural network (SNN) [35], is employed to identify the uncertain force F_u. In contrast to traditional neural networks such as MLP (Multilayer Perceptron), CNN, and LSTM, SNNs possess inherent temporal dynamics and sparse computation capabilities, allowing them to effectively process time-varying signals such as time-varying sensor creep and temperature drift. Furthermore, the event-driven nature of SNNs contributes to high computational efficiency, and, together with their low power consumption, makes them well-suited for robotic tool manipulation applications. Leveraging these advantages, this study adopts SNNs for error compensation to enhance both the accuracy and responsiveness of tool-end interaction force estimation during dynamic manipulation tasks.

The proposed SNN-based uncertainty errors estimation method is shown in Figure 2. The output of the training model is defined as F_u, which is calculated from the reading of the wrist sensor and the identified force F in Equation (7). The inputs include the 10

dynamic parameters ${}^{s}\varphi$ of the payload and the measured wrist 6-dimensional velocity and acceleration vector, respectively. The optimizer employs the Adam algorithm as implemented in PyTorch (Version 2.2.1), with a learning rate of 0.001. The number of hidden layer neurons is set to 128, and the output layer dimension corresponds to the output parameter dimensions. The membrane potential decay rate β and the threshold V_{thr} are randomly initialized. The iteration count is set to 50,000, along with the optimizer and loss function. The regression model is trained iteratively, where forward propagation computes the error loss and updates the model weights W and bias b; backward propagation adjusts the membrane potential decay rate β and the threshold V_{thr}.

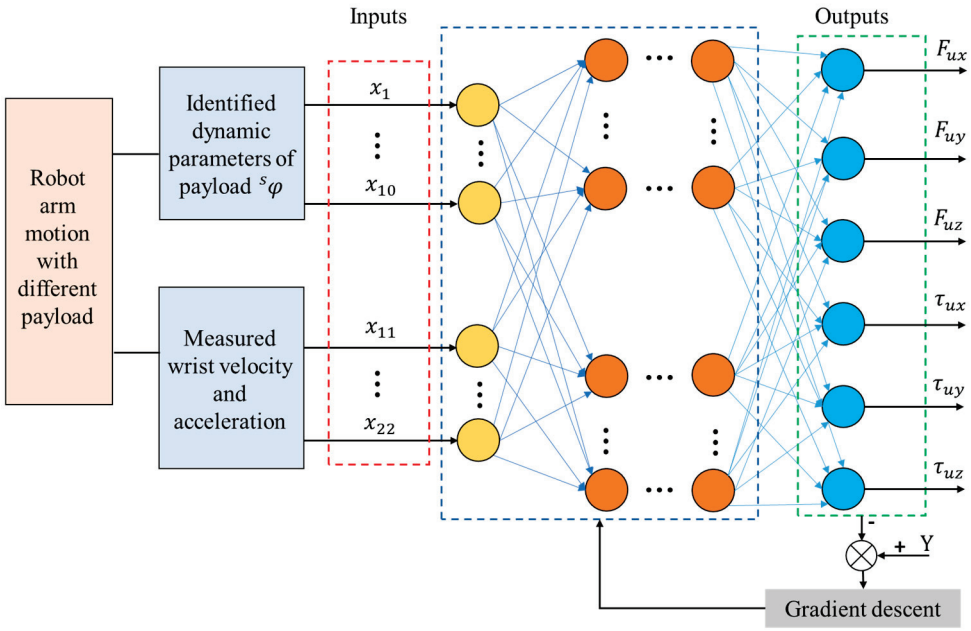


Figure 2. Sensor uncertainty error estimation with Spiking Neural Network.

The loss function for training is set as follows:

$$fit = |F_u - F_{pu}|, \tag{13}$$

where F_{pu} denotes the output of the Spiking Neural Network model, which is the predicted value of the uncertainty error.

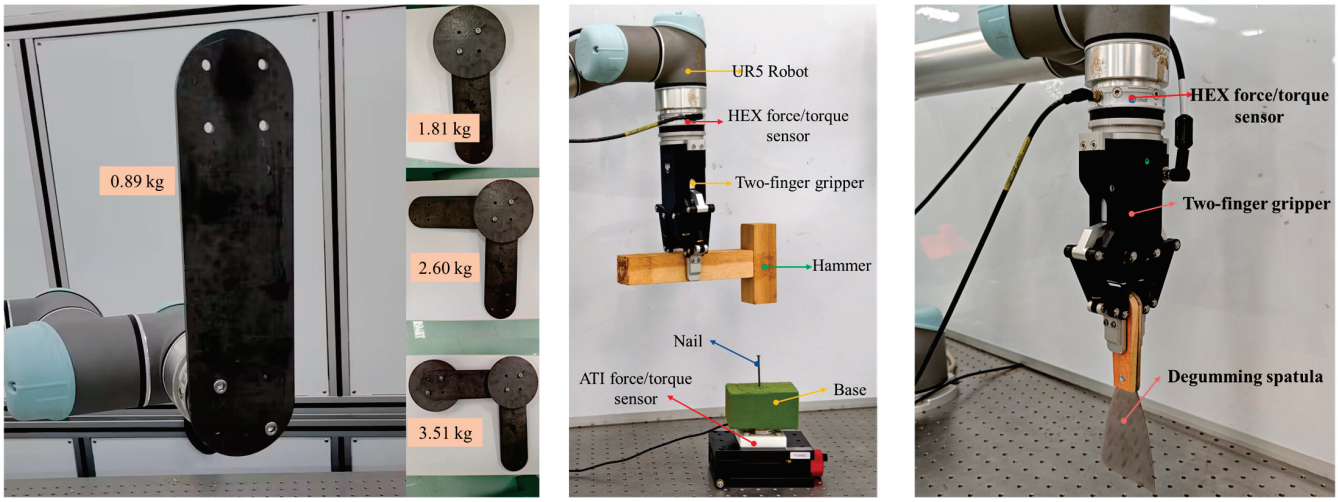
To increase the sample size and fully reflect the creep characteristics of the sensor under different motion states, a 60-second robot arm motion trajectory was commanded as the sample trajectory for data collection. Four sets of payloads (0.89 kg, 1.81 kg, 2.60 kg, and 3.51 kg) were attached to the robot. The wrist velocity, acceleration, and wrist force/torque data were collected at a frequency of 100 Hz, resulting in a total of 240,000 datasets, all of which were filtered. These data were aggregated into a training dataset incorporating temporal information and fed into the SNN for training.

To mitigate its effect on model accuracy, we employ L1 regularization to minimize model complexity and K-fold cross-validation to assess the model's performance in generalizing to noisy data. We use L1 loss (absolute error) and the root mean square (RMS) of the difference as evaluation metrics and calculate the average of the K-fold cross-validations to determine the final evaluation metric. During cross-validation, we select the model parameters that perform best overall on the validation set as the final regression model. This method significantly improves the robustness of the model and effectively avoids overfitting by utilizing all the data.

3. Experimental Verification

3.1. Experimental Setups

To validate the effectiveness of the proposed tool force estimation method, a series of experiments were designed and conducted. The experimental setup, as shown in Figure 3, consists of a robotic arm (UR5, Universal Robots, Odense, Denmark), a six-axis force/torque sensor (HEX-E, OnRobot, Odense, Denmark), and a two-finger gripper (AG95, Dahuan Robotics, Suzhou, China).



(**a**) (**b**) (**c**)

Figure 3. Experimental setups: (**a**) tool dynamic parameter identification and verification, (**b**) tool-end interaction force validation in a robotic nail-hammering task, and (**c**) robotic glue-scooping manipulation in which the estimated tool-end interaction force is used as feedback to the impedance controller.

Figure 3a shows the standard payloads of different masses, which were used for tool inertial parameters identification experiments. Figure 3b shows the robotic nail-hammering experiment design, which was employed to evaluate the effectiveness of the proposed tool-end interaction force estimation method during a tool manipulation task. A high-precision force sensor (Nano43, ATI Industrial Automation, Apex, NC, USA) with a sensing resolution of 0.008 N was used to measure the actual tool-end interaction force, that is, the hammering force applied to the nail. And Figure 3c presents a glue scooping manipulation task, in which the estimated tool-end interaction force serves as feedback to the impedance controller. Each experiment, including the hammering and adhesive-scraping tasks, was repeated at least three times under identical conditions to verify the consistency. The results from one representative trial are presented in the figures for clarity.

3.2. Identification of Tool Inertial Parameters

To verify the effectiveness and accuracy of the proposed tool dynamic parameters identification method, four sets of standard payloads (0.89 kg, 1.81 kg, 2.60 kg, and 3.51 kg) were used by running the excitation trajectory. These parameters are used to calculate the inertial force of the tool. Table 1 displays the identified and theoretical values of the ten dynamic parameters for the four standard loads. The identification errors for the load mass parameters are 6.7%, 2.8%, 2.3%, and 2.0%, respectively. The average identification errors for the three center-of-mass coordinate parameters of the loads are 8.1%, 4.3%, 7.2%, and 5.3%, respectively. It should be noted that these center-of-mass identification errors also include installation and the geometric alignment errors of the standard payloads, meaning that the true identification uncertainty is smaller than the reported theoretical values. The

average time required for parameter identification across four standard payloads was 5.78 s, demonstrating high real-time performance. These results demonstrate that the method employed in this study is time-efficient and highly accurate with minimal error.

Table 1. Standardized payload inertial parameter identification results.

	0.89 kg		**1.81 kg**		**2.60 kg**		**3.51 kg**	
	Identified	**Theoretical**	**Identified**	**Theoretical**	**Identified**	**Theoretical**	**Identified**	**Theoretical**
m (kg)	0.95	0.89	1.86	1.81	2.66	2.60	3.61	3.51
c_x (mm)	−81.8	−89	−127.4	−130	−0.9	0	−132.4	−137
c_y (mm)	−0.35	0	−0.95	0	−153.5	−156	0.66	0
c_z (mm)	2.66	2.9	6.93	6.50	8.63	9.80	16.1	15
I_{xx} (kg·cm^2)	3.34	3.94	−72.25	−523.6	499.80	753.58	−123.3	−965.21
I_{xy} (kg·cm^2)	1.76	0	−130.56	0	19.08	0	−419.01	0
I_{xz} (kg·cm^2)	−2.58	−2.27	−11.47	−38.28	41.91	0	−140.8	−96.46
I_{yy} (kg·cm^2)	144.50	119.16	−517.74	−156.74	26.76	41.73	−882.71	−526.64
I_{yz} (kg·cm^2)	4.10	0	−39.29	0	43.04	−45.41	31.403	0
I_{zz} (kg·cm^2)	154.78	122.91	−486.99	−146.79	73.94	773.39	−804.65	−426.39
Identification time (s)	5.6		5.8		6.1		5.6	

It is observed that some theoretical values are 0 in Table 1, while their identified values are not. This is mainly due to the manufacturing errors of the homemade payloads and alignment error caused by mounting. Moreover, as the values of the inertia parameters of the payload are relatively small and primarily driven by acceleration, and considering the speed requirements for robotic tool operations, the components caused by the inertia parameters are generally minor in Equation (2). Nevertheless, the uncertainty introduced by the identified inertial terms may still propagate into the estimation of the gravitational and inertial force F_g, thereby affecting the overall accuracy of the estimated tool-end interaction force F_e. This effect remains limited under our low-speed experimental conditions but represents an inherent source of uncertainty within the identification process. This uncertainty could be further reduced through the optimization of the excitation trajectory design and the conditioning of the least-squares identification procedure.

3.3. Tool-End Interaction Force Estimation

The purpose of this experiment is to evaluate the accuracy of the proposed tool-end interaction force estimation method in a robotic nail-hammering scenario, as shown in Figure 3b. First, the robot grasps the wooden hammer without contacting the nail, indicating that the tool and its in-hand posture have been confirmed and that the tool-end interaction force F_e is zero at this stage. We then proceed with gripper–tool dynamic parameter identification, inertial force decoupling, and SNN-based uncertainty error compensation. The total duration of the entire recognition and computation process is approximately 6.5 s.

Figure 4 shows the reading of the wrist force/torque sensor and the estimated tool-end interaction force after inertial force decoupling. It can be seen that, after the decoupling of inertial forces, the tool-end interaction force approaches the actual value of zero; yet, a maximum estimation error of approximately 1.8 N still remains. Figure 5 demonstrates the estimated tool-end interaction force with and without uncertainty error compensation, which shows the maximum tool-end interaction force error can be further reduced to 0.8 N after uncertainty error compensation. We compare the average estimation error (RMSE) with the measurement accuracy of the wrist sensor in Table 2. These results partially show that our tool-end force estimation method achieves a precision comparable to the in situ measurement.

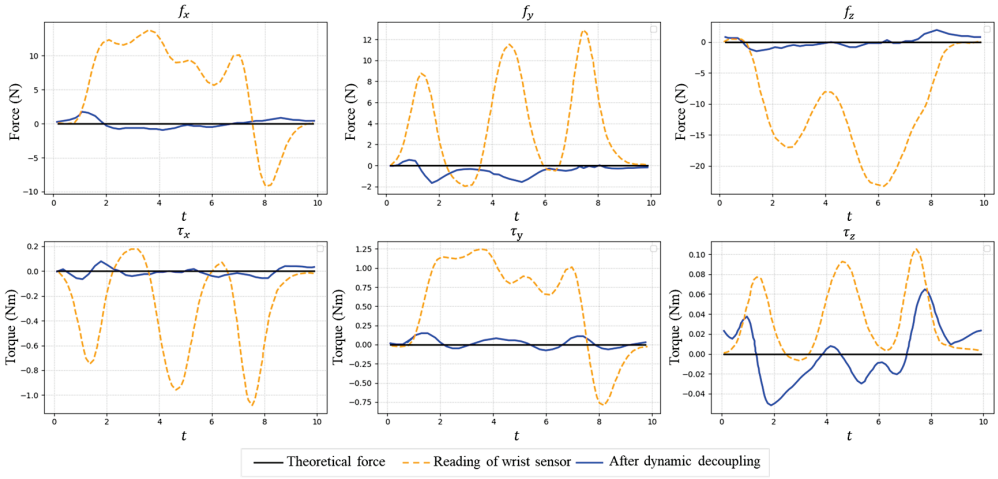


Figure 4. The reading of the wrist force/torque sensor and the estimated tool-end interaction force after inertial force decoupling.

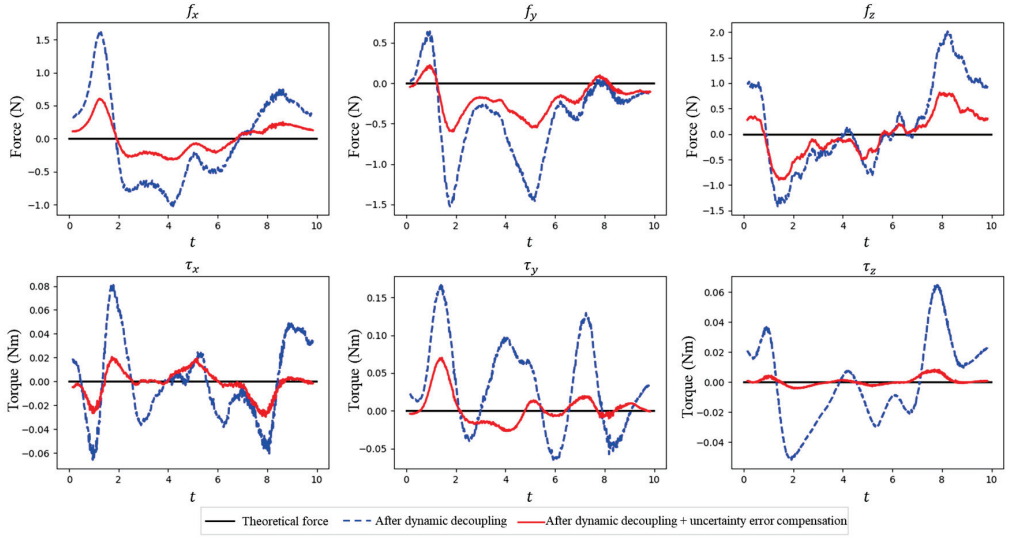


Figure 5. The estimated tool-end interaction force with and without uncertainty error compensation.

Table 2. The RMSEs of the tool-end force after decoupling and after SNN compensation when the robot is in a non-contact state ($F_e = 0$).

	Reading of Sensor	**After Decoupling**	**After SNN Compensation**	**Resolution of Wrist Sensor**
f_x(N)	8.41	0.75	0.27	0.2
f_y(N)	5.70	0.74	0.31	0.2
f_z(N)	12.84	1.02	0.49	0.8
τ_x(N*m)	0.477	0.057	0.023	0.01
τ_y(N*m)	0.791	0.081	0.028	0.01
τ_z(N*m)	0.049	0.031	0.008	0.002

In order to verify the superiority of our method, we present the RMSE comparison of the six-axis force decoupling result between our method and other approaches when the robot is in a non-contact state ($F_e = 0$) in Table 3, where the Double-LSM is the method proposed in the literature [16], which considers the sensor zero-drift and robot installation error. DL-LSM (Deep Learning—Least Squares Method) is the method proposed in the literature [28], and Bias with Non-linear Optimization is the method proposed in the literature [29], which considers not only the bias, center of mass, robot inclination, but

also the crosstalk of the sensor. And the SNN-LSM is our method. In Table 3, our method achieves lower RMSE values on the f_x, f_y, and f_z axes compared with other approaches. The RMSE values for the torque are slightly higher than other methods, particularly Ref. [29]. This is primarily because Ref. [29] estimated the torque using large sample sizes of torque-generating factors, including attitude-related zero points, mounting inclination angles, and the gravitational effects of tool. Therefore, only minor random noise remains in the contactless state, resulting in torque errors appearing very close to zero.

Table 3. Comparison of RMSE of six-axis force/torque error with other studies.

	f_x	f_y	f_z	τ_x	τ_y	τ_z	**Method**
Ours	0.27	0.31	0.49	0.023	0.028	0.008	SNN-LSM
[16]	0.593	0.751	0.822	0.013	0.301	0.003	Double-LSM
[28]	0.693	1.069	1.285	0.095	0.110	0.022	Bias with Non-liner Optimization
[29]	0.377	0.583	0.629	0.0082	0.0138	0.003	DL-LSM

Given that these methods employ sensors with differing precision and measurement ranges, a simple comparison of force RMSE data post-identification may not accurately reflect the methodological superiority. Therefore, we calculated %FS values based on the RMSE values obtained from each method and the sensor parameters for a more reasonable comparison, as shown in Table 4. It is evident that our method exhibits a high accuracy on the force estimation of f_x, f_y, and f_z, while the gap with other methods on the torque is significantly narrowed.

Table 4. Comparison of %FS of 6-axis force/torque error with other studies.

	f_x **(%FS)**	f_y **(%FS)**	f_z **(%FS)**	τ_x **(%FS)**	τ_y **(%FS)**	τ_z **(%FS)**	**Method**	**Sensor Resolution**
Ours	0.135	0.155	0.245	0.230	0.280	0.123	SNN-LSM	HEX-E OnRobot ($f_{x,y}$ 0.2, f_z 0.8, $\tau_{x,y}$ 0.01, τ_z 0.002)
[16]	0.473	0.571	0.328	0.218	0.428	0.049	Double-LSM	Omega190 ($f_{x,y}$ 0.375, f_z 0.628, $\tau_{x,y}$ 0.05, τ_z 0.03)
[28]	0.347	0.535	0.643	0.633	0.733	0.147	Bias with Non-liner Optimization	AFT200 ($f_{x,y,z}$ 0.15, $\tau_{x,y,z}$ 0.015)
[29]	0.377	0.583	0.315	0.164	0.276	0.060	DL-LSM	Self-developed tandem force sensor ($f_{x,y}$ 0.003, f_z 0.006, $\tau_{x,y,z}$ 0.00015)

Considering all six force/torque axes, our method demonstrates a superior overall accuracy in the six-axis force/torque estimation. This improvement results from our two-stage sensor error identification method. The method addresses the errors induced by varying tool types and grasping postures: the stable zero-drift is incorporated into the dynamic model for identification, while time-varying uncertainties from creep and temperature drift are compensated for by the SNN.

We then allow the robot to strike the nail along the z-axis under two different acceleration conditions (50 mm/s^2 and 100 mm/s^2). Figure 6 shows the force trajectories of the estimated hammer–nail interaction force and the actual measured values from the ATI (ATI Industrial Automation, Apex, NC, USA) force sensor.

It is observed that the transient impact force generated upon contact between the hammer and the nail is effectively captured, manifested as a sharp increase in interaction force during the hammering task. For the two accelerations (50 mm/s^2 and 100 mm/s^2), the average errors in f_z over the entire hammering process are only 0.32 N and 0.21 N, respectively. Both values are within the established accuracy envelope (z-axis RMSE = 0.49 N, 0.245%FS), corresponding to 0.160%FS and 0.105%FS on f_z. These results

indicate that the proposed tool-end force estimation maintains its precision under dynamic impact conditions.

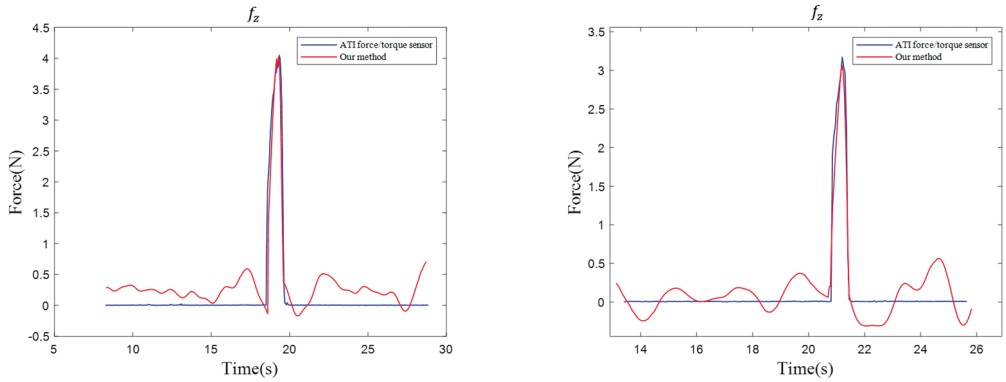


Figure 6. The force trajectories of the estimated hammer–nail interaction force and the actual measured values: (**a**) striking acceleration 50 mm/s^2; and (**b**) striking acceleration 100 mm/s^2.

3.4. Tool-End Interaction Force for Manipulation Controller

The purpose of this experiment is to evaluate the effectiveness and responsiveness of the proposed tool-end interaction force estimation method in a robotic adhesive scraping manipulation task. The 'scrape off the adhesive' task was chosen as the validation scenario because it requires the tool-end interaction force, as shown in Figure 7a, to be quickly and accurately estimated and serves as feedback to the manipulation impedance controller. The impedance control framework is detailed in [36].

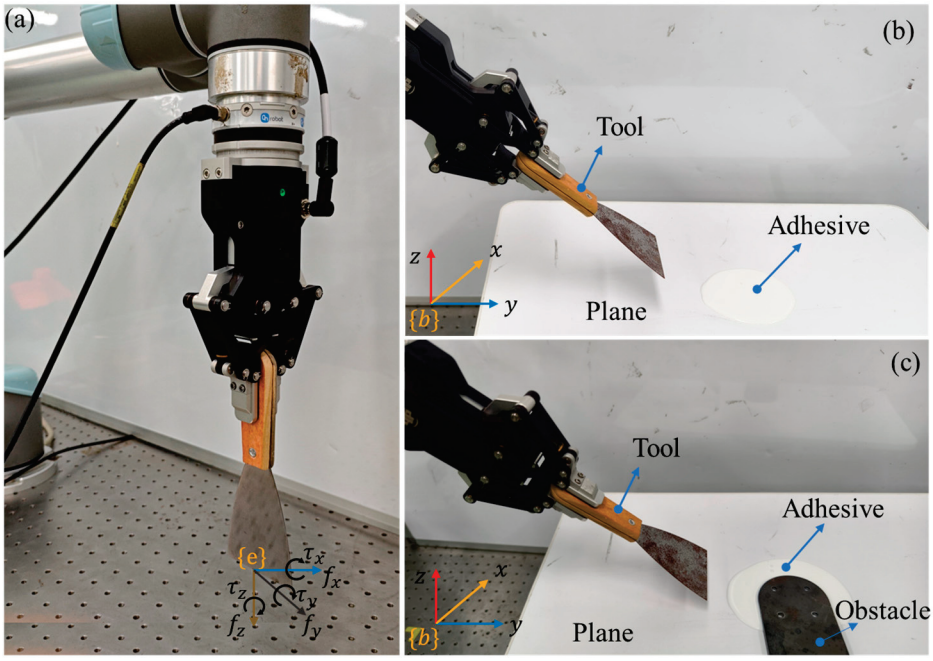


Figure 7. Robotic adhesive scraping manipulation task: (**a**) robot grasping a scraper; (**b**) scraping off the adhesive on a flat surface; and (**c**) scraping adhesive from an obstacled flat surface.

During the adhesive scraping task, the robot grasps the scraper and moves in the positive y-direction of the world coordinate system {b}. The desired contact force of 5 N along the y-axis (motion direction) and z-axis (contact normal) are set as the input of the controller. Figure 8a,b show the regulated tool-end force and position trajectories during

a robotic scraping manipulation control when the adhesive is on a flat surface and on an obstacled flat surface. It can be seen that, after establishing stable contact with a surface at the time point marked by dashed line 1, the controller is triggered and regulates the force values along the y-axis and z-axis near 5 N. In Figure 8b, the time point marked by dashed line 2 indicates the moment the scraper contacts the obstacle. This contact results in sudden changes in the force values along the x-axis and y-axis. The contact force in the x-axis direction causes the robot to generate a lateral displacement response within 0.02 s. This offset is approximately 0.9 cm, enabling the robot to successfully navigate around the obstacle. After navigating around the obstacle, the force values on the y- and z-axes quickly return to the pre-set contact force 5 N. For both Figure 8a,b, dashed line 3 indicates the end of the 'scraping the adhesive' task. Obviously, the estimated tool-end force can successfully be applied to a robotic scraping manipulation control.

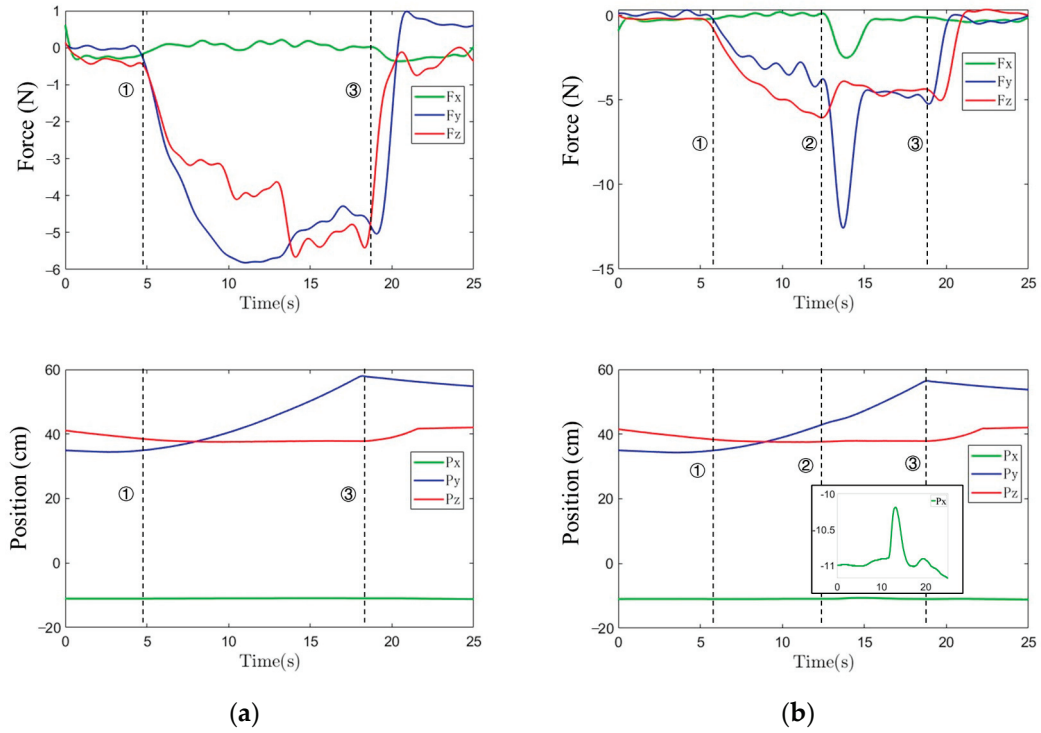


Figure 8. Regulated tool-end force and position trajectories during a robotic scraping manipulation control: (**a**) scraping off the adhesive from a flat surface; and (**b**) scraping off the adhesive from an obstacled flat surface.

4. Discussion

The above experiments verified that the proposed two-stage compensation method for zero-drift and creep error can precisely and rapidly decouple the tool-end interaction force from a wrist force/torque sensor, enabling robots to perform tool manipulation control. The tested four different payloads ranged from 0.89 kg to 3.51 kg, covering the typical mass range of most tools. As shown in Table 1, the identification error is largest when the payload is 0.89 kg, suggesting that the accuracy may degrade further for lighter tools due to the increased signal-to-noise ratio and quantization effects. Since the tool is grasped by a gripper, the identified payload includes both the gripper weight (1 kg) and the tool mass, making it incapable of identifying some light tools. In our nail-hammering and adhesive-scraping experiments, the wooden mallet and scraper employed weighed 246 g and 60 g, respectively. In robotic tool manipulation, consideration must be given not only to the tool selection but also to the permissible load of the robotic arm and the maximum load capacity of the gripper. As the maximum payload capacity of the gripper is 3 kg, and

the robot itself is limited to 5 kg in the experiments, these constrains limit the performance verification beyond this range.

The method currently assumes a rigid tool model and fixed grasp configuration. Flexible or deformable tools could introduce dynamic coupling effects that are not captured by the rigid-body dynamics model, thereby reducing the estimation fidelity. Moreover, the SNN-based compensator was trained and executed on a single sensor configuration, and its long-term robustness to sensor aging or hardware degradation remains to be validated. The current method also neglects the possible crosstalk among different error components during dynamic identification. A further investigation of such crosstalk effects, from both theoretical and material perspectives, may further improve the accuracy of the interaction force estimation.

5. Conclusions

In this paper, we present an online identification method for the six-axis interaction forces at the end of any gripper and tool. This method combines the LSM and SNN to enhance the accuracy of tool-end interaction force identification through gripper–tool dynamic parameter identification, sensor zero-drift identification, and sensor uncertainty error compensation. The experimental results validated the accuracy of our proposed method for identifying the tool-end interaction forces, essentially achieving the in situ measurement of robotic tool-end interaction forces. We have also demonstrated that the force provided by this method exhibits a high precision and quick response within the impedance control framework. Future work will focus on extending the proposed online 6-DoF force/torque identification method to humanoid robots with multi-fingered hands. In addition, we will validate its performance in complex manipulation tasks involving multiple grasping modes and multi-contact configurations.

Author Contributions: Methodology, J.X.; validation and data curation, J.X. and D.W.; writing—original draft preparation, D.W. and J.X.; writing—review and editing, Y.X.; visualization, Y.X.; supervision, Y.X., T.L. and J.Z.; funding acquisition, Y.X. and W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 52375027 and by the National Engineering Research Center for Petroleum Refining Technology and Catalyst of China (Grant No. ZC0609-0014).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Billard, A.; Kragic, D. Trends and challenges in robot manipulation. *Science* **2019**, *364*, eaat8414. [CrossRef]
2. Huan, Z.; Ma, W.; Wang, J.; Wu, F. Path planning and optimization for micro-robot in a vessel-mimic environment. *Front. Neurorobot.* **2022**, *16*, 923348. [CrossRef]
3. Yang, C.; Xie, Y.; Liu, S.; Sun, D. Force Modeling, Identification, and Feedback Control of Robot-Assisted Needle Insertion: A Survey of the Literature. *Sensors* **2018**, *18*, 561. [CrossRef]
4. Wang, R.; Xie, Y.; Liu, H.; Zhou, W. Center-of-Mass-Based object regrasping: A reinforcement learning approach and the effects of perception modality. *IEEE/ASME Trans. Mechatron.* **2024**, *30*, 1356–1365. [CrossRef]
5. Tang, L.; Jia, Y.-B.; Xue, Y. Robotic manipulation of hand tools: The case of screwdriving. In Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 13–16 May 2024; pp. 13883–13890.

6. Xia, Z.; Deng, Z.; Fang, B.; Yang, Y.; Sun, F. A review on sensory perception for dexterous robotic manipulation. *Int. J. Adv. Robot. Syst.* **2022**, *19*, 17298806221095974. [CrossRef]
7. Liu, X.; Zhao, F.; Ge, S.S.; Wu, Y.; Mei, X. End-effector force estimation for flexible-joint robots with global friction approximation using neural networks. *IEEE Trans. Ind. Inform.* **2018**, *15*, 1730–1741. [CrossRef]
8. Le, D.P.; Choi, J.; Kang, S. External force estimation using joint torque sensors and its application to impedance control of a robot manipulator. In Proceedings of the 2013 13th International Conference on Control, Automation and Systems (ICCAS 2013), Gwangju, Republic of Korea, 20–23 October 2013; pp. 1794–1798.
9. Choi, J.; Kang, S. External force estimation using joint torque sensors for a robot manipulator. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, St. Paul, MN, USA, 14–18 May 2012; pp. 4507–4512.
10. He, W.; Yan, Z.; Sun, C.; Chen, Y. Adaptive neural network control of a flapping wing micro aerial vehicle with disturbance observer. *IEEE Trans. Cybern.* **2017**, *47*, 3452–3465. [CrossRef] [PubMed]
11. He, W.; Meng, T.; He, X.; Sun, C. Iterative learning control for a flapping wing micro aerial vehicle under distributed disturbances. *IEEE Trans. Cybern.* **2018**, *49*, 1524–1535. [CrossRef]
12. Leboutet, Q.; Roux, J.; Janot, A.; Guadarrama-Olvera, J.R.; Cheng, G. Inertial parameter identification in robotics: A survey. *Appl. Sci.* **2021**, *11*, 4303. [CrossRef]
13. Lutter, M. A differentiable newton–euler algorithm for real-world robotics. In *Inductive Biases in Machine Learning for Robotics and Control*; Siciliano, B., Khatib, O., Eds.; Springer Tracts in Advanced Robotics; Springer Nature: Cham, Switzerland, 2023; pp. 9–34. Available online: https://link.springer.com/chapter/10.1007/978-3-031-37832-4_2 (accessed on 24 October 2025).
14. Huang, W.; Min, H.; Guo, Y.; Liu, M. A review of dynamic parameters identification for manipulator control. *Cobot* **2022**, *1*, 5. [CrossRef]
15. Xie, K.; Cao, Q.; Luan, N.; Zhu, X. External Forces/Torques perception of robot endeffector tip based on a 6-axis force sensor. In Proceedings of the 2021 International Conference on Intelligent Computing, Automation and Systems (ICICAS), Chongqing, China, 29–31 December 2021; pp. 440–444.
16. Zhang, L.-J.; Hu, R.-Q.; Yi, W.-M. Research on force sensing for the end-load of industrial robot based on a 6-axis force/torque sensor. *Acta Autom. Sin.* **2017**, *43*, 439–447. [CrossRef]
17. Farsoni, S.; Bonfè, M. Complete and consistent payload identification during human-robot collaboration: A safety-oriented procedure. In Proceedings of the Human-Friendly Robotics 2021: HFR: 14th International Workshop on Human-Friendly Robotics, Bologna, Italy, 28–29 October 2021; pp. 15–28.
18. Nadeau, P.; Giamou, M.; Kelly, J. Fast object inertial parameter identification for collaborative robots. In Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; pp. 3560–3566.
19. Duan, J.; Liu, Z.; Bin, Y.; Cui, K.; Dai, Z. Payload identification and gravity/inertial compensation for six-dimensional force/torque sensor with a fast and robust trajectory design approach. *Sensors* **2022**, *22*, 439. [CrossRef] [PubMed]
20. Su, H.; Yang, C.; Mdeihly, H.; Rizzo, A.; Ferrigno, G.; De Momi, E. Neural network enhanced robot tool identification and calibration for bilateral teleoperation. *IEEE Access* **2019**, *7*, 122041–122051. [CrossRef]
21. Su, H.; Qi, W.; Yang, C.; Sandoval, J.; Ferrigno, G.; De Momi, E. Deep neural network approach in robot tool dynamics identification for bilateral teleoperation. *IEEE Robot. Autom. Lett.* **2020**, *5*, 2943–2949. [CrossRef]
22. Cao, M.Y.; Laws, S.; Ferdinando, R.y.B. Six-axis force/torque sensors for robotics applications: A review. *IEEE Sens. J.* **2021**, *21*, 27238–27251. [CrossRef]
23. Kebede, G.A.; Ahmad, A.R.; Lee, S.-C.; Lin, C.-Y. Decoupled six-axis force–moment sensor with a novel strain gauge arrangement and error reduction techniques. *Sensors* **2019**, *19*, 3012. [CrossRef]
24. Chavez, F.J.A.; Nava, G.; Traversaro, S.; Nori, F.; Pucci, D. Model based in situ calibration with temperature compensation of 6 axis force torque sensors. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 5397–5403.
25. Zhao, Y.; Liu, Y.; Li, Y.; Hao, Q. Development and application of resistance strain force sensors. *Sensors* **2020**, *20*, 5826. [CrossRef] [PubMed]
26. Xue, X.; Huang, H.; Zuo, L.; Wang, N. A compliant force control scheme for industrial robot interactive operation. *Front. Neurorobot.* **2022**, *16*, 865187. [CrossRef]
27. Yu, Y.; Shi, R.; Lou, Y. Bias estimation and gravity compensation for wrist-mounted force/torque sensor. *IEEE Sens. J.* **2021**, *22*, 17625–17634. [CrossRef]
28. Shin, J.; Shin, J.; Jung, H.-r.; Won, J.; Auh, E.; Moon, H. On-Site Implementation of External Wrench Measurement via Non-Linear Optimization in Six-Axis Force–Torque Sensor Calibration and Crosstalk Compensation. *Appl. Sci.* **2025**, *15*, 1510. [CrossRef]
29. Yao, L.; Gao, Q.; Zhang, D.; Zhang, W.; Chen, Y. An integrated compensation method for the force disturbance of a six-axis force sensor in complex manufacturing scenarios. *Sensors* **2021**, *21*, 4706. [CrossRef]
30. Liu, M.; Li, P.; Duan, J.; Liu, L.; Shen, Y.; Tian, W.; Ji, Y. A two-stage payload dynamic parameter identification method for interactive industrial robots with large components. *IEEE Trans. Autom. Sci. Eng.* **2025**. [CrossRef]

31. Wu, J.; Huang, T.; Zhu, Z.; Song, K. Cold starting temperature time-related compensation model of inertial sensors based on particle swarm optimization algorithm. *Rev. Sci. Instrum.* **2021**, *92*, 065106. [CrossRef]
32. Sun, Y.; Liu, Y.; Liu, H. Temperature compensation for a six-axis force/torque sensor based on the particle swarm optimization least square support vector machine for space manipulator. *IEEE Sens. J.* **2015**, *16*, 798–805. [CrossRef]
33. Gao, Z.; Wen, Y.; Li, P.; Lei, H.; Xiao, Q.; Fan, Y. Creep compensation for mechanical sensors: Considering historical loading memory. *Measurement* **2025**, *249*, 116901. [CrossRef]
34. Sousa, C.D.; Cortesao, R. Physical feasibility of robot base inertial parameter identification: A linear matrix inequality approach. *Int. J. Robot. Res.* **2014**, *33*, 931–944. [CrossRef]
35. Lotfi Rezaabad, A.; Vishwanath, S. Long short-term memory spiking networks and their applications. In Proceedings of the International Conference on Neuromorphic Systems, Oak Ridge, TN, USA, 28–30 July 2020; pp. 1–9.
36. Jiangtao, X.; Yu, X.; Renpeng, W.; Jinhui, Z.; Wei, Z.; Yi, Z. A Novel Robot Tool Manipulation Method for Battery Degumming Task. In Proceedings of the 2024 7th International Symposium on Autonomous Systems (ISAS), Chongqing, China, 7–9 May 2024; pp. 1–5.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

sensors

MDPI

Article

A Sensor Based Waste Rock Detection Method in Copper Mining Under Low Light Environment

Jianing Ding [1,2], Fuming Qu [1,*], Weihua Zhou [1,2], Jiajun Xu [1,2], Lingyu Zhao [1,3] and Yaming Ji [1,2]

[1] Institute of Minerals Research, University of Science and Technology Beijing, Beijing 100083, China; m202320067@xs.ustb.edu.cn (J.D.); m202420145@xs.ustb.edu.cn (W.Z.); m202510094@xs.ustb.edu.cn (J.X.); lingyulanazhao@xs.ustb.edu.cn (L.Z.); m202320089@xs.ustb.edu.cn (Y.J.)
[2] School of Resources and Safety Engineering, University of Science and Technology Beijing, Beijing 100083, China
[3] School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China
* Correspondence: qufuming@ustb.edu.cn

Abstract

During production, copper mining could generate substantial waste rock that impacts land use and the environment. Advances in deep learning have enabled efficient, cost-effective intelligent sorting, where vision sensor performance critically determines sorting accuracy and efficiency. However, the sorting environment of copper mine waste rock is inherently complex, particularly within the conveyor belt section of the sorting machine, where insufficient and uneven lighting significantly impairs the performance of vision-based detection systems. To address the above challenges, a deep-learning-based copper mine waste rock detection algorithm under low-light environments is proposed. Firstly, an Illumination Adaptive Transformer (IAT) module is added as a preprocessing layer at the beginning of the Backbone to enhance the brightness of the images acquired by the vision sensor. Secondly, a Local Enhancement-Global Modulation (LEGM) module is integrated after the A2C2f and C3k2 modules in the Neck to enhance the detection accuracy. Finally, to further improve the model performance, MPDIoU is introduced to optimize the original loss function CIoU. As a result, the proposed algorithm achieved an mAP@0.5 of 0.957 and an mAP@0.5:0.95 of 0.689, outperforming advanced methods by 1.9% and 8.6%, respectively.

Keywords: mineral processing; vision sensor; deep learning; object detection; low light

1. Introduction

Copper plays a crucial role in supporting the development of modern society, technology and infrastructure [1]. It is widely used in various industries, such as electric power, machinery, construction, transportation, etc. [2], making the efficient and sustainable utilization of copper ore resources particularly important. During the production process, waste rock should be separated from copper minerals as early as possible to conserve energy and time in subsequent processing steps. Therefore, sorting of copper ore waste rock is not only an urgent need for environmental protection but also an important strategy to achieve sustainable development of the mining industry [3–5].

With the rapid development of the global technological revolution and the advancement of industrialization, technologies in waste rock sorting are also evolving continuously. The traditional manual sorting model, characterized by high labor intensity, low efficiency, and poor precision under harsh working conditions, can no longer meet the demands of

high-quality industry development. With the widespread use of deep learning in object detection, computer-vision-based intelligent sorting technology has emerged, providing an efficient solution for the automated sorting of copper mine waste rock [6,7]. However, practical implementation still faces significant technical challenges, the most significant being low and uneven lighting inside sorting machines. Such conditions make it difficult for image sensors to capture ore features accurately, leading to lower image quality, loss of detail, and frequent detection errors. Therefore, developing a real-time detection method for waste rocks on conveyor belts under low-light environments is critical for mineral processing development [8,9].

Current research on mineral waste rock sorting is primarily based on computer vision classification algorithms. With various image-based learning methods [10], it can preprocess input mineral images and construct classification models to achieve mineral sorting. For example, Liu Y et al. [11] designed a lightweight ore sorting network (LOSN), which integrates lightweight structures, attention mechanisms, and multi-scale feature fusion techniques to achieve end-to-end localization and classification of ore particles. Shatwell et al. [12] developed an algorithm based on image processing and machine learning, specifically designed for classifying high-grade ore rocks and low-grade waste rocks in gold and silver mines, enabling real-time ore sorting. Tsangaratos et al. [13] proposed a real-time mineral classification system that integrates classical computer vision techniques with advanced deep learning algorithms. This system can identify multiple minerals within a single frame and output probabilities for multiple mineral types, facilitating real-time mineral classification. However, existing methods heavily rely on the quality of input image. This dependence not only makes image acquisition challenging under harsh working conditions but also significantly increases recognition time and reduces sorting efficiency. Low-light or uneven-lighting conditions in actual sorting environments further degrade image quality, which makes feature extraction more difficult. In recent years, numerous scholars from various fields have developed methods to enhance real-time target detection in low-light environments by improving the input image quality, focusing primarily on image enhancement, small-target detection, noise suppression, and target deblurring.

Regarding image enhancement, Hui Y et al. [14] proposed the WSA-YOLO algorithm, a weakly supervised adaptive object detection algorithm for low-light environments based on YOLOv7. This algorithm decomposes images into reflectance and illumination maps for separate enhancement through a decomposition network, addressing performance degradation in object detection caused by the loss of fundamental image features under low-light conditions and the imbalance in low-level semantic information between low-light and normal images. Zhou Y et al. [15] introduced an improved spatial object detection method based on YOLO11, which enhances image quality by applying bilateral filtering to remove noise while preserving edge texture details and using Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve image contrast. Zhang Q et al. [16] proposed the multi-module network LLD-YOLO for robust vehicle detection in low-light conditions. It enhances low-light images by integrating a DarkNet module based on self-calibrated illumination learning for adaptive illumination adjustment.

In the small target detection field, Zhang M et al. [17] proposed a Gaussian Curvature-Inspired Network (GCI-Net) for single-frame infrared small target detection, addressing the difficulty of separating small infrared targets from backgrounds with high brightness and strong edges in complex environments where they share similar physical characteristics. Wang S et al. [18] introduced a lightweight infrared small target detection algorithm called PHSI-RTDETR for drone aerial imagery, tackling issues of low model accuracy and high computational complexity caused by complex ground environments and uneven target scales in infrared image target detection. Wang Z et al. [19] proposed the PC-YOLO11s

model to overcome detection difficulties in small target scenarios arising from factors such as small size, low resolution, weak contrast, and significant background interference. Fang, W. et al. [20] proposed a lightweight TBF-YOLOv8n tea bud detection model based on an improved YOLOv8n to address issues such as the high computational complexity of deep learning detection models and the difficulty of detecting tea buds due to their small size, complex background, and high density.

Regarding noise suppression and target blurring issues. Peng D et al. [21] proposed NLE-YOLO, a novel low-light object detection method based on YOLOv5 with fused feature enhancement, to overcome accuracy issues caused by insufficient illumination and noise interference in low-light environments. Das P P et al. [22] introduced YOLO-D, a domain-adaptive low-light object detection framework based on YOLOv8, designed to mitigate performance degradation in conventional detection systems due to low visibility, poor contrast, and high noise in low-light conditions. Liu Q et al. [23] developed an efficient object detection network architecture suitable for complex low-light environments, tackling low accuracy issues caused by blurring and low contrast during target detection under weak light. Han Z et al. [24] proposed 3L-YOLO, a lightweight low-light object detection algorithm based on YOLOv8n, addressing performance degradation and high computational resource consumption resulting from low contrast, high noise, and blurred boundaries in low-light detection. Ye S et al. [25] presented YES, a novel low-light object detection framework, to resolve the "target-background confusion" issue caused by uneven illumination in dim environments. Qi Z et al. [26] proposed the HSG-Net model, based on the lightweight YOLOv8, aiming to enhance detection accuracy and reduce computational costs in variable lighting conditions and complex backgrounds.

Additionally, some researchers have focused on improving real-time performance and computational efficiency from a methodological perspective rather than relying solely on input quality. Zhao Y et al. [27] proposed a Real-time Efficient Enhancement Network (RTEE-Net) for low-light images, addressing the trade-off between model complexity and performance in low-light image enhancement. Qiu Z et al. [28] introduced GS-YOLO-Seg, an improved YOLO11-Seg-based method, to resolve issues of low efficiency in traditional approaches and performance bottlenecks on edge devices during low-grade graphite ore sorting. Li J et al. [29] developed an enhanced model based on YOLOv5, featuring a parallel convolutional architecture, which improves the accuracy and efficiency of photovoltaic panel detection by addressing challenges arising from low resolution, small target sizes, and complex backgrounds in remote sensing imagery.

Based on a comprehensive analysis of existing research, various algorithms demonstrate certain advantages within their respective low-light application contexts. However, some inadequately extract comprehensive features when sorting copper minerals in complex and dim environments. Others are tailored to particular scenarios and struggle to adapt to specialized operational conditions such as conveyor belts, while certain methods impose demanding hardware requirements due to high model complexity, failing to meet real-time sorting needs.

This study proposes a novel method for identifying waste rock in copper ore within conveyor systems, specifically designed for operation under low-light conditions. Industrial experiments were conducted in environments free from overexposure or significant noise interference. Furthermore, the use of high-frame-rate industrial cameras minimized image distortion caused by abnormal exposure and noise. Consequently, the core objective of this research is to overcome recognition challenges in low-light environments. To address issues such as low detection accuracy, high false negative rates, and high false positive rates caused by blurred images resulting from dim lighting and uneven illumination, we introduce the YOLO-ILM algorithm. This algorithm enables real-time detection within low-

light conveyor belt environments. Building upon the advanced YOLOv12 framework [30], YOLO-ILM leverages its efficient feature extraction and target localization capabilities. The algorithm further integrates a series of image enhancement and feature optimization techniques specifically designed for low-light and blurry scenarios. Through synergistic module collaboration, it significantly enhances adaptability to complex lighting conditions while ensuring detection speed meets the industrial requirements. The model introduces the following innovations:

(1) To address image quality degradation caused by low-light conditions, an Illumination-Adaptive Transformer module is proposed and integrated as a preprocessing layer at the network front-end to enhance the brightness of subsequent input images.
(2) To improve the detection accuracy of copper ore waste rocks, a method is introduced that integrates local feature embedding into global feature extraction modules following the A2C2f and C3k2 modules.
(3) To further enhance the detection accuracy of copper ore waste rocks, the original loss function has been refined, optimizing the network.

2. Materials and Methods

The proposed YOLO-ILM algorithm for low-light copper mine waste rock detection is depicted in Figure 1, which consists of three core components: the Backbone, the Neck, and the Head.

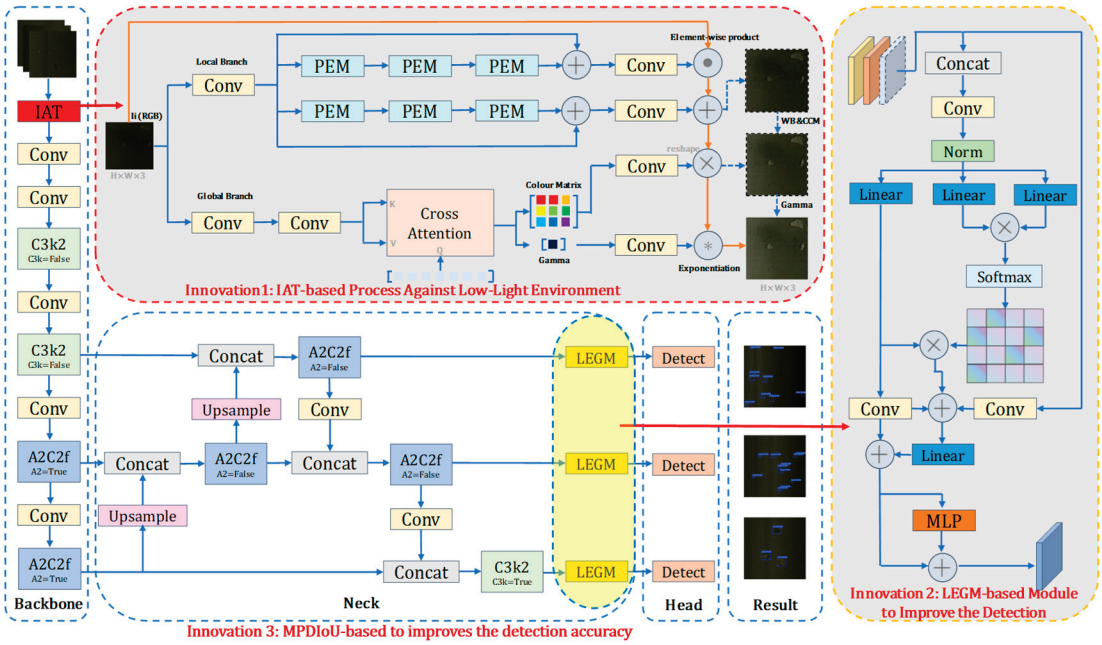


Figure 1. YOLO-ILM Model Architecture Diagram (The red arrow represents the newly added module, whereas the remaining arrows indicate connections between the existing workflow steps.).

In the Backbone section, to address the issue of poor input quality under low-light conditions, the Illumination Adaptive Transformer module (IAT) [31] is incorporated as a preprocessing layer at the front-end of the Backbone. This module effectively mitigates the adverse effects of uneven illumination, thereby providing high-quality input for subsequent copper mine waste rock detection tasks. Its lightweight design also ensures efficient operation under computational resource constraints.

In the Neck section, the preprocessed waste rock images from copper mines are input into the Backbone for feature extraction after IAT preprocessing. The extracted features subsequently enter the Neck for feature fusion. To more comprehensively integrate

multi-scale features and enhance the model's discriminative capability for waste rock characteristics, the Local Feature-Embedded Global Feature Extraction Module (LEGM) [32] is integrated after both the A2C2f and C3k2 modules within the Neck. By combining self-attention and convolution, it integrates local and global features; by aggregating multi-source inputs and depth information, it achieves multi-scale feature fusion. This enables the model to adaptively learn dependencies between different features, enhances focus on critical regions, and suppresses noise interference. The fused features are then forwarded to the Head for final prediction.

In the Head section, the Minimum Points Distance IoU (MPDIoU) loss function [33] is introduced, which places greater emphasis on the positional relationships between bounding boxes and their overlapping regions. This further improves the detection accuracy for copper mine waste rock.

2.1. Illumination Adaptive Transformer Module

Copper mine waste rock detection in real-world environments suffers from low-light and insufficient-light conditions, resulting in degraded visual quality that adversely affects downstream computer vision tasks. Images captured under weak or insufficient lighting exhibit darkness due to limited photon counts and in-camera noise. Therefore, it is necessary to pre-integrate an illumination-adaptive preprocessing module. The method of "optimizing input quality first and then improving model performance" has been verified in many fields [34]. Numerous solutions have been proposed in existing research, such as low-light enhancement or methods improving robustness in low-light and over-exposed conditions with oversaturation. However, these approaches have limitations and shortcomings. Low-light enhancement methods, while recovering details and suppressing accompanying noise, may inadvertently obscure fine details. Furthermore, when local image areas are severely over-exposed (pure white) or under-exposed (pure black), the original scene's brightness information is irretrievably lost. Methods enhancing robustness to oversaturation under low-light and over-exposure conditions suffer from high parameter complexity and significant computational cost.

In contrast, the Illumination Adaptive Transformer (IAT) addresses both low-light enhancement and exposure correction through its designed lightweight dual-branch structure. IAT simulates the camera Image Signal Processor (ISP) pipeline, as illustrated in Figure 2, dividing image enhancement into two complementary components: local pixel-level illumination adjustment and global dynamic adjustment of ISP parameters.

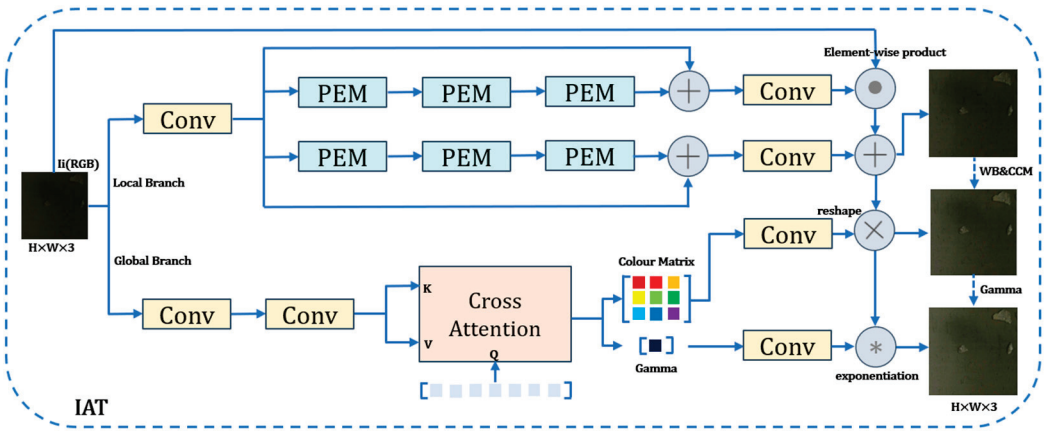


Figure 2. IAT Structure Diagram.

For an input sRGB image Ii captured under lighting condition Li, illumination-induced degradation is addressed through the Illumination Adaptive Transformer (IAT) module via its two core components: local pixel-level adjustment and global ISP parameter adjustment.

First, illumination effects are corrected while preserving image details via the Local Branch. This branch employs a transformer-style architecture, processing the image while maintaining the input resolution. Its core consists of stacked Pixel Enhancement Modules (PEMs). Each PEM first encodes positional information using a 3×3 depth-wise convolution. Subsequently, a Pointwise Convolution (PWConv)—Depth-wise Convolution (DWConv)—Pointwise Convolution (PWConv) structure enhances local details. Two separate 1×1 convolutions then strengthen feature representations. Innovatively, standard Layer Normalization is replaced with Illumination Normalization (I Norm), which learns scaling factor a and bias term b applied before channel fusion. Three PEM modules are stacked within the branch. Their outputs are fused with the input features via skip connections (element-wise addition), effectively preserving original details. Finally, a 3×3 convolution reduces dimensionality, followed by the application of ReLU (yielding a multiplicative map M) and Tanh (yielding an additive map A) activation functions. This design achieves a balance between lightweight implementation and detail enhancement.

Second, parameter modulation is performed by the Global Branch. This branch forms the core of the IAT model for handling ISP pipeline operations and generating the target sRGB image. It focuses on capturing global interactions while maintaining lightweight efficiency. The process is as follows: First, a lightweight encoder, composed of two stacked convolutional layers, encodes the input image. This design reduces feature map resolution to minimize computational cost while extracting high-dimensional global features. The encoded features are then fed into the Global Prediction Module (GPM). Within the GPM, a globally initialized component query (Q) attends to keys (K) and values (V) generated from the encoded features. Notably, positional encoding for K and V is implemented via depth-wise convolution, enabling flexible adaptation to varying input resolutions. The features are subsequently processed by a Feed-Forward Network (FFN) containing two linear layers. Finally, two specially initialized parameters are appended, outputting the joint color transformation matrix (W) and the gamma correction parameter (γ). The initialization strategy (setting the color matrix to the identity matrix and γ to 1) effectively ensures training stability. This entire structure dynamically controls global ISP-related parameters (color correction, gamma correction) through a transformer-style query mechanism. This enables adaptation to diverse lighting conditions while rigorously maintaining the model's lightweight nature and computational efficiency.

When the IAT module is not incorporated into the model, the image input is expressed as shown in Equation (1):

$$Input = I_i = I_{raw} \tag{1}$$

After incorporating the IAT module into the model, the preprocessed image output by the IAT module is expressed as shown in Equation (2):

$$I_t = max(\sum_{cj} W_{ci,cj} \, (I_i \odot M + A), 0)^{\gamma} \tag{2}$$

Functioning as a preprocessing layer, the IAT module transforms the raw input I_{raw} into an illumination-enhanced image, facilitating subsequent feature extraction by convolutional layers. The enhancement effect is visualized in Figure 3. Figure 3a,b depicts the original image and the IAT-processed result (denoted as IAT Output), respectively. Across diverse scenarios—including Isolated (single waste rock), Plural (multiple waste rocks), Small (smaller waste rocks), Large (larger waste rocks), and notably Extremely Dark (severe low-light conditions commonly encountered in areas like conveyor belt edges

where illumination is insufficient and ambient brightness is critically low)—the comparison between original and processed images demonstrates the IAT module's pronounced enhancement effect on low-illumination imagery. Additionally, this study employs Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as primary metrics for evaluating image enhancement performance. Higher PSNR values indicate smaller differences between the enhanced image and the reference image, while SSIM values closer to 1 signify stronger consistency in image structure and texture. Table 1 presents experimental comparisons between IAT and representative enhancement methods on this dataset. These results further validate IAT's superior image enhancement performance in this study.

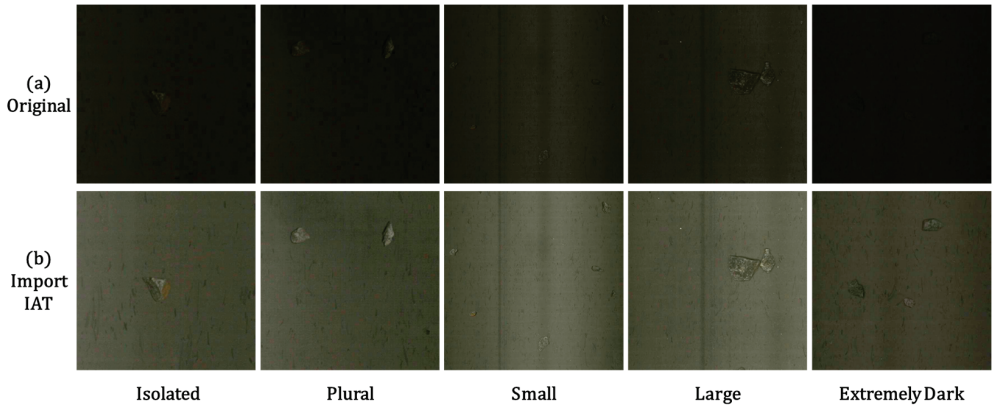


Figure 3. Effect Diagram Before and After IAT Processing.

Table 1. Quantitative Comparison of IAT and Other Mainstream Methods.

Method	Test 1		Test 2		Test 3		Test 4		Test 5	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Zero-DCE [35]	9.207	0.515	9.429	0.528	9.969	0.602	9.372	0.573	10.449	0.225
RetinexNet [36]	6.705	0.324	6.714	0.450	7.009	0.311	6.805	0.454	6.584	0.197
RT-X Net [37]	10.235	0.415	10.521	0.448	10.874	0.432	10.236	0.529	11.145	0.375
IAT	15.163	0.679	15.485	0.633	15.487	0.705	15.407	0.686	16.021	0.669

2.2. LEGM Attention Mechanism

Building upon the base model, the LEGM attention mechanism is introduced to enhance both detail preservation and global context modeling. The fundamental principle of self-attention mechanisms is that when processing sequential data, each element can establish relationships with any other element in the sequence, rather than relying solely on adjacent elements. It adaptively captures long-range dependencies by calculating the relative importance between elements. Consequently, the self-attention module computes association weights between all positions in the feature map, dynamically aggregating global information.

However, conventional self-attention suffers from a key limitation: it focuses exclusively on global dependencies while neglecting local details, often resulting in structurally blurred restored images. To address this, the LEGM module integrates convolutional layers for local feature embedding. This design combines the strengths of convolutional layers—excellent at local feature extraction—with the self-attention module's ability to model global context. By fusing these capabilities, LEGM effectively models global haze distribution patterns while preserving fine details, thereby significantly enhancing the visual realism of the dehazed results.

The LEGM module integrates a self-attention mechanism with a convolutional module. It concurrently receives and concatenates shallow, mid-level, and deep features. During

the feature processing stage, the concatenated features first pass through convolutional layers and normalization layers to explicitly extract local structures, thereby overcoming the inherent local receptive field limitations of traditional self-attention. These features subsequently transform a Multilayer Perceptron (MLP), are further processed by a linear layer, and finally pass through a Softmax function to generate the attention weight matrix. In the feature fusion stage, this attention weight matrix is element-wise multiplied with the input features. The resulting features are then processed through convolutional and linear layers. To enhance the precision of attention focus—particularly on distant, high-haze-concentration areas—these processed features are additively combined with the original input features via a residual connection. Following this, a filtering operation is applied, and the features are finally output through another Multilayer Perceptron (MLP). By synergistically fusing convolution-based local feature embedding with depth-prior-guided self-attention, LEGM effectively achieves a balance between preserving local details and capturing global structural information for the image dehazing task, as illustrated in Figure 4.

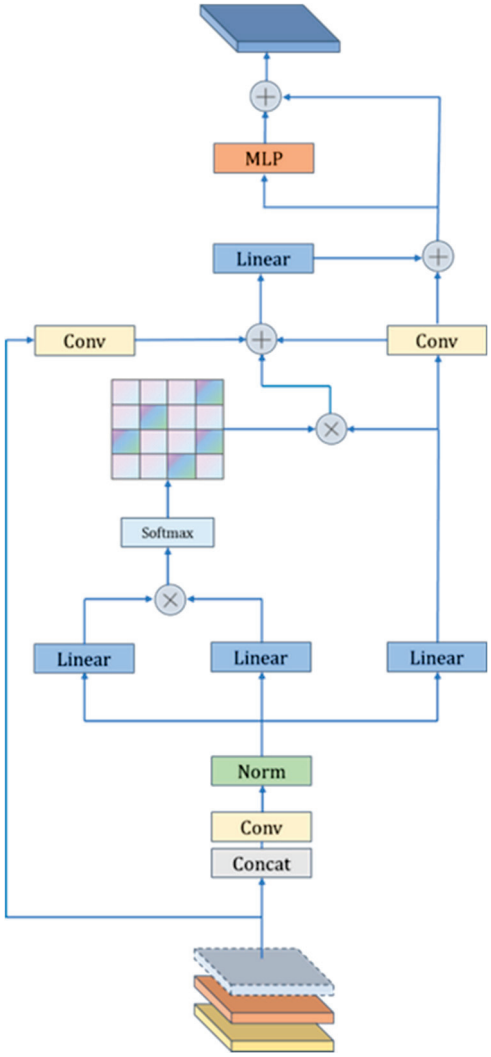


Figure 4. LEGM Structure Diagram.

Fusion of input features:

$$F_{in} = Concat(Conv1 \times 1(F_{unet}), Conv3 \times 3(F_{unet}), DRDB(F_{depth})) \quad (3)$$

Self-attentive module:

$$Q = Conv(F_{in}), K = Conv(F_{in}), V = Conv(F_{in}), d_K = dim(K) \tag{4}$$

$$F_{att} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V \tag{5}$$

MLP transformation:

$$F_{mlp} = MLP(F_{att}) \tag{6}$$

Output Equation:

$$F_{out} = Conv\Big(LayerNorm\Big(F_{in} + F_{mlp}\Big)\Big) \tag{7}$$

Following LEGM integration, the network output is:

$$\begin{cases} F_{out1} = LEGM(A2C2f(F_{in1})) \\ F_{out2} = LEGM(A2C2f(F_{in2})) \\ F_{out3} = LEGM(C3K2(F_{in3})) \end{cases} \tag{8}$$

After incorporating the LEGM module into the original model, the self- attentive module integrates the locally extracted convolutional features with global contextual information, enhancing feature representation capabilities and improving the utilization efficiency of image structure and depth information.

2.3. MPDIoU Loss Function

The original CIoU [38] loss function in the model considers both center-point distance and aspect ratio. However, it focuses solely on relative aspect ratio deviation rather than absolute size differences, severely constraining regression accuracy and convergence efficiency when predicted and ground-truth bounding boxes share identical aspect ratios but differ significantly in scale. In contrast, the MPDIoU loss function currently employed in our model fundamentally resolves this limitation by innovatively minimizing the distances between corresponding top-left and bottom-right corner points of predicted and ground-truth bounding boxes. This approach unifies the geometric properties of bounding boxes (defined by four corner coordinates) into a concise metric that comprehensively incorporates critical factors, including overlap area, center-point distance, and width-height deviations, while directly optimizing absolute size differences. Simultaneously, it streamlines the computational process, delivering an efficient and robust solution for bounding box regression.

The final CIoU loss is computed as follows:

$$L_{CIOU} = L_{base} + \alpha \times \frac{4}{\pi^2}\left(\arctan\frac{w_1}{h_1} - \arctan\frac{w_2}{h_2}\right)^2 \tag{9}$$

where L_{base} denotes the fundamental geometric error, α represents the dynamic weighting parameter, and their computational formulations are given, respectively, as follows:

$$L_{base} = (1 - IoU) + \frac{\rho^2(M_2, M_1)}{c^2} \tag{10}$$

$$\alpha = \begin{cases} 0, & \text{if } IoU < 0.5, \\ \frac{V}{(1-IoU)+V}, & \text{if } IoU \geq 0.5. \end{cases} \tag{11}$$

where $M_1 = (x^A, y^A)$ and $M_2 = (x^B, y^B)$ denote the center coordinates of ground-truth box A and predicted box B, respectively; C represents the minimum enclosing rectangle covering both boxes, ρ corresponds to the Euclidean distance between M_1 and M_2. w_1 and w_2 indicate the widths of boxes A and B; h_1 and h_2 signify their corresponding heights.

The MPDIoU loss function is formulated as follows:

$$L_{MPDIoU} = 1 - MPDIoU \tag{12}$$

The core computational formulation of MPDIoU is defined as follows:

$$MPDIoU = IoU - \left(\frac{d_1^2}{w^2 + h^2} + \frac{d_2^2}{w^2 + h^2} \right) \tag{13}$$

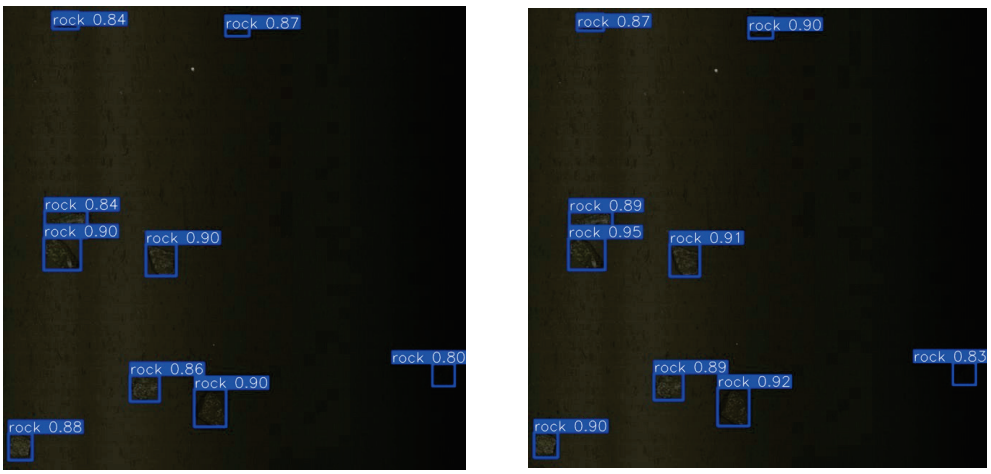
where IoU measures the overlap, and ${d_1}^2$ and ${d_2}^2$ represent the distances from the top-left corner and the bottom-right corner, respectively. These are calculated as follows:

$$IoU = \frac{I}{U} \tag{14}$$

$$d_1^2 = \left(x_1^B - x_1^A\right)^2 + \left(y_1^B - y_1^A\right)^2 \tag{15}$$

$$d_2^2 = \left(x_2^B - x_2^A\right)^2 + \left(y_2^B - y_2^A\right)^2 \tag{16}$$

Consequently, when employed in object detection, MPDIoU offers significant advantages. By comprehensively accounting for overlapping and non-overlapping regions, center point distance, and width-height deviations, it effectively resolves optimization inefficiencies that occur when bounding boxes share identical aspect ratios but differ in absolute dimensions, enabling precise discrimination between predicted boxes located inside and outside the ground truth. Furthermore, its computational efficiency stems from determining relevant factors directly from the four bounding box coordinates, minimizing computational overhead, and facilitating straightforward integration into detection algorithms. Finally, as shown in the Figure 5, MPDIoU accelerates training convergence, thus achieving higher accuracy in shorter timeframes while also delivering superior final detection precision.



(**a**) CIoU (**b**) MPDIoU

Figure 5. Prediction Box Comparisons.

2.4. Evaluation Metrics

To better evaluate the comprehensive performance of each algorithm, based on the specific characteristics of the copper ore waste rock detection scenario, this paper selects

three primary evaluation metrics: Precision (P), Recall (R), and Mean Average Precision (mAP), where:

(1) Precision (P) refers to the proportion of true positive samples among all samples predicted as positive by the model. A higher precision indicates a lower probability of misjudgment by the model. Its calculation formula is:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{17}$$

In the formula, TP (True Positive) represents the number of samples correctly predicted as positive by the model, i.e., samples that are positive and predicted as positive; FP (False Positive) represents the number of samples incorrectly predicted as positive by the model, i.e., samples that are negative but predicted as positive.

(2) Recall (R) refers to the proportion of actual positive samples correctly predicted as positive by the model among all actual positive samples. A higher recall indicates a lower probability of missed detections by the model. Its calculation formula is:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{18}$$

In the formula, FN (False Negative) represents the number of samples incorrectly predicted as negative by the model, that is, samples that are positive but predicted as negative.

(3) Average Precision (AP) refers to the area under the Precision-Recall curve. A higher AP indicates better model precision, recall, and overall performance. Mean Average Precision (mAP) is the average of AP values across all categories. The calculation formulas for AP and mAP are, respectively, as follows:

$$\text{AP} = \int_0^1 P(R)dR \tag{19}$$

$$\text{mAP} = \frac{1}{M}\sum_{i=1}^{M} \text{AP}_i \tag{20}$$

In the formula, AP_i represents the average detection accuracy of the model for a specific category; M represents the total number of categories.

3. Results

3.1. Experimental Dataset

The dataset used in this experiment is a copper ore waste rock dataset collected from the sorting site, which contains a total of 1200 images of copper ore waste rocks on the conveyor belt. To fully evaluate the performance of the algorithm and model, the dataset is divided into a training set, a validation set, and a test set in an 8:1:1 ratio. Specifically, the training set includes 960 images, the validation set contains 120 images, and the test set comprises 120 images. Some samples of the dataset are shown in Figure 6.

3.2. Experimental Environment and Parameters

In this experiment, four mainstream object detection algorithms were selected: Faster R-CNN [39], YOLOv8 [40], YOLOv11 [41], and YOLOv12. During the experiment, each algorithm was loaded onto the copper ore waste rock dataset for training, ultimately resulting in four distinct object detection models.

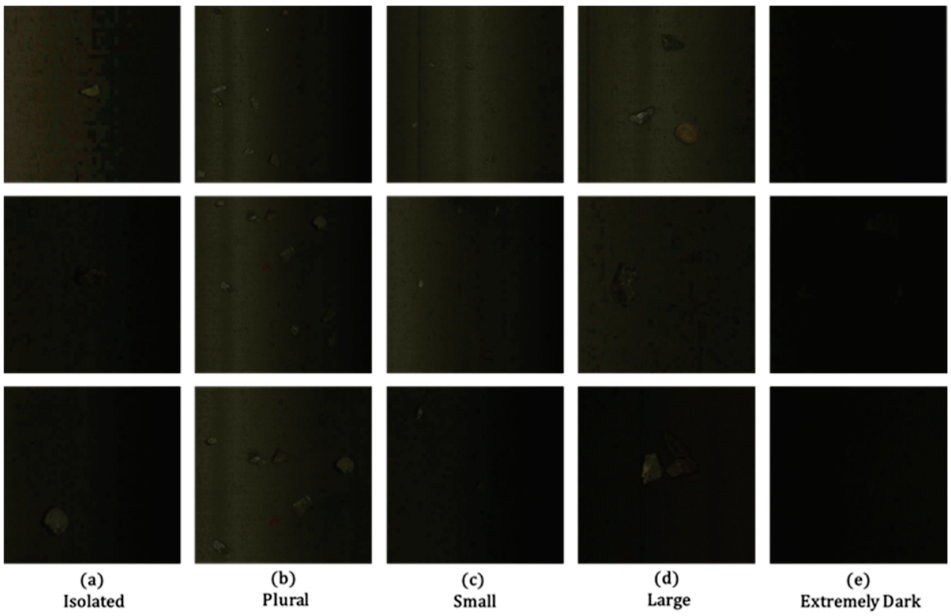


Figure 6. Instance Diagram.

The hardware and software environment for this experiment are detailed in Table 2. In terms of hardware, the core computing device featured a central processing unit (CPU) of Intel Core Ultra 9-275HX (Intel Corporation, Santa Clara, CA, USA) and a graphics processing unit (GPU) of NVIDIA GeForce RTX 5080 (NVIDIA Corporation, Santa Clara, CA, USA) discrete graphics card, equipped with the CUDA platform and its corresponding cuDNN library. On the software side, the experiment was conducted on the Windows 11 operating system, utilizing the Python programming language and the PyTorch deep learning framework. PyCharm was chosen as the integrated development environment for writing, debugging, and executing the code.

Table 2. Experimental Environment.

Hardware/Software Used	Description
CPU	Intel Core Ultra 9-275HX
GPU	NVIDIA GeForce RTX 5080
GPU Parallel Processing Platform	CUDA 12.8
CUDA Deep Neural Network Library	cuDNN v 8.9.7
Operating System	Windows 11
Programming Language	Python 3.12.7
Deep Learning Framework	PyTorch 2.7.1
Integrated Development Environment	PyCharm 2025.1.3.1-Windows

To fully leverage the computational capabilities of the graphics processor used in the experiment and balance model training speed, this experiment sets the input image size to 640×640, the number of training epochs to 200, and the batch size to 32, ensuring efficient model training.

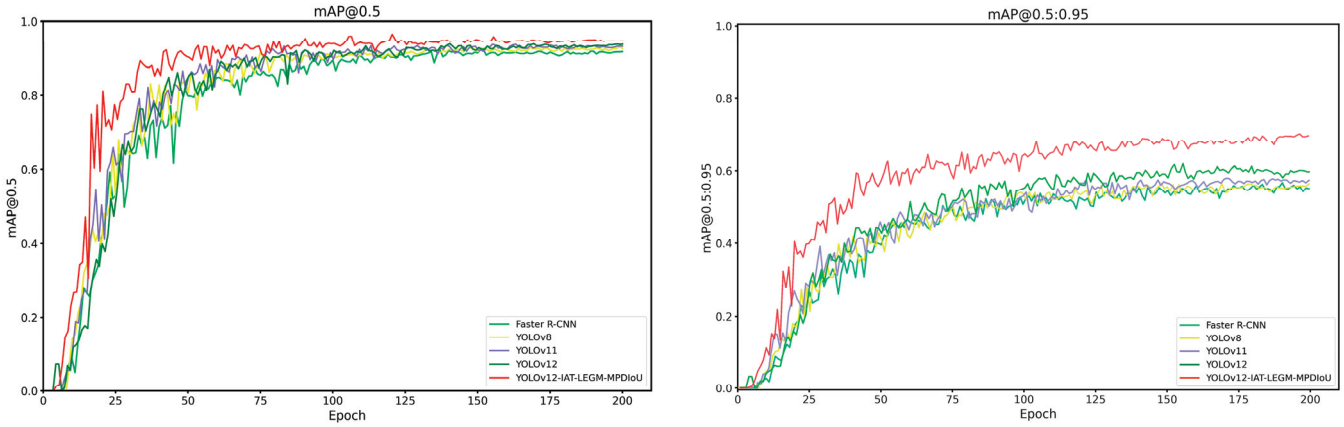
To optimize the training process, the stochastic gradient descent (SGD) algorithm was selected as the optimizer. To enhance the convergence speed and stability of SGD, the momentum mechanism was employed, with the momentum parameter set to 0.937. To prevent the computer from allocating excessive computational resources at once—which could lead to exceeding the upper limit of the graphics processor's memory and cause Out of Memory (OOM) errors—the total number of GPU worker threads was set to 4. This configuration balances computational resource allocation and ensures stable model training.

3.3. Comparative Experiments

To fully validate the superiority of the YOLO-ILM model algorithm, the experiment selected Faster R-CNN, YOLOv8, YOLOv11, and YOLOv12 as comparative models. The experimental results of training and prediction on the copper ore waste rock dataset were used for comparison. The experimental outcomes from training different models are presented in Table 3 and Figure 7. Table 3 displays the Precision (P), Recall (R), and mAP evaluation metrics for each model, while Figure 7 illustrates the mAP@0.5 and mAP@0.5:0.95 results curves across the models.

Table 3. Experimental Results under Different Models.

Model	Precision (P)	Recall (R)	mAP@0.5	mAP@0.5:0.95	GFLOPs	Memory (G)	Inference Times (ms)
Faster R-CNN	0.852	0.876	0.915	0.574	181.4	36.2	29.56
YOLOv8	0.874	0.881	0.923	0.582	28.80	11.7	11.41
YOLOv11	0.880	0.884	0.935	0.594	21.70	7.39	9.85
YOLOv12	0.903	0.910	0.938	0.603	19.60	7.02	6.73
YOLO-ILM	0.948	0.941	0.957	0.689	45.2	9.92	9.43

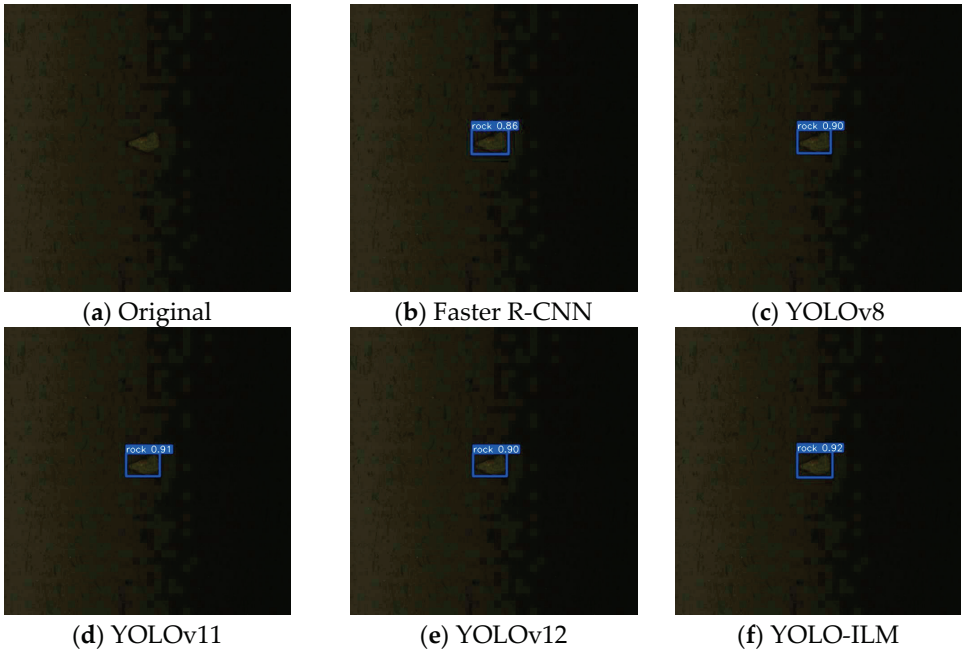


(**a**) mAP@0.5 (**b**) mAP@0.5:0.95

Figure 7. Experimental Plots under Different Models.

Combining Table 3 and Figure 7, it is evident that among all evaluated models, the YOLO series models collectively outperform Faster R-CNN in precision, recall, and mean average precision (mAP). Specifically, YOLO-ILM achieved optimal performance across four key evaluation metrics: precision (p = 0.948), recall (R = 0.941), mAP@0.5 (0.957), and mAP@0.5:0.95(0.689). Compared to the original model, it improved mAP@0.5 and mAP@0.5:0.95 by 1.9% and 8.6%, respectively. Additionally, as evidenced by inference time, although the YOLO-ILM model is 2.7 ms slower than the baseline model, our model can meet real-time detection requirements.

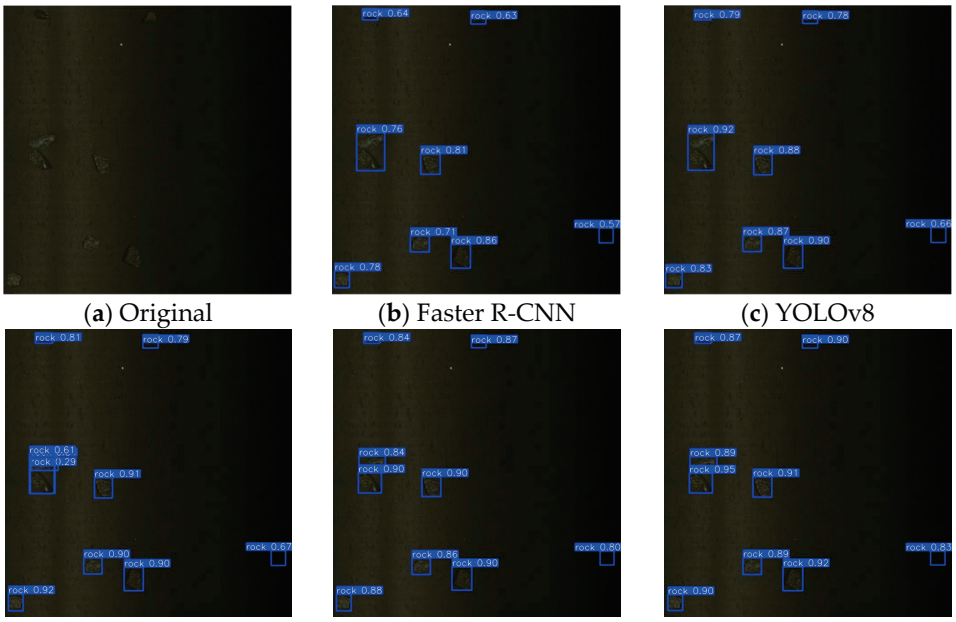
To further validate the effectiveness of the YOLO-ILM algorithm, the trained model is deployed to an on-site sorting machine and conducted field tests in complex low-light environments. The experimental results are shown in Figures 8–12, where "rock" denotes detected copper ore waste. The results demonstrate that for multiple targets, Faster R-CNN and YOLOv8 exhibited partial missed detections (Figure 9) with low confidence, while YOLOv11 and YOLOv12 detected all targets, their confidence was lower than YOLO-ILM. Through comparison with other algorithm models, this fully demonstrates YOLO-ILM's advanced performance in low-light target detection and its suitability for this task.



(**a**) Original (**b**) Faster R-CNN (**c**) YOLOv8

(**d**) YOLOv11 (**e**) YOLOv12 (**f**) YOLO-ILM

Figure 8. Isolated Detection Results.



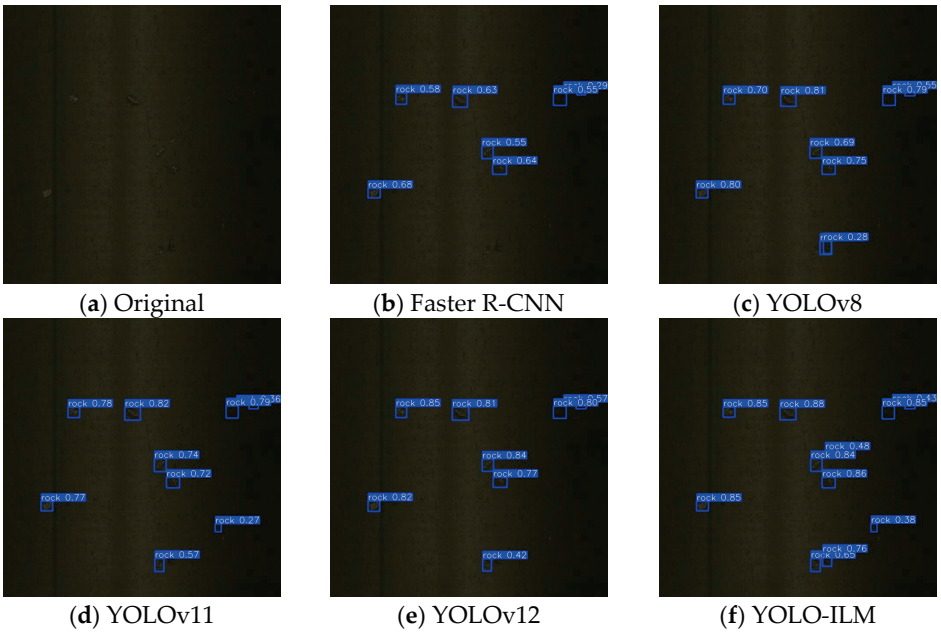
(**a**) Original (**b**) Faster R-CNN (**c**) YOLOv8

(**d**) YOLOv11 (**e**) YOLOv12 (**f**) YOLO-ILM

Figure 9. Plural Detection Results.

3.4. Ablation Experiments

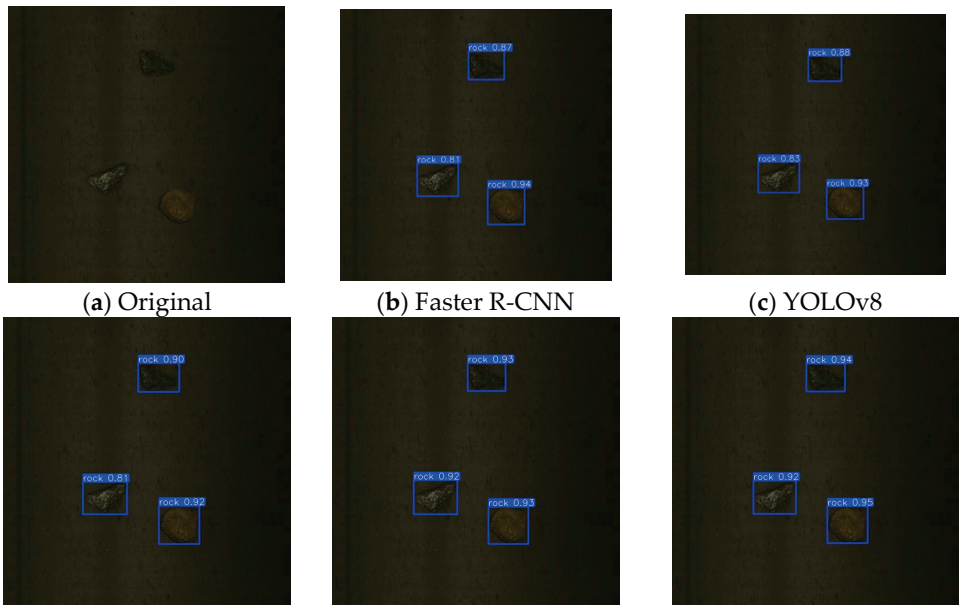
To further validate the effectiveness of each module (IAT, LEGM, MPDIoU) in the proposed algorithm, ablation experiments were conducted on the added modules. The experimental results are shown in Table 4 and Figure 13. In Table 4, using the YOLOv12 model as the baseline, the effectiveness of each improvement method was evaluated by incrementally adding one enhancement at a time. The symbol "$\surd$" indicates that the corresponding improvement module has been incorporated into the model. Figure 13 presents a comparative analysis of mAP@0.5 and mAP@0.5:0.95 results between various model combinations in the ablation experiments and the baseline model.



(**a**) Original (**b**) Faster R-CNN (**c**) YOLOv8

(**d**) YOLOv11 (**e**) YOLOv12 (**f**) YOLO-ILM

Figure 10. Small Detection Results.



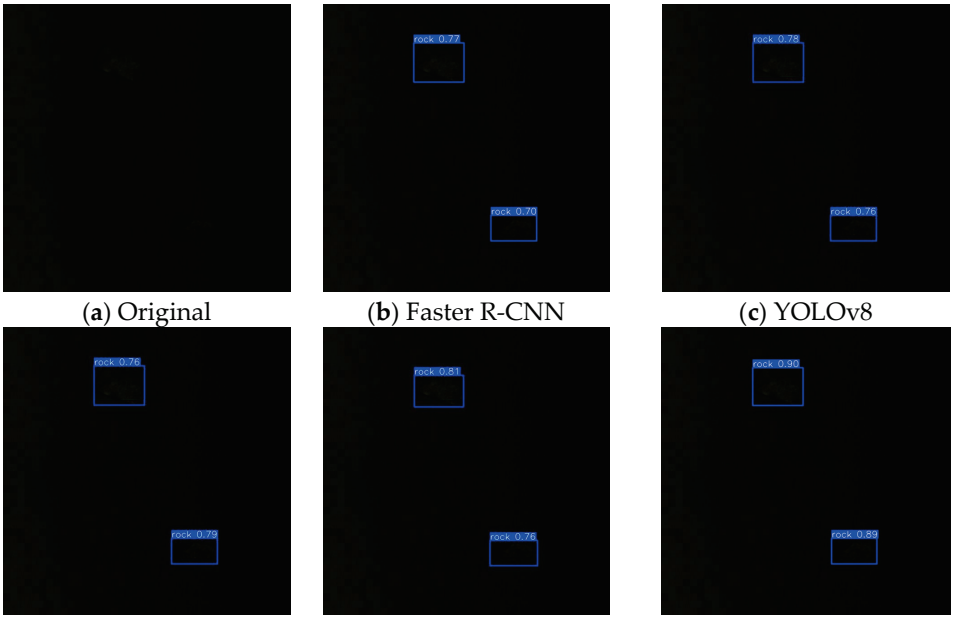
(**a**) Original (**b**) Faster R-CNN (**c**) YOLOv8

(**d**) YOLOv11 (**e**) YOLOv12 (**f**) YOLO-ILM

Figure 11. Large Detection Results.

Table 4. Ablation Test Results.

Test	IAT	LEGM	MPDIoU	Precision (P)	Recall (R)	mAP@0.5	mAP@0.5:0.95	GFLOPs	Inference Times (ms)
1				0.903	0.910	0.938	0.603	19.60	6.73
2	√			0.909	0.912	0.942	0.632	24.20	6.88
3		√		0.916	0.914	0.944	0.649	27.30	8.87
4			√	0.914	0.915	0.940	0.641	19.60	7.13
5	√	√		0.929	0.925	0.958	0.657	45.20	9.31
6	√		√	0.930	0.934	0.952	0.663	24.20	8.92
7		√	√	0.936	0.929	0.949	0.676	27.30	9.34
8	√	√	√	0.948	0.941	0.957	0.689	45.20	9.43

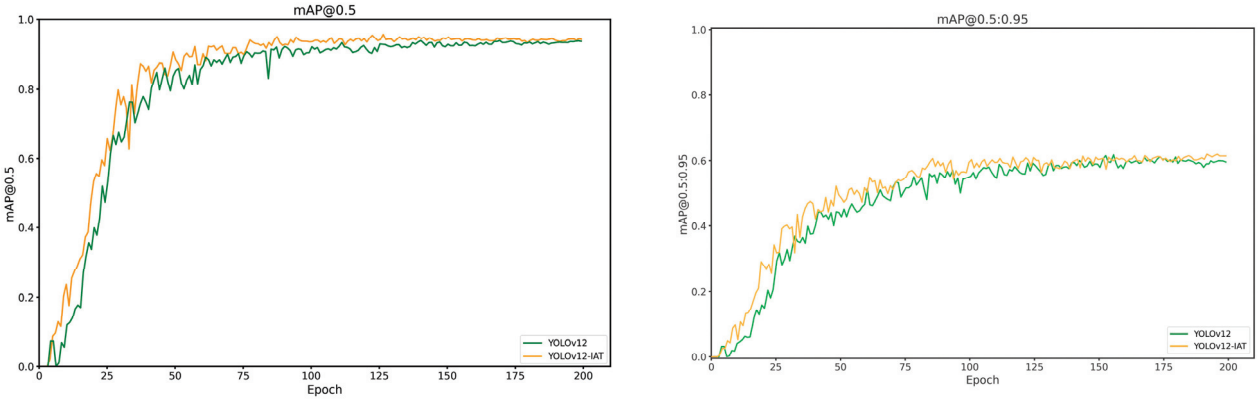


(**a**) Original (**b**) Faster R-CNN (**c**) YOLOv8

(**d**) YOLOv11 (**e**) YOLOv12 (**f**) YOLO-ILM

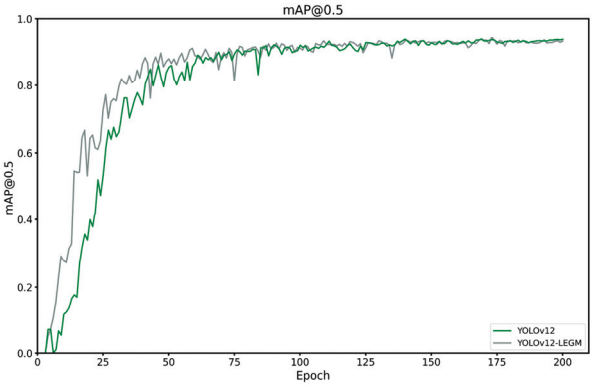
Figure 12. Extremely Dark Detection Results.

Combining Tables 4 and 5, and Figure 13, initially, the baseline model YOLOv12 was tested, achieving mAP@0.5 of 0.938 and mAP@0.5:0.95 of 0.603. Separately introducing the IAT module, LEGM attention mechanism, and MPDIoU loss function each increased both mAP@0.5 and mAP@0.5:0.95 compared to the baseline model, indicating individual contributions from each module. Next, pairwise combinations of modules demonstrated further improvements in mAP@0.5 and mAP@0.5:0.95, with a greater magnitude than single modules alone. This demonstrates synergistic effects when modules are combined pairwise, significantly enhancing the model's object detection performance. Finally, integrating all three improved modules achieved peak results: precision (p = 0.948), recall (R = 0.941), mAP@0.5(0.957), and mAP@0.5:0.95(0.689), representing improvements of 1.9% and 8.6% in mAP@0.5 and mAP@0.5:0.95, respectively, over the baseline model. These results confirm maximal synergistic gains when all three modules operate jointly, delivering comprehensive optimal performance across all evaluation metrics.

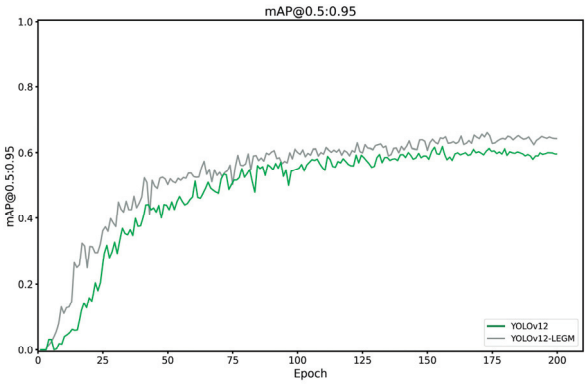


(**a1**) YOLOv12-IAT mAP@0.5 (**b1**) YOLOv12-IAT mAP@0.5:0.95

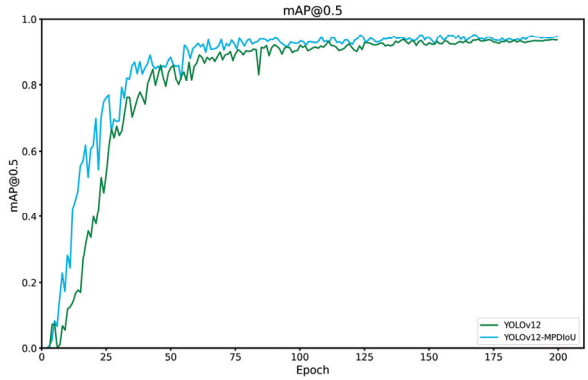
Figure 13. *Cont.*



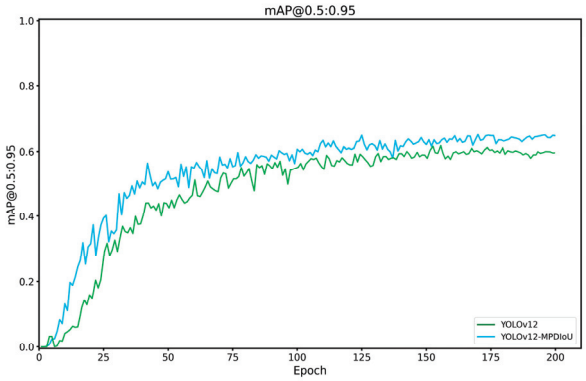
(**a2**) YOLOv12-LEGM mAP@0.5



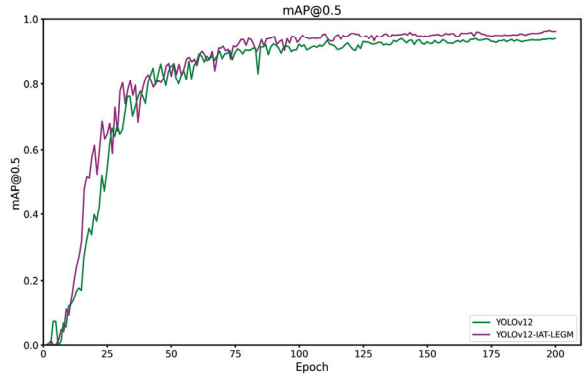
(**b2**) YOLOv12-LEGM mAP@0.5:0.95



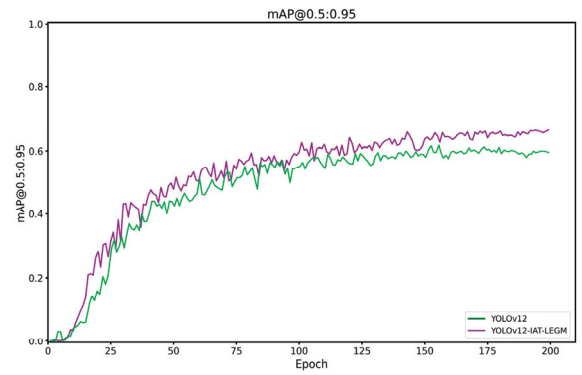
(**a3**) YOLOv12-MPDIoU mAP@0.5



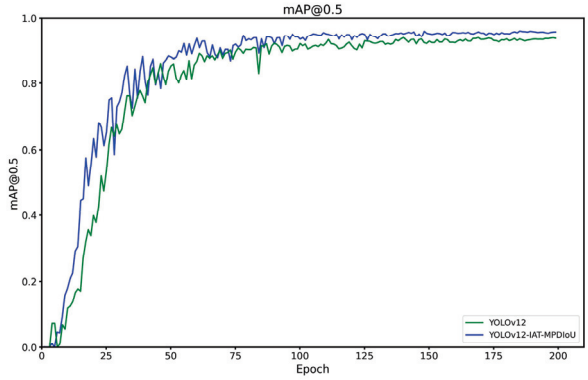
(**b3**) YOLOv12-MPDIoU mAP@0.5:0.95



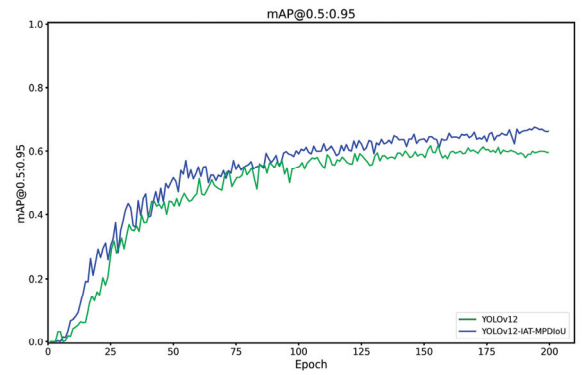
(**a4**) YOLOv12-IAT-LEGM mAP@0.5



(**b4**) YOLOv12-IAT-LEGM mAP@0.5:0.95

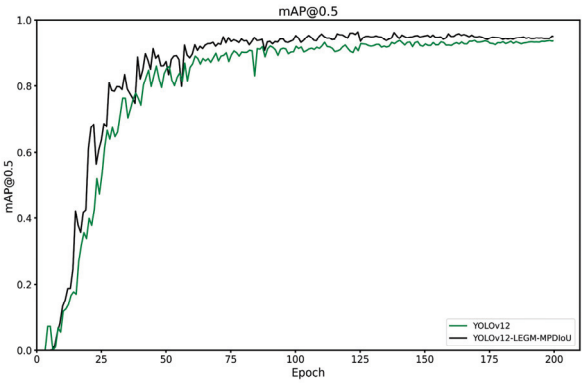


(**a5**) YOLOv12-IAT-MPDIoU mAP@0.5

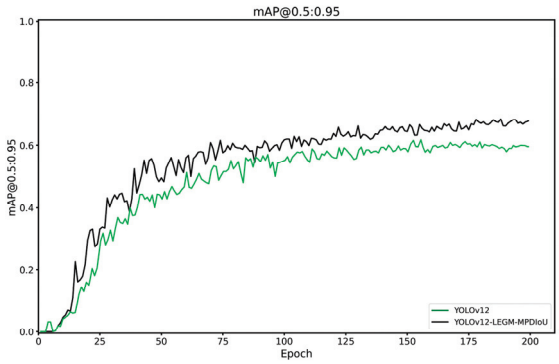


(**b5**) YOLOv12-IAT-MPDIoU mAP@0.5:0.95

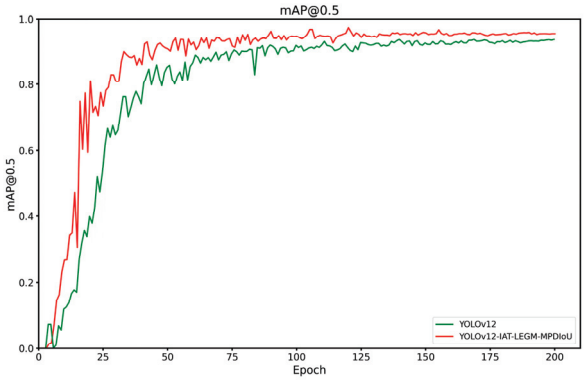
Figure 13. *Cont.*



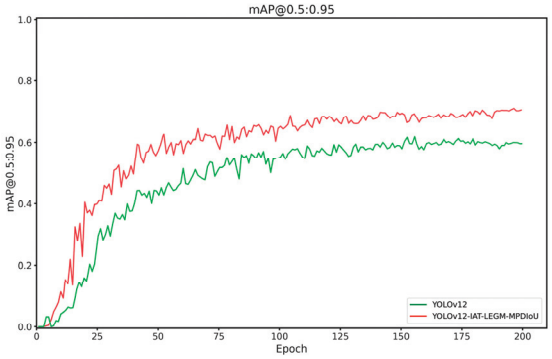
(**a6**) YOLOv12-LEGM-MPDIoU mAP@0.5



(**b6**) YOLOv12-LEGM-MPDIoU mAP@0.5:0.95



(**a7**) YOLO-ILM mAP@0.5



(**b7**) YOLO-ILM mAP@0.5:0.95

Figure 13. Performance Comparison of Ablation Experiments.

Table 5. Improvement in Each Module of the Ablation Experiments.

Module Combination	ΔPrecision(%)	ΔRecall (%)	ΔmAP@0.5 (%)	ΔmAP@0.5:0.95 (%)
Baseline	–	–	–	–
IAT	0.60	0.20	0.40	2.90
LEGM	1.30	0.40	0.60	4.60
IAT + LEGM + MPDIoU	1.10	0.50	0.20	3.80
IAT + LEGM	2.60	1.50	2.00	5.40
IAT + MPDIoU	2.70	2.40	1.40	6.00
LEGM + MPDIoU	3.30	1.90	1.10	7.30
IAT + LEGM + MPDIoU	4.50	3.10	1.90	8.60

4. Conclusions

This study addresses the issue of low target detection accuracy caused by poor image quality in low-light environments during copper ore waste rock processing. YOLO-ILM, a real-time detection algorithm for low-light copper ore waste, is proposed with core contributions including:

(1) Integrating the YOLOv12 network with the Illumination Adaptive Transformer module IAT to enhance model adaptability to low-light conditions;
(2) Introducing the Locally Embedded Global-feature attention mechanism LEGM to improve target feature capture capability under complex environmental interference;
(3) Introducing an efficient and accurate bounding box regression loss function MPDIoU to reduce computational load while optimizing localization accuracy.

This method satisfies real-time detection requirements for copper ore waste in low-light scenarios. The experimental results demonstrate that through comparison with

different detection algorithms, the proposed YOLO-ILM model shows significant advantages over Faster R-CNN, YOLOv8, YOLOv11, and YOLOv12 in both subjective visual detection effects and objective evaluation metrics. Compared to the baseline model, YOLO-ILM improves mAP@0.5 and mAP@0.5:0.95 by 1.9% and 8.6%, respectively, confirming the method's superiority for real-time copper ore waste detection under complex conditions.

Author Contributions: Conceptualization, J.D. and F.Q.; methodology, W.Z., J.X., L.Z. and Y.J.; validation, J.D. and W.Z.; investigation, J.X., L.Z. and Y.J.; data curation, J.D., W.Z. and J.X.; writing—original draft preparation, J.D., W.Z. and J.X.; writing—review and editing, F.Q., W.Z. and L.Z.; supervision, J.D. and F.Q.; project administration, J.D. and F.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China under Grant 2023YFC3904203.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in this article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Moreau, K.; Laamanen, C.; Bose, R.; Shang, H.; Scott, J.A. Environmental impact improvements due to introducing automation into underground copper mines. *Int. J. Min. Sci. Technol.* **2021**, *31*, 1159–1167. [CrossRef]
2. Tang, J.; Yang, H.; Chen, H.; Li, F.; Wang, C.; Liu, Q.; Zhang, Q.; Zhang, R.; Yu, L. Potential and future direction for copper resource exploration in China. *Green Smart Min. Eng.* **2024**, *1*, 127–131. [CrossRef]
3. Wang, Y.; Chen, Q.; Dai, B.; Wang, D. Guidance and review: Advancing mining technology for enhanced production and supply of strategic minerals in China. *Green Smart Min. Eng.* **2024**, *1*, 2–11. [CrossRef]
4. Liang, G.; Liang, Y.; Niu, D.; Shaheen, M. Balancing sustainability and innovation: The role of artificial intelligence in shaping mining practices for sustainable mining development. *Resour. Policy* **2024**, *90*, 104793. [CrossRef]
5. Shiryayeva, O.; Suleimenov, B.; Kulakova, Y. Sustainable Mineral Processing Technologies Using Hybrid Intelligent Algorithms. *Technologies* **2025**, *13*, 269. [CrossRef]
6. Quan, Z.; Sun, J. A Feature-Enhanced Small Object Detection Algorithm Based on Attention Mechanism. *Sensors* **2025**, *25*, 589. [CrossRef]
7. Korie, J.; Gabriela, C.-A.; Ezeonyema, C.; Oshim, F. Machine Learning Innovations for Improving Mineral Recovery and Processing: A Comprehensive Review. *Int. J. Econ. Environ. Geol.* **2024**, *15*, 26–31.
8. Corrigan, C.C.; Ikonnikova, S.A. A review of the use of AI in the mining industry: Insights and ethical considerations for multi-objective optimization. *Extr. Ind. Soc.* **2024**, *17*, 101440. [CrossRef]
9. Tao, M.; Lv, S.; Feng, S. Study on the Evaluation of the Development Efficiency of Smart Mine Construction and the Influencing Factors Based on the US-SBM Model. *Sustainability* **2023**, *15*, 5183. [CrossRef]
10. Shen, X.; Li, L.; Ma, Y.; Xu, S.; Liu, J.; Yang, Z.; Shi, Y. VLCIM: A vision-language cyclic interaction model for industrial defect detection. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 1–13. [CrossRef]
11. Liu, Y.; Wang, X.; Zhang, Z.; Deng, F. LOSN: Lightweight ore sorting networks for edge device environment. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106191. [CrossRef]
12. Shatwell, D.G.; Murray, V.; Barton, A. Real-time ore sorting using color and texture analysis. *Int. J. Min. Sci. Technol.* **2023**, *33*, 659–674. [CrossRef]
13. Tsangaratos, P.; Ilia, I.; Spanoudakis, N.; Karageorgiou, G.; Perraki, M. Machine learning approaches for real-time mineral classification and educational applications. *Appl. Sci.* **2025**, *15*, 1871. [CrossRef]
14. Hui, Y.; Wang, J.; Li, B. WSA-YOLO: Weak-supervised and Adaptive object detection in the low-light environment for YOLOV7. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–12. [CrossRef]
15. Zhou, Y.; Zhang, T.; Li, Z.; Qiu, J. Improved Space Object Detection Based on YOLO11. *Aerospace* **2025**, *12*, 568. [CrossRef]
16. Zhang, Q.; Guo, W.; Lin, M. LLD-YOLO: A multi-module network for robust vehicle detection in low-light conditions. *Signal Image Video Process.* **2025**, *19*, 271. [CrossRef]

17. Zhang, M.; Yue, K.; Li, B.; Guo, J.; Li, Y.; Gao, X. Single-Frame Infrared Small Target Detection via Gaussian Curvature Inspired Network. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–13. [CrossRef]
18. Wang, S.; Jiang, H.; Li, Z.; Yang, J.; Ma, X.; Chen, J.; Tang, X. PHSI-RTDETR: A Lightweight Infrared Small Target Detection Algorithm Based on UAV Aerial Photography. *Drones* **2024**, *8*, 240. [CrossRef]
19. Wang, Z.; Su, Y.; Kang, F.; Wang, L.; Lin, Y.; Wu, Q.; Li, H.; Cai, Z. PC-YOLO11s: A Lightweight and Effective Feature Extraction Method for Small Target Image Detection. *Sensors* **2025**, *25*, 348. [CrossRef] [PubMed]
20. Fang, W.; Chen, W. TBF-YOLOv8n: A Lightweight Tea Bud Detection Model Based on YOLOv8n Improvements. *Sensors* **2025**, *25*, 547. [CrossRef]
21. Peng, D.; Ding, W.; Zhen, T. A novel low light object detection method based on the YOLOv5 fusion feature enhancement. *Sci. Rep.* **2024**, *14*, 4486. [CrossRef]
22. Das, P.P.; Ganguly, T.; Chaudhuri, R.; Deb, S. YOLO-D: A Domain Adaptive approach towards low light object detection. *Procedia Comput. Sci.* **2025**, *258*, 3042–3051. [CrossRef]
23. Liu, Q.; Huang, W.; Hu, T.; Duan, X.; Yu, J.; Huang, J.; Wei, J. Efficient network architecture for target detection in challenging low-light environments. *Eng. Appl. Artif. Intell.* **2025**, *142*, 109967. [CrossRef]
24. Han, Z.; Yue, Z.; Liu, L. 3L-YOLO: A Lightweight Low-Light Object Detection Algorithm. *Appl. Sci.* **2025**, *15*, 90. [CrossRef]
25. Ye, S.; Huang, W.; Liu, W.; Chen, L.; Wang, X.; Zhong, X. YES: You should Examine Suspect cues for low-light object detection. *Comput. Vis. Image Underst.* **2025**, *251*, 104271. [CrossRef]
26. Qi, Z.; Wang, J.; Yang, G.; Wang, Y. Lightweight YOLOv8-Based Model for Weed Detection in Dryland Spring Wheat Fields. *Sustainability* **2025**, *17*, 6150. [CrossRef]
27. Zhao, Y.; Li, W.; Yang, R.; Liu, Y. Real-time efficient image enhancement in low-light condition with novel supervised deep learning pipeline. *Digit. Signal Process.* **2025**, *165*, 105342. [CrossRef]
28. Qiu, Z.; Huang, X.; Sun, Z.; Li, S.; Wang, J. GS-YOLO-Seg: A Lightweight Instance Segmentation Method for Low-Grade Graphite Ore Sorting Based on Improved YOLO11-Seg. *Sustainability* **2025**, *17*, 5663. [CrossRef]
29. Li, J.; Meng, X.; Wang, S.; Lu, Z.; Yu, H.; Qu, Z.; Wang, J. Enhanced Parallel Convolution Architecture YOLO Photovoltaic Panel Detection Model for Remote Sensing Images. *Sustainability* **2025**, *17*, 6476. [CrossRef]
30. Tian, Y.; Ye, Q.; Doermann, D. Yolov12: Attention-centric real-time object detectors. *arXiv* **2025**, arXiv:2502.12524.
31. Cui, Z.; Li, K.; Gu, L.; Su, S.; Gao, P.; Jiang, Z.; Qiao, Y.; Harada, T. You only need 90k parameters to adapt light: A light weight transformer for image enhancement and exposure correction. *arXiv* **2022**, arXiv:2205.14871. [CrossRef]
32. Zhang, Y.; Zhou, S.; Li, H. Depth information assisted collaborative mutual promotion network for single image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 2846–2855.
33. Ma, S.; Xu, Y. Mpdiou: A loss for efficient and accurate bounding box regression. *arXiv* **2023**, arXiv:2307.07662. [CrossRef]
34. Li, Y.; Fu, C.; Liu, T.; Hu, Z.; Song, G. Two-stage remaining useful life prediction method across operating conditions based on small samples and novel health indicators. *Reliab. Eng. Syst. Saf.* **2025**, *264*, 111290. [CrossRef]
35. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1780–1789.
36. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep retinex decomposition for low-light enhancement. *arXiv* **2018**, arXiv:1808.04560. [CrossRef]
37. Jha, R.; Lenka, A.; Ramanagopal, M.; Sankaranarayanan, A.; Mitra, K. RT-X Net: RGB-Thermal cross attention network for Low-Light Image Enhancement. *arXiv* **2025**, arXiv:2505.24705.
38. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Trans. Cybern.* **2022**, *52*, 8574–8586. [CrossRef]
39. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
40. Varghese, R.; Sambath, M. Yolov8: A novel object detection algorithm with enhanced performance and robustness. In Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 18–19 April 2024; pp. 1–6.
41. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

sensors

MDPI

Review

Comprehensive Review of Deep Learning Approaches for Single-Image Super-Resolution

Zirun Liu [1,2,†], Shijie Jiang [3,†], Shuhan Feng [3], Qirui Song [3] and Ji Zhang [1,3,*]

[1] Longmen Laboratory, Luoyang 471000, China; 23s134148@stu.hit.edu.cn
[2] School of Architecture and Design, Harbin Institute of Technology, Harbin 150001, China
[3] School of Information Engineering, Henan University of Science and Technology, Luoyang 471000, China; 9906685@haust.edu.cn (S.J.); 241404180104@stu.haust.edu.cn (S.F.); 241404180117@stu.haust.edu.cn (Q.S.)
* Correspondence: zhangji@haust.edu.cn
† These authors contributed equally to this work.

Abstract

Single-image super-resolution (SISR) is a core challenge in the field of image processing, aiming to overcome the physical limitations of imaging systems and improve their resolution. This article systematically introduces the SISR method based on deep learning, proposes a method-oriented classification framework, and explores it from three aspects: theoretical basis, technological evolution, and domain-specific applications. Firstly, the basic concepts, development trajectory, and practical value of SISR are introduced. Secondly, in-depth research is conducted on key technical components, including benchmark dataset construction, a multi-scale upsampling strategy, objective function optimization, and quality assessment indicators. Thirdly, some classic SISR model reconstruction results are listed and compared. Finally, the limitations of SISR research are pointed out, and some prospective research directions are proposed. This article provides a systematic knowledge framework for researchers and offers important reference value for the future development of SISR.

Keywords: single-image super-resolution; deep learning; image quality assessment

1. Introduction

Image super-resolution (SR), especially single-image super-resolution reconstruction (SISR), is an important research direction in the field of image processing. As shown in Figure 1, the core goal of SISR is to restore high-fidelity high-resolution (HR) images from degraded low-resolution (LR) inputs through advanced algorithm models. This technology has demonstrated significant value in multiple cutting-edge fields, such as image enhancement in security monitoring, medical imaging reconstruction, real-time video resolution enhancement, and precise segmentation tasks in computer vision.

Early research work focused on traditional interpolation-based methods, including bicubic interpolation and Lanczos interpolation [1]. However, due to certain limitations in the use of SISR, the phenomenon of different HR images corresponding to the same LR image always occurs. Subsequently, numerical methods such as edge-based methods [2] and image statistical models [3] have been proposed to address this deficiency. Meanwhile, with the continuous research on SISR, learning-based methods such as neighbor embedding [4] and sparse coding [5] have emerged to learn the mapping between LR and HR image blocks and improve the reconstruction quality.

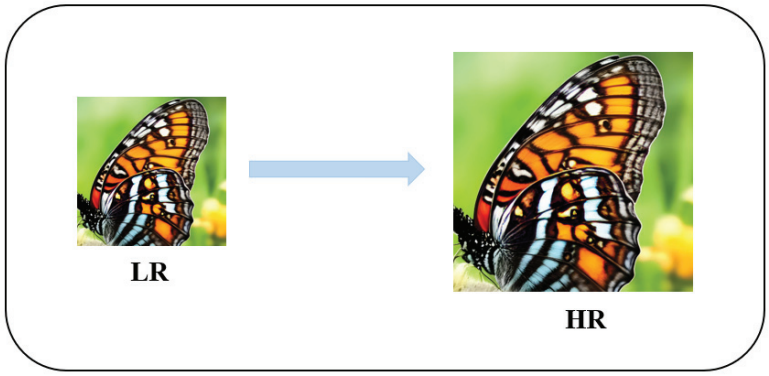


Figure 1. SISR aims to reconstruct a HR image from its degraded LR one.

With the emergence and development of deep learning (DL) [6], compared with other models, it has shown unique advantages and is in a dominant position in many artificial intelligence fields, such as computer vision [7], speech recognition [8], and natural language processing [9]. Based on this, SISR is also constantly developing. The end-to-end framework that can be implemented through DL integrates and unifies the three key steps in SISR tasks—feature extraction, nonlinear mapping, and image reconstruction. In this framework, the three main processes of SISR tasks can be easily implemented, while also helping to improve the performance of SISR and effectively reducing the cost of manual design and calculation. In addition, due to the efficient hardware and complex algorithm support of DL, in order to further promote the development of SISR, more and more researchers are focusing on SISR based on deep learning.

With the development of SISR, more and more SISR methods have been proposed. This article intends to provide a comprehensive review of SISR methods. At present, most existing comments are classified according to supervised methods, while there are relatively few classifications guided by other directions. This article adopts a goal-oriented classification method to classify deep learning-based SISR methods into three categories: "simulated SISR", "real-world SISR", and "domain-specific applications". This classification has wide applicability. It covers simulated scenarios, real-life scenarios, and applications in specific fields. This classification structure based on different objectives is clearer, allowing for a better understanding of the correlations between similar SISR models, making it more interpretable. For example, in the "Efficient Network Design Methods" section, 'Transformer-based methods" and "Mamba-based methods" are listed, which not only clearly demonstrates their technical relevance but also highlights the core objectives of these two methods in SISR. These two methods not only enable technological innovation in SISR but also strike a balance between efficiency and accuracy. In the "Specific Field Applications" classification section, SISR methods for specific fields, such as medical imaging and remote sensing imaging, are classified. This classification helps researchers to understand that methods in specific fields need to be combined with scene features, thereby enhancing the practical application of SISR models.

In this article, firstly, the key content of SISR is systematically integrated, and relevant work is analyzed in depth. Secondly, the SISR method is analyzed from the perspectives of simulating SISR, real-world SISR, and specific field applications, as shown in Figure 2. Finally, we discuss the key challenges in SR reconstruction and propose potential research directions by integrating the current development status. The main contributions of this review can be summarized in terms of four aspects:

(1) This article provides an overview of DL-based SISR methods and introduces various SR methods, elaborating in detail on the current status of SR method development;

(2) This article analyzes emerging SISR methods and lists various applications in specific fields, such as interdisciplinary SISR methods related to medical imaging and remote sensing;
(3) We compare the reconstruction results of some SISR models and compare their performance to provide an analysis that is simple and intuitive;
(4) We analyze the technical bottlenecks in the SISR field and explore the future development directions of SR technology by combining strategic frameworks such as cross-modal fusion.

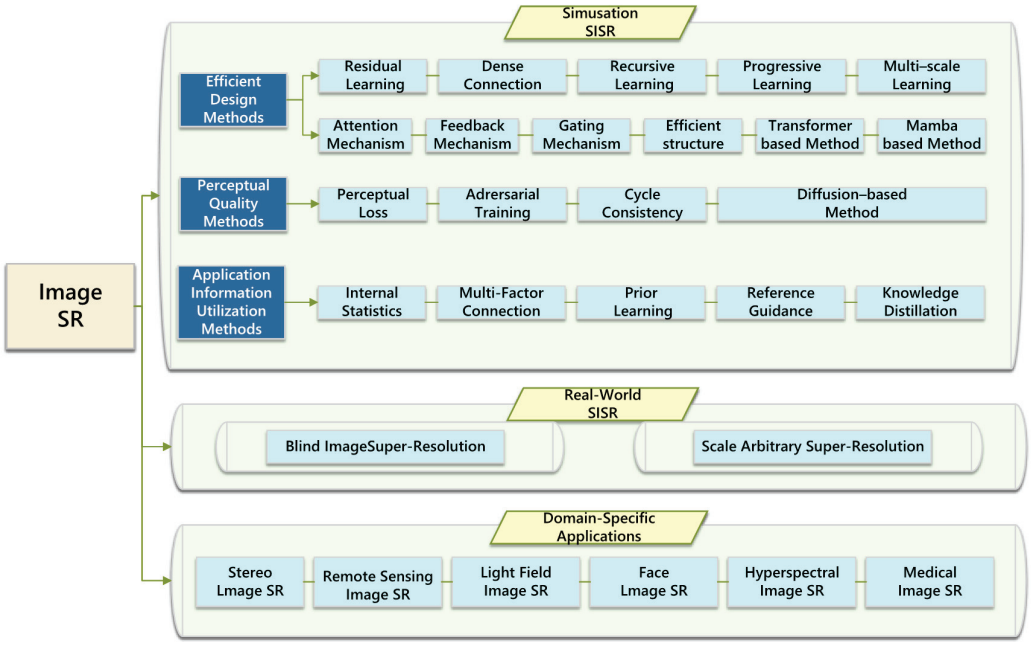


Figure 2. The basic SISR method classification in this review is mainly based on three types: simulated SISR, real-world SISR, and applications in specific fields.

2. Problem Statement and Related Works

2.1. Problem Statement

Image SR technology aims to improve the resolution of imaging systems, and it can be divided into two types: SISR and multi-image super-resolution (MISR). The core difference between these two types of image SR lies in the number of input images. Comparing the two, it can be found that SISR faces greater technical challenges because it relies entirely on the intrinsic information of a single frame for feature reconstruction, while MISR can utilize complementary information from multiple frames to assist in recovery.

Let the LR image be $I_x \in \mathbb{R}^{h \times w}$ and its corresponding ground-truth HR image be represented as $I_y \in \mathbb{R}^{H \times W}$, where the spatial resolution satisfies $H > h$ and $W > w$. In a typical SISR framework, LR image I_x is represented by the degradation process $I_x = \mathcal{D}(I_y; \theta_\mathcal{D})$, where $\mathcal{D} : \mathbb{R}^{H \times W} \to \mathbb{R}^{h \times w}$ defines the degradation function parameterized by θ_D. The most widely used formula in existing degradation models is

$$\mathcal{D}(I_y; \theta_\mathcal{D}) = (I_y \otimes k) \downarrow_s + n \tag{1}$$

Among them, LR images are usually generated from high-resolution HR images through blurring, downsampling, and noise interference. $I_y \otimes k$ corresponds to the convolution operation between the blur filter k and the HR image input I_y. $\downarrow_s$ represents the subsequent downsampling process with a scaling factor s, and n usually refers to additive white Gaussian noise (AWGN), characterized by the standard deviation σ.

For SISR, the main objective involves reconstructing SR image I_SR from its low-resolution counterpart I_x. This reconstruction problem is mathematically represented as $I_S = F(I_x; \theta_F)$, which defines the end-to-end SR model, and θ_F represents the trainable parameter set that controls the reconstruction process. In recent research, SISR has been redefined as an end-to-end learning task and relies on the unique advantages of convolutional neural networks (CNNs), further driving the development of DL-based models. Therefore, SISR can be mathematically expressed as the following optimization problem:

$$\hat{\theta}_{\mathcal{F}} = \arg\min_{\theta_{\mathcal{F}}} \mathcal{L}(I_{SR}, I_y) + \lambda\Phi(\theta), \tag{2}$$

where $\mathcal{L}$ represents the difference between the SR output I_{SR} and the ground-truth high-resolution image I_y, also known as the loss function. $\Phi(\theta)$ imposes constraints on the model parameters θ, and λ serves as a regularization hyperparameter that balances the trade-off between reconstruction fidelity and model complexity.

2.2. Benchmark Datasets

In deep learning-based SISR models, data are crucial in achieving excellent reconstruction performance (as shown in Figure 3). In recent years, both academia and industry have released numerous datasets applicable to SISR.

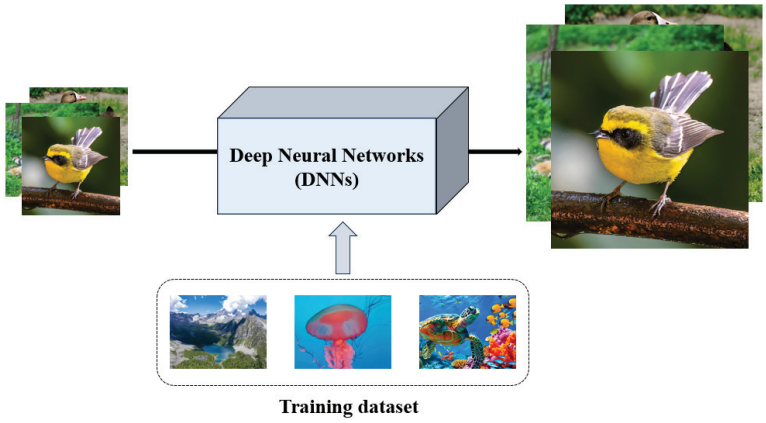


Figure 3. The training procedure for data-driven deep neural networks.

2.2.1. Degradation Mode

In research, SR datasets still face significant challenges in terms of scale, authenticity, and pairwise matching. In order to integrate data, the existing method is to degrade the dataset to obtain degraded images and then pair them with the original dataset to obtain paired data. In order to improve the authenticity of the dataset and efficiently train models, three commonly used degradation modes have been proposed, namely BI, BD, and DN. With the deepening of research, new degradation modes continue to emerge. For example, some studies consider more complex imaging processes and incorporate factors such as lens distortion and illumination changes into the degradation models.

2.2.2. Training and Test Datasets

In recent years, there have been many datasets used for SISR tasks, such as BSDS300 [10], DIV2K [11], and Flickr2K [12]. At the same time, there are also some test datasets, such as Set5 [13], Set14 [14], Urban100 [15], and Manga109 [16], which can be used to test the performance of the model. These datasets that can be used for SISR are shown in Table 1. Regarding these datasets, DIV2K [11] is widely used for training models, with a total of 1000 images, including 800 training images, 100 validation images,

and 100 test images. The Flickr2K [12] training dataset has a large scale, mainly consisting of 2650 images. Urban100 [15] belongs to the group of testing datasets and mainly consists of real-world images, with a total of 100 images.

Table 1. Some commonly used datasets.

Name	Usage	Amount	Format
BSDS300 [10]	Train	300	JPG
DIV2K [11]	Train	1000	PNG
Flickr2K [12]	Train	2650	PNG
Set5 [13]	Test	5	PNG
Set14 [14]	Test	14	PNG
Urban100 [15]	Test	100	PNG
Manga109 [16]	Test	109	PNG

Recently, new datasets have been proposed. In [17], the authors focus on SISR tasks in specific fields, such as medical image SR. Their dataset consists of three different low-resolution and high-resolution image pairs, promoting the application and success of scanning confocal microscopy in medicine. This provides strong support for the training of more professional and accurate SISR models.

2.3. Upsampling Methods

The framework of the SISR method consists of two parts: nonlinear mapping learning and an upsampling module. According to the different stages of upsampling, the framework of the method is divided into four components: pre-upsampling SR, post-upsampling SR, progressive upsampling SR, and iterative up-and-down sampling SR.

The pre-upsampling SR passes through the upsampling module and the nonlinear mapping learning module successively. It avoids nonlinear mapping learning in the low-dimensional space, but it will generate large amounts of blurring and noise, and it has a high cost. Post-upsampling SR is the opposite to pre-upsampling SR. It first performs nonlinear mapping learning in the low-dimensional space and then passes through the upsampling module, greatly improving the efficiency. Progressive upsampling SR combines multiple modules and gradually enlarges the image. It is powerful, but the design is complex and the training is difficult. Iterative up-and-down sampling SR repeatedly performs LR-HR mapping learning. It is complex in design and not very practical. This section mainly explains common upsampling methods.

2.3.1. Based on Interpolation Upsampling Methods

Interpolation is the most commonly used upsampling method, and there are three main methods: nearest-neighbor interpolation, bilinear interpolation, and bicubic interpolation.

Nearest-neighbor interpolation is the simplest interpolation algorithm. It selects the pixel value located at the nearest point to the insertion pixel as the pixel value of the insertion pixel. It simply enlarges the image, and there is no gradual change process between blocks. Its advantage is that the algorithm is simple and fast, but it will generate block effects and obvious visible jagged edges.

Bilinear interpolation involves performing linear interpolation in both the horizontal and vertical directions. Bilinear interpolation needs to find the four nearest pixels around the pixel point to be calculated and calculate the pixel value of the pixel point based on their pixel values and the distance between them. As shown in the Figure 4, in the horizontal direction, two linear interpolations are performed to calculate the pixel values of R0 and R1;

then, in the vertical direction, based on the pixel values of R0 and R1, linear interpolation is performed to calculate the pixel value of point P. The formula is as follows:

$$f(P) = \frac{(x_2 - x)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} f(Q_{11}) + \frac{(x - x_1)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} f(Q_{21}) + \frac{(x_2 - x)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} f(Q_{12}) + \frac{(x - x_1)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} f(Q_{22}) \quad (3)$$

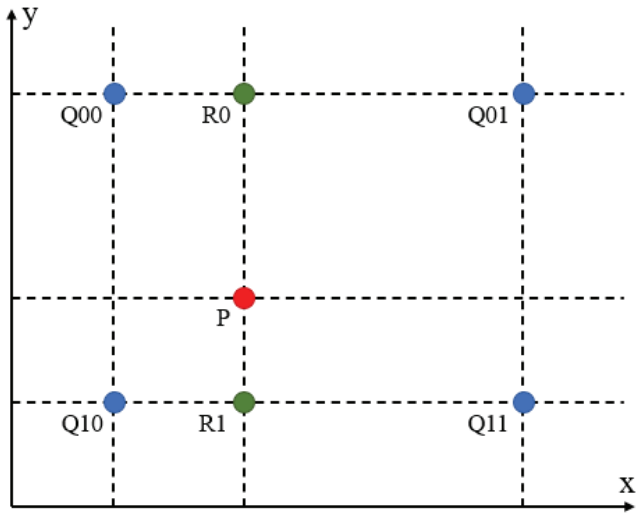


Figure 4. Bilinear interpolation.

Bicubic interpolation is a more complex algorithm based on polynomial interpolation. It calculates the pixel value based on 16 pixels around the pixel point to be calculated. It further considers the influence of the change rate of gray values of adjacent pixels. Compared with the previous two interpolation methods, bicubic interpolation can generate smoother and more detail-retained images, but it is relatively complex and slow.

2.3.2. Based on Learning Upsampling Methods

Unlike interpolation methods, which are fixed, frameworks based on learning methods are trained autonomously and can be continuously adjusted to find suitable convolution kernels.

Transposed convolution layers, based on the size of the output pixel and the original pixel, insert zeros in the original pixels around and between (pixels between can also not be inserted) by setting different strides, paddings, and kernels. The relationship between the set stride (s), padding (p), and kernel (k) and the insertion of zeros is as follows: fill s-1 rows and columns with zeros between the elements of the feature map, and fill k-p-1 rows and columns with zeros around the feature map, as shown in Figure 5. Then, through convolution, HR is obtained. However, the transposed convolution layer has a "non-uniform overlap effect", which will cause the generated image to have a checkerboard texture.

Sub-pixel convolutional layers, as shown in Figure 6, use multiple convolution kernels to perform convolution on the input feature map. First, through convolution, the input feature map of size H × W × C is changed to a feature map of size H × W × $(C * r^2)$. Through a PixelShuffle operation, the r^2 channels of each pixel point are rearranged to form an r × r-sized region. Then, each pixel point formed region is combined to output a feature map of Hr × Wr × C. The sub-pixel convolution layer has high computational efficiency and resolves the drawback of transposed convolution generating checkerboard textures,

but there will be unnatural textures between small regions, especially for images with severely lost information, and the recovery effect of sub-pixel convolution layers is limited.

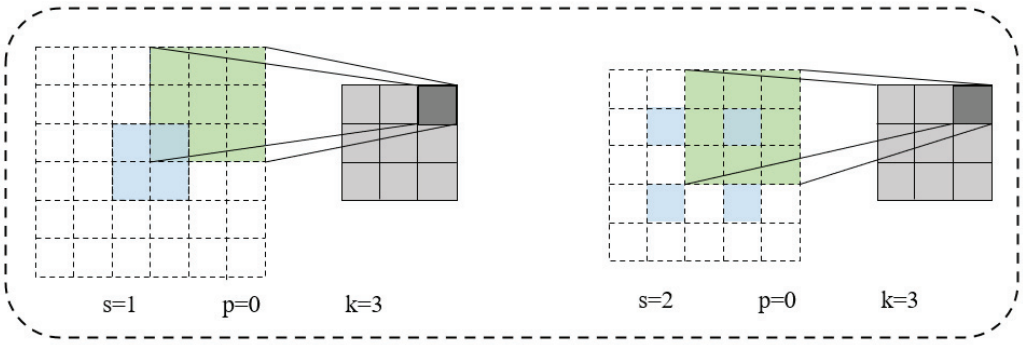


Figure 5. Transposed convolutional layer.

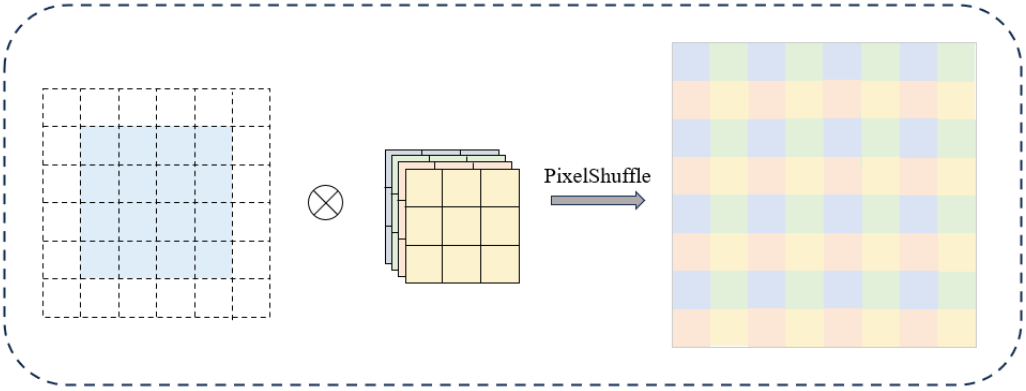


Figure 6. Sub-pixel convolutional layer.

2.4. Optimization Objective

Evaluation and parameter updates are critical steps for all DL-based models. In this section, we will detail the necessary procedures during model training, including the design of learning strategies and loss functions.

2.4.1. Learning Strategy

In SISR tasks, the choice of learning strategy is crucial for model performance. Based on whether paired low-resolution (LR) and high-resolution (HR) images are used for training, SISR models are mainly divided into two categories: supervised learning methods and unsupervised learning methods.

Supervised Learning. In supervised learning, models are trained using paired LR-HR images. In simulated SISR tasks, LR images are obtained by downsampling HR images, while, in real-world SISR tasks, LR and HR image pairs are acquired by adjusting the camera zoom. The core goal of supervised learning is to minimize the error between the reconstructed image and the ground-truth HR image. The optimization objective of supervised learning can be expressed as

$$\theta_f = \arg\min_{\theta} \mathcal{L}(I_{SR}, I_{HR}) \tag{4}$$

where $\mathcal{L}$ is the loss function, and the reasonable selection of the loss function is crucial for model performance.

Unsupervised Learning. Due to the difficulty of obtaining paired LR-HR images, unsupervised learning methods have gradually gained attention. These methods do

not require paired LR-HR images but instead use unpaired LR images or self-generated HR images for training. Common unsupervised learning methods include generative adversarial network (GAN)-based methods and self-supervised learning methods. Relevant examples include zero-shot SR (ZSSR) [18], which constructs training data using the test image and its downscaled versions, optimizing the model through data augmentation and loss functions, and CycleGAN [19], which learns bidirectional mappings from LR to HR and HR to LR to achieve unsupervised image SR.

The optimization objectives of unsupervised learning typically involve adversarial loss, cycle consistency loss, and so on, to generate visually realistic HR images. The advantage of unsupervised learning lies in its ability to better adapt to real-world scenarios, although its performance often depends on the accuracy of the degradation model.

2.4.2. Loss Function

The loss function plays a guiding role in optimizing models in SISR tasks. A single loss function often fails to comprehensively reflect the quality of image reconstruction, so researchers usually combine multiple loss functions to guide model optimization, thereby better reflecting the image restoration effect. Below are several commonly used loss functions.

Pixel Loss. Pixel loss is the most basic loss function, used to measure the pixel-level differences between the reconstructed image and the HR image. Common pixel losses include the L1 loss, mean squared error (MSE) loss, and Charbonnier loss. Among them, the L1 loss calculates the absolute difference in pixel values but is less sensitive to outliers, while the MSE loss calculates the mean squared difference in pixel values and is sensitive to outliers. The Charbonnier loss is a differentiable variant of the L1 loss with better numerical stability, often used for noise handling. The pixel loss can be expressed as follows:

$$\mathcal{L}_{\text{pixel}} = \frac{1}{HWC} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{k=1}^{C} |I_{\text{SR}}(i,j,k) - I_{\text{HR}}(i,j,k)| \tag{5}$$

where H, W, and C are the height, width, and the number of channels of the image. The pixel loss is still widely adopted. However, this type of loss function has certain limitations. The reconstructed images lose certain high-frequency details, such as subtle changes in texture and clear contours of edges, resulting in a lack of layered perception.

Adversarial Loss. Adversarial loss is introduced through generative adversarial networks (GANs) [20] to generate more realistic HR images. A GAN consists of a generator and a discriminator, where the generator is responsible for generating fake samples, and the discriminator evaluates the authenticity of the samples. For example, SRGAN [21] proposes a discriminator loss function based on cross-entropy. The adversarial Loss can be expressed as follows:

$$\mathcal{L}_{adv} = \mathbb{E}_{I_{HR}}[\log D(I_{HR})] + \mathbb{E}_{I_{LR}}[\log(1 - D(G(I_{LR})))] \tag{6}$$

where $\mathbb{E}$ represents the mathematical expectation, G stands for the generator, and D denotes the discriminator.

Fourier Space Loss. The Fourier space loss transforms the image into the frequency domain using the fast Fourier transform (FFT) [22] and calculates the differences in the magnitude and phase of frequency components. This loss function can better handle high-frequency information in images, while enhancing the details of reconstructed images. The Fourier space loss can be expressed as follows:

$$\mathcal{L}_{\text{fourier}} = \frac{1}{HW} \sum_{u=1}^{H} \sum_{v=1}^{W} \left| Y_{u,v}^{SR} - Y_{u,v}^{HR} \right| + \angle\left(Y_{u,v}^{SR}, Y_{u,v}^{HL} \right) \tag{7}$$

where $(Y_{u,v})$ represents the frequency spectrum of the image.

Prior Loss. Prior loss guides model optimization by incorporating prior knowledge of the image, such as sparsity, gradients, and edges. Common prior losses include the gradient prior loss and edge prior loss. The prior loss can be expressed as follows:

$$\mathcal{L}_{\text{prior}} = \sum_{i=1}^{N} \lambda_i \mathcal{L}_{\text{prior}_i}(I_{SR}, I_{HR}) \tag{8}$$

where N represents the number of prior losses, and λ_i represents the weight coefficient of the ith prior loss.

Mixed Loss. To balance the quality, detail, and visual perception of the generated images, researchers often combine multiple loss functions. Below, several commonly used combinations of loss functions are listed.

L1+ perceptual loss combines the pixel-level loss and content loss to generate images with rich details; L1+TV loss combines the pixel-level loss and total variation (TV) loss to reduce blocky artifacts in images; content loss+adaptive loss combines the content loss and adaptive weight loss to generate images with better visual consistency.

Combining the commonly used loss functions mentioned above, such as the pixel loss, content loss, adversarial loss, Fourier space loss, and prior loss, the mixed loss function can be expressed as follows:

$$\mathcal{L}_{\text{mixed}} = \lambda_1 \mathcal{L}_{\text{pixel}} + \lambda_2 \mathcal{L}_{\text{content}} + \lambda_3 \mathcal{L}_{\text{adv}} + \lambda_4 \mathcal{L}_{\text{prior}} \tag{9}$$

where λ_1, λ_2, λ_3, λ_4 represent different weight coefficients, which are used to balance the importance of various loss terms. Usually, they need to be adjusted to appropriate values through experiments.

In recent years, with the continuous development of deep learning techniques, the design of loss functions in SISR tasks has also evolved. The latest advancements in SISR technology have opened up new paths for its future development. Below, we list some of the latest research advancements.

Multi-Scale Loss: By calculating loss functions at different scales, one can better capture global and local information from the image, improving the quality of reconstructed images. Adaptive Loss Weighting: This dynamically adjusts the weights of different loss functions to better balance image quality and detail. Self-Supervised Loss: This utilizes unpaired LR and HR images for training through self-supervised learning, reducing the reliance on paired data. Frequency Domain Loss: Loss functions incorporating frequency-domain information (for example, Fourier transform, wavelet transform) can better handle high-frequency details in images.

These latest loss function design methods have achieved significant performance improvements. In the SISR task, especially when dealing with complex scenes and low-quality input images, the design of learning strategies and loss functions has a significant impact on model performance. In addition, based on the latest loss function design methods, it is also possible to combine multiple loss functions to more effectively balance image quality, detail, and visual perception.

2.5. Assessment Methods

2.5.1. Objective Evaluation Methods

Objective evaluation refers to analyzing the objective features of an image through certain algorithms to obtain relevant quality scores. It is relatively low-cost, fast, and straightforward. At present, the two main objective quality evaluation indices are the peak signal-to-noise ratio and structural similarity.

Peak signal-to-noise ratio. The peak signal-to-noise ratio (PSNR) is a commonly used objective criterion for evaluating image quality. It first needs to calculate the mean square error, i.e., calculate the difference between the individual pixels, and then average these differences to obtain the mean square error. Finally, by incorporating this mean square error into the ratio of the peak signal, and through a logarithmic transformation, an intuitive number is obtained, which directly reflects the quality of the image. The mathematical expression for the PSNR is as follows:

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{10}$$

$$\mathrm{MSE}(X,Y) = \frac{1}{W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H}\left(X(i,j) - Y(i,j)\right) \tag{11}$$

where MSE is the mean square error of each pixel in the original image X and the reconstructed image Y, MAX is the maximum possible pixel value of the image, and H and W represent the height and width of the image, respectively. $PSNR$ is measured in decibels (dB), and the magnitude of the value reflects the quality of the image. When the $PSNR$ value is above 40 dB, the quality of the image is close to that of the original high-resolution image, and the image quality is very high. When the $PSNR$ value is at 30–40 dB, the quality of the image is good, and the degree of distortion of the image can be observed. When the $PSNR$ value is 20–30 dB, it indicates that the image has a high degree of distortion and poor image quality. When the PSNR value is below 20, the quality of the image is very poor and the image information loss is serious.

The peak signal-to-noise ratio (PSNR) is a widely used and easy-to-understand image quality evaluation metric. By calculating the mean square error between the original image and the reconstructed image, the PSNR can quantify the similarity between them. Higher values indicate better image quality, which makes it possible to visually compare the quality of different image processing methods. The advantage of the PSNR is that the calculation process is simple and clear; it is easy to understand and explain. This intuitive and quantitative evaluation method has led to the widespread acceptance and use of the PSNR in the fields of image processing and communication. However, it is important to note that the PSNR, while simple and easy to understand, does not always fully reflect the perception of image quality by the human visual system, so, in some cases, it is necessary to combine other evaluation indicators to comprehensively evaluate the image quality.

Structural similarity. The structural similarity index (SSIM) [23] is a measure of both original and reconstructed images. Unlike the PSNR, the SSIM is a similarity index designed to more comprehensively consider human perceptions of image quality. Focusing only on luminance information, the contrast and structural information of the image is also considered. The calculation of the SSIM is based on brightness and contrast. The brightness part measures the average brightness of the image, the contrast part considers the range of brightness variation, and the structure part focuses on the texture and structure information of the image. By combining these three aspects, the SSIM can provide more granular and comprehensive information in image quality assessments. The formula for calculating the SSIM based on brightness, contrast, and structure is

$$l(I_{SR}, I_{HR}) = \frac{2\mu_{SR}\mu_{HR} + c_1}{\mu_{SR}^2 + \mu_{HR}^2 + c_1} \tag{12}$$

$$c(I_{SR}, I_{HR}) = \frac{2\sigma_{SR}\sigma_{HR} + c_2}{\sigma_{SR}^2 + \sigma_{HR}^2 + c_2} \tag{13}$$

$$s(I_{SR}, I_{HR}) = \frac{\sigma_{SRHR} + c_3}{\sigma_{SR}\sigma_{HR} + c_3} \tag{14}$$

$$\begin{aligned} \mathrm{SSIM}(I_{SR}, I_{HR}) &= [l(I_{SR}, I_{HR})]^{\alpha}[c(I_{SR}, I_{HR})]^{\beta}[s(I_{SR}, I_{HR})]^{\gamma} \\ &= \frac{(2\mu_{SR}\mu_{HR} + c_1)(\sigma_{SRHR} + c_2)}{(\mu_{SR}^2 + \mu_{HR}^2 + c_1)(\sigma_{SR}^2 + \sigma_{HR}^2 + c_2)} \end{aligned} \tag{15}$$

where I_{SR} and I_{HR} represent SR and HR images, respectively; μ_{SR} and μ_{HR} represent the mean pixels of the SR and HR images; σ_{SR} and σ_{HR} represent standard deviations; σ_{SRHR} represents the covariance of the SR and HR images; c_1, c_2, and c_3 are constants to avoid errors caused by a zero denominator. $SSIM$ can range from zero to one, where a value of 1 indicates that the two images are identical, and a value close to 0 indicates that they are very different. Therefore, the closer the $SSIM$ value is to 1, the more similar the reconstructed image is to the original image, and the higher the image quality. Conversely, the closer the value is to 0, the more significant the differences between the images and the lower the quality.

2.5.2. Subjective Evaluation Methods

In addition to objective evaluation methods, subjective evaluation methods also play a crucial role in image quality evaluation. Although the PSNR and SSIM provide numerical measures, they cannot fully simulate the perception of images by the human visual system. Human subjective perception is influenced by many factors, including cognition, subjective preferences, and esthetics. Therefore, the subjective evaluation method provides an assessment of image quality that more closely reflects real perception by inviting observers to perform actual observations and evaluations. A commonly used subjective evaluation criterion is the mean opinion score (MOS), which is obtained by inviting a group of observers to evaluate an image, and their scores are averaged to obtain a composite score. These scores, which typically range from 1 to 5, represent the observer's subjective perception of image quality, with higher scores indicating better image quality and lower scores indicating worse quality, as shown in Table 2.

Table 2. Subjective assessment indicator rating scale.

Quality Score	Degree of Image Distortion	Image Quality
1	The level of distortion is so great that vision is hindered	Very poor
2	Noticeable distortion can be observed in the image	Poor
3	Distortion is observable in the image	Ordinary
4	The degree of distortion is low	Good
5	Visual distortion is not easily noticeable	Excellent

2.5.3. Fine-Grained Error Analyses

Traditional indicators cannot reveal specific behavioral differences in texture, edges, and other aspects during the model image reconstruction process. Researchers call for fine-grained error analysis to further analyze SISR models. In this way, one can devise a classification method for common artifacts such as texture distortion, edge blurring, noise amplification, and false detail generation. Then, experts annotate the main artifacts in the SR output on the annotated benchmark dataset. This method can use confusion matrices to quantify SISR models. For example, the confusion matrix may display the confusion between texture artifacts and edge artifacts, thereby revealing the limitations of the model, considering aspects that were not reflected in the overall score. This can provide

precise guidance for improving models and help to solve specific problems, such as texture restoration and edge sharpness.

2.6. Ablation Studies for Component Validation

To quantify the contributions of key components in the proposed SISR framework, we conducted ablation experiments using the SRCNN [7] model as an example. This included dataset construction strategies, upsampling modules, loss function design, and evaluation methods.

Specifically, we adjusted the dataset construction strategy, upsampling method, loss function, and evaluation method separately. All experiments were conducted under the same training (T91 dataset) and testing (Set5 dataset, $\times 2$) conditions. The results showed that, when adjusting the dataset construction strategy or changing the degradation kernel, its performance significantly decreased. For the upsampling method, the performance slightly decreases after transpose convolution. For the loss function, using the L1 loss can allow an improvement in performance. Overall, for SRCNN, the dataset construction strategy is the most critical, and the loss function design has a significant impact on the image quality. Upsampling methods are related to efficiency and performance, and diverse evaluation methods can comprehensively reflect changes in model performance. This experimental framework can provide a reference for other models, but different models have different characteristics and they still need to be analyzed to improve model performance.

3. Image Super-Resolution

In 2014, the SR Convolutional Neural Network (SRCNN) proposed by Dong et al. [7] was of milestone significance in the field of SISR. As the first SISR model based on a convolutional neural network (CNN), SRCNN revealed the equivalence between the deep CNN architecture and the traditional example-based learning method based on sparse coding in the mathematical framework through theoretical analysis, successfully introducing deep learning into the field of SR research. This groundbreaking work promoted a shift in the research paradigm of SISR. Subsequent studies gradually framed the SR task as an end-to-end deep learning problem, directly establishing a nonlinear mapping relationship between LR images and HR images through deep neural networks. Since then, CNN-based SR methods have entered a stage of rapid development. Innovative techniques such as residual learning and attention mechanisms have been successively introduced, continuously surpassing performance records in objective indicators such as the PSNR and SSIM, as well as in visual perception quality, and continuously driving technological progress in this field.

3.1. Simulation SISR

In recent years, SISR reconstruction technology has been developing rapidly, with many innovative algorithm models emerging. It is worth noting that the current mainstream research methods generally adopt the training and validation paradigm based on simulated datasets, which is defined as the "simulated SISR" technical path in the academic community. The core feature of this technical route is to degrade HR images through a preset fixed degradation model (such as bicubic downsampling) and then generate paired LR images as training data. However, the difference between these simulated data and the actual scenario may lead to a performance decline in the model in real-world applications.

It should be particularly pointed out that, despite the above limitations, the simulated SISR method has still had a profound impact on the development of the field. According to the differences in technical routes, existing methods can be systematically divided into three research branches: innovation in efficient network architectures and computational mechanisms, perceptual quality-oriented methods, and multi-modal information fusion

technology. These three research directions complement each other and jointly promote theoretical breakthroughs and application expansions in SISR technology.

3.1.1. Efficient Network/Mechanism Design Methods

The techniques that have emerged in the last few years have focused on the efficient and precise design of network architectures and mechanisms, which can be modeled to achieve better performance with fewer parameters. Therefore, the choice of an appropriate research methodology is particularly crucial when studying the dynamical behavior of nonlinear systems.

Residual learning In the development and research of SR convolutional neural networks, it is known that increasing the number of layers in the convolutional neural network can improve the accuracy of model training and thus lead to better results. However, directly stacking too many layers may lead to problems such as "gradient vanishing" and "gradient explosion" [24].

In ResNet [25], He et al. introduced a residual learning framework designed to learn a residual mapping instead of directly approximating the entire underlying function (Figure 7). In SISR, there is a lot of common information between low-resolution (LR) images and high-resolution (HR) images, making it easy to model the relationships between high-resolution and low-resolution images directly. The advantages of residual learning are obvious. It not only allows the number of layers in the neural network to increase but also alleviates problems such as gradient vanishing and network degradation. Kim et al. [26] proposed another deep SR network, VDSR. With the in-depth development of deep learning, to make the network design simpler, residual blocks have gradually become the basic components for building network structures. For example, a convolutional branch usually contains two 3×3 convolutional layers, two batch normalization layers, and a ReLU activation function in the middle. However, Lim et al. [27] found through research that, in SISR tasks, batch normalization layers consume a lot of memory but do not significantly improve model performance, so batch normalization layers are generally removed from the network structure. In addition, Lu et al. [28] proposed MARNet, which combines a residual network structure and attention mechanism. It can effectively extract image information and fuse multi-layer features, further improving the reconstruction accuracy.

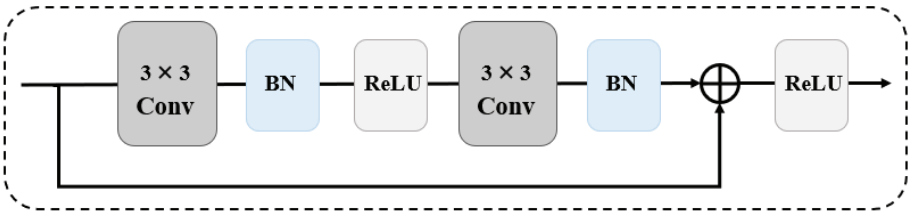


Figure 7. Sketch of residual learning architecture/residual block.

Global and local residual learning: Global residual learning optimizes the learning process of the entire network by establishing a connection between the input and the final reconstruction layer, improving information transmission, and it can improve the network performance and reduce information loss. However, residual networks usually contain parameterized layers, making the network model more complex and potentially causing the loss of image details. Therefore, local residual learning has been proposed. Local residual learning uses residual connections at the local level rather than from input to output. This multi-path learning method can carry rich image details and enhance the network's expressive ability. Of course, local residual learning can also be applied in the feature extraction module to enhance the learning ability of the module [29,30]. In modern

times, combining global residual learning and local residual learning is highly applicable. Using both can effectively improve the effectiveness of the model.

Residual scaling: Lim et al. [27] proposed an enhanced deep super-resolution network (EDSR) that can improve the performance by expanding the model size, and they also proposed a new multi-scale deep super-resolution system (MDSR) that can reconstruct high-resolution images with different magnification factors in a single model. In GAMSRN [31], this method utilizes global attention multi-scale residual blocks. This method can not only extract features at different scales but also improve the performance without increasing the model complexity.

Dense connection: This connection method is widely used in dense connection convolutional networks and in neural networks. It enables each layer in the dense block to be connected to all previous layers, thereby obtaining the features of the previous layers. The short paths established between these layers can alleviate gradient vanishing and gradient explosion and enhance the information exchange between these layers, further improving the precision of reconstruction.

In the field of deep learning image reconstruction, Tong et al. [32] proposed SR-DenseNet. SR-DenseNet not only employs hierarchical dense connections to enable the exchange and integration of features at different levels but also introduces block-level dense connections, ensuring that the outputs of each dense block are interconnected through dense connections (Figure 8). This design allows for the effective combination and utilization of low-level and high-level features during the image reconstruction process, thereby enhancing the reconstruction effect. Kuldeep et al. [33] proposed a hybrid dense connection block (MDCB) that combines the advantages of residuals and dense connections, which can achieve the SR of multiple factors, improve the performance, and achieve high factor parameter efficiency. Besides SR-DenseNet and MDCB, the dense connection mechanism has been widely applied in other models, such as MemNet [34], RPMNet [35], and MFNet [36]. By using the dense connection mechanism, the information flow between different depths of the network can be efficiently facilitated and utilized, allowing the model to fully exploit data features and resulting in superior image reconstruction performance.

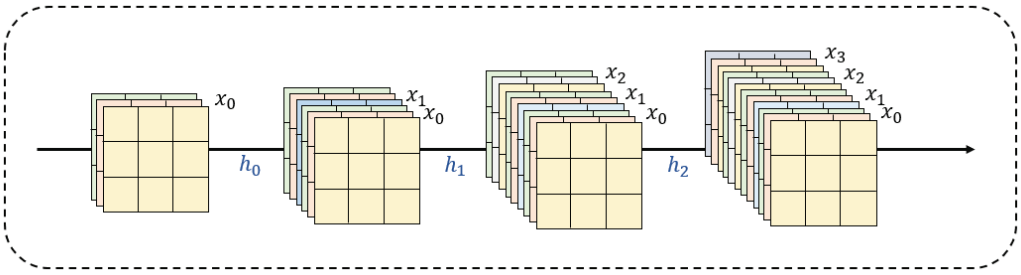


Figure 8. Dense connection module.

Recursive learning: Recursive learning methods involve repeatedly applying the same submodules in the same model and using the same data. Among them, each recursive module is composed of many recursive units, which have the same structure and data. However, during the use of recursive learning methods to construct models, if there are many stacked layers, problems such as vanishing or exploding gradients can easily occur. To address this issue, a deep recursive residual network based on residual learning can be constructed to construct recursive blocks. In DSRN [37], it is possible to enhance sequence data processing by introducing a two-state mechanism, a structure that allows the model to maintain two different internal states within the same time step, thus better capturing

long-term dependencies and improving the performance. In recent years, an increasing number of models have adopted the idea of residual learning in recursive units, such as MDCB [33], MemNet [34], CARN [38], and SRRFN [39].

Progressive learning: Progressive learning is a strategy that aims to help learners or models to adapt and develop better by gradually increasing the difficulty of a task. It shows excellent performance when dealing with complex sequence prediction and decision-making tasks. It not only shortens the training cycle but also significantly enhances the generalization ability of the model. SISR is a complex and unstable problem, with challenges such as large-scale factorization, undetermined degradation kernels, insufficient data, and noise interference. By adopting an incremental learning strategy, we can split the SISR task into a series of shallow to deep subtasks, starting from simple small-scale problems and gradually increasing their difficulty, which can enhance the performance and generalization ability of the model during continuous challenges and thus effectively simplify the learning process and improve the learning efficiency.

In PRAN [40], it is shown that introducing progressive residual attention for optimization can improve the performance. In the design of ProSR [41], the layers of the pyramid were gradually blended to minimize the impact on the previously trained layers by gradually introducing training pairs for various scales. Wong et al. [42] proposed a progressive adversarial network (PAN) that utilizes U-Net and a neural structure for progressive growth; this can enhance the texture details of U-Net while ensuring the accuracy of image reconstruction. Through progressive learning, the complex problem is decomposed into several smaller subtasks, so that it is possible start from simple problems and gradually accumulate experience, thus accelerating the convergence of the algorithm and showing better results in the reconstruction stage.

Multi-scale learning: Enriched image features are a key factor in achieving high-quality image SR reconstruction. In recent years, many research works [40,43,44] have pointed out that images exhibit diverse features in different scenes and conditions. Therefore, fully exploiting and utilizing these diverse features can significantly improve the performance of the model and enhance the effects of image reconstruction. Li et al. [29] proposed a multi-scale residual block (MSRN) for feature extraction; it introduces convolution kernels of different sizes on top of the residual block and adaptively detects image features of different scales while letting these features interact with each other to obtain the most effective image information (Figure 9). After this, Li et al. [45] further optimized the structure and proposed the multi-scale dense cross-network (MDCN); this can maximize the utilization of image feature flows at different scales to improve the performance.

In recent years, multi-scale SISR models have been proposed one after another. For example, Qin et al. [46] proposed a multi-scale feature fusion residual network (MSFFRN) to better exploit the image features in SISR. Chang et al. [47] organically integrated multi-scale learning and dense connectivity techniques to form a novel dual-scale dense network (MSDN). Cao et al. [48] proposed a novel multi-scale residual channel attention network (MSRCAN) and introduced the channel attention mechanism (CAM, Figure 10) in MSRCAN. Yazıcı et al. [49] proposed a multi-scale lightweight hybrid network (GLIMS), which combines the bottleneck structure of Swin Transformer with attention guidance and multi-scale feature aggregation to effectively capture local and global features. Li et al. [50] proposed the remote multi-scale fusion network (LMFN). This network is capable of discovering multi-scale priors in images and utilizing interactive fusion modulation modules to fuse multi-scale features. These examples also show that effectively extracting and utilizing the features of multi-scale images is the key to further enhancing the reconstructed images.

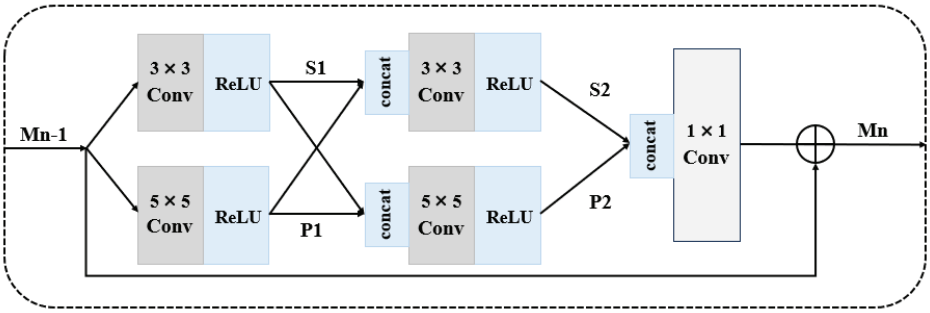


Figure 9. Structure of multi-scale residual block (MSRB) [29].

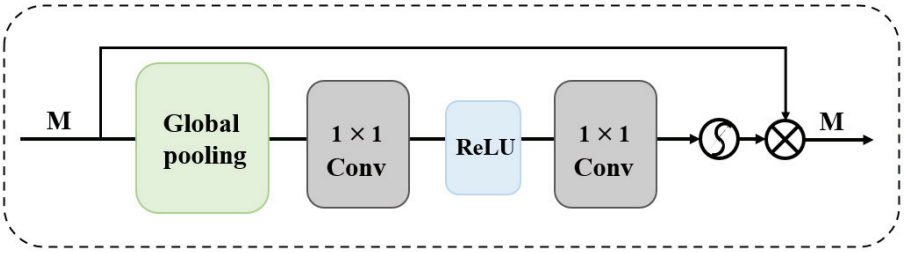


Figure 10. Core operational framework of CAM.

Attention Mechanism: An attention mechanism is a tool that prioritizes the allocation of resources to the most informative input regions, thereby enhancing the learning efficiency. With the increasing popularity of attention mechanisms, similar attention mechanisms have been introduced in the field of SISR. Mei et al. [51] proposed a deep residual network based on squeeze-and-excitation blocks, which can improve the performance and enhance the accuracy of reconstructed images. In order to fully utilize the potential of convolutional neural networks in the field of SR, Yan et al. [52] proposed a multi-scale attention network (MAN) that combines classical multi-scale mechanisms with emerging large kernel attention. The MAN network is comparable to SwinIR in terms of performance and computational efficiency, achieving a different trade-off between optimal performance and computational efficiency.

Traditional attention mechanisms have significantly improved the performance of SISR but often lead to complex network architectures and many parameters, resulting in slow inference speeds and large model sizes. To address this issue, SPAN [53] was proposed, an efficient SISR model that balances the number of parameters, inference speed, and image quality. SPAN employs a novel parameter-free attention mechanism that leverages symmetric activation functions and residual connections to enhance highly contributing information and suppress redundant information. This further demonstrates the effectiveness of attention mechanisms, which help to further improve model performance. Majid et al. [54] proposed a spatial edge residual attention model. This model combines edge and spatial attention to achieve better performance in image reconstruction. Yu et al. [55] proposed an algorithm that utilizes a dual attention module to capture implicit weight information and employs residual connections to fuse global features. This algorithm can allocate computing resources more effectively and accelerate the computing speed.

Feedback Mechanism: Feedback mechanisms enhance the SR reconstruction accuracy by introducing recurrent structures into the model, allowing it to use previous output information to guide subsequent computations. Feedback networks have been applied in SR, enhancing the network's feature extraction and reconstruction capabilities through feedback mechanisms. In SRFBN [56], the authors introduce feedback connections that allow the network to share and reuse features across different layers. These feedback

connections help the network to better learn the global and local features of images, and SRFBN can achieve better performance than existing methods with fewer residual dense blocks.

Gating Mechanism: Gating mechanisms improve the model's expressive power and computational efficiency by introducing gated units to control the flow of information. Specifically, gating mechanisms can dynamically adjust the transmission of information within the network based on the input features, allowing the model to focus more on important features and suppress unimportant ones, thereby enhancing the reconstructed image. Trishna et al. [57] proposed that a gating module with attention mechanism can be learned to effectively fuse relevant feature images and improve the accuracy of image reconstruction. Yao et al. [58] proposed a lightweight SISR guided by gated feedback and LatticeFormer. This method can improve the feature extraction ability and effectively obtain global features at different levels.

3.1.2. Efficient Structures

Although increasing the depth of the model is a direct method to improve SR performance, its enormous computational cost makes it difficult to deploy deep models in scenarios with limited computing power, such as mobile devices. In order to solve this problem, many efficient SISR solutions have emerged in the academic community in recent years, mainly achieving breakthroughs in two directions: network architecture optimization and knowledge transfer.

In terms of network architecture innovation, Hui et al. [59] were the first to propose the Information Distillation Network (IDN), which effectively compresses the number of parameters by constructing feature distillation channels. Subsequent research further upgraded it to the Information Multi-Distillation Network (IMDN), which significantly improves the feature fusion efficiency by adopting a cascaded distillation block architecture. The RFDN developed by Liu's team [60] reoptimizes the computational efficiency through lightweight feature distillation connections, while VapSR proposed by Zhou et al. [61] enhances the network's representational ability through a dynamic routing mechanism. It is worth noting that Liu et al. [62] innovatively designed a collaborative architecture consisting of an efficient residual block (ERB) and a high-frequency attention block (HFAB). The ERB module realizes deep feature learning through reconfigurable skip connections, and the HFAB uses the Laplacian operator for high-frequency feature detection and dynamically enhances detail reconstruction through pixel-level scaling factors.

In the lightweight technology route, the knowledge distillation method shows unique advantages. Lee et al. [63] pioneered the introduction of the teacher–student learning framework into the SISR field. By constructing a lightweight student model to inherit the knowledge distilled from the teacher network, a good balance between computational efficiency and reconstruction quality is achieved. In terms of hardware-adapted architectures, researchers have drawn on the successful experiences of MobileNet and ShuffleNet. Lin et al. [64] replaced standard convolution operations with depth-separable convolutions, which greatly reduces the parameter scale. Sun et al. [65] proposed ShuffleMixer, which has demonstrated the latest breakthrough. This technology, through the collaborative design of large-sized convolution kernels and channel shuffling operations, significantly expands the receptive field while maintaining mobile-end compatibility.

These innovative methods effectively resolve the computational bottlenecks of traditional deep models during mobile-end deployment through architecture improvements and algorithm optimizations, providing important technical support for the practical application of real-time SR technology. Future research can further explore directions such

as dynamic network architectures and the joint optimization of hardware awareness to continuously promote the development of lightweight SR technology.

3.1.3. Transformer-Based Methods

The core idea of the Transformer is the "self-attention" mechanism, which can capture the long-range information correlations between token elements. Recently, Chen et al. [66] made innovative improvements to the "self-attention" mechanism and proposed the recursive generalization self-attention. It integrates image features into representative feature maps through a recursive generalization module, achieving the aggregation of global key information. As a result, it can effectively capture global spatial information while maintaining a low computational cost, enhancing the model's perception of global features.

In recent years, Transformer-based models have emerged in an endless stream. For example, Liu et al. [67] proposed Swin Transformer, which is a highly innovative model. It introduced the sliding window mechanism, including non-overlapping local windows and overlapping cross-windows. By restricting the attention calculation within the windows, the computational amount is greatly reduced. Meanwhile, with the help of the sliding window operation, the attention mechanism can pay attention to global features, improving the performance of the model. Chen et al. proposed the Image Processing Transformer (IPT [68]), which was pretrained on large datasets, and introduced contrastive learning for different image post-processing tasks. Therefore, the pretrained model can be fine-tuned on the required tasks, showing good adaptability.

Meanwhile, Liang et al. proposed SwinIR [69] for image restoration. The Swin Transformer block proposed in this model is used for feature extraction, and DIV2K-Flicker2K was selected as the training set, reducing the dependence on large-scale datasets and being applicable to different scenarios. Zamir et al. [70] proposed Restormer, which reconstructs high-quality images by embedding convolutional neural networks (CNNs) within the Transformer. Subsequently, Chen et al. [71] used overlapping cross-attention modules and combined these with a pretraining strategy to enhance the performance of the Transformer model. Jiang et al. [72] mentioned an image SR reconstruction method based on a lightweight symmetric CNN–Transformer. Through its unique module design, it improves the processing ability for local and global features. Guo et al. [73] proposed the CMTNet model with a dual-branch structure. This model consists of a feature extraction module and an edge enhancement module. The feature extraction module is further divided into a Transformer high-frequency feature extraction branch and a feature extraction branch based on ConvLSTM. The feature fusion of the two branches ensures the order of spectral information, providing new ideas and methods for image SR reconstruction. In addition, Zhang et al. [74] proposed CVIformer in 2025, which is a new architecture that integrates cross-view interaction into Transformer. This architecture can effectively combine input images to restore the details of stereo SR.

3.1.4. Mamba-Based Methods

Mamba is a novel deep learning architecture that enables fast inference and accelerates computation speeds. To some extent, it can replace Transformer and improve efficiency. Recently, He et al. [75] proposed the Mamba Pixel Sequential Interactive Network (MPSI). This network effectively enhances long-distance connections and improves the effectiveness of image reconstruction by using channel Mamba blocks (CMB) and Mamba channel recursive modules (MCRM). Wang et al. [76] proposed a collaborative network consisting of Mamba and a CNN (CNMC), which is mainly used for lightweight image SR. It can reduce model complexity, improve the computational speed, and extract deep features that are beneficial for reconstruction. Romina et al. [77] proposed MambaLiteSR, which

mainly utilizes the Vision Mamba architecture and effectively reduces power consumption and performance loss through methods such as knowledge distillation. It is a lightweight image SR model. Jiang et al. [78] proposed Rep Mamba, an architecture based on the RSISR state space model. They designed a cross-scale state propagation mechanism (CSSP) and constructed a lightweight progressive fusion module (LPFM), which could improve the reconstruction accuracy while achieving high computational efficiency.

3.2. Perceptual Quality Methods

The PSNR and SSIM are two important metrics commonly used for image quality assessment judgment. Recently, Hsu et al. [79] proposed the Dense Residual Connected Transformer (DRCT), which has achieved good results in achieving a high PSNR and SSIM. Blau et al. [80] first revealed the existence of perceptual distortion trade-offs. Subsequently, Xue et al. [81] proposed the unreachable area as a new indicator for evaluating and comparing the trade-off between distortion and perception, achieving better results. However, the question of how to improve the perceptual quality while avoiding distortion still needs to be addressed. Therefore, this section proposes several methods to address this trade-off dilemma, aiming to achieve the dual goals of reducing image distortion and maintaining good perceptual quality.

Perceptual Loss: Although pixel-based losses like the L1 loss (mean absolute error) and L2 loss (mean square error) have been heavily utilized in the pursuit of high-quality images, the loss function focuses only on pixel matching, without paying attention to the visual quality of the reconstructed image, and it is usually difficult to capture the differences between SR images and high-resolution images at the perceptual level. Therefore, to address this issue, methods for content loss, texture loss, object perception loss, dual perception loss, and wavelet domain loss have been proposed, and these methods enable the loss function to more accurately reflect the perceptual and semantic differences between images.

Content loss: This involves ensuring that the reconstructed image is highly consistent with the target image in terms of content. For example, aligning SR images with human visual perception features can be used to integrate content loss metrics into SISR, and a mechanism for content loss can be implemented through pretrained classification networks. Generally speaking, pretrained classification networks such as VGG [82] and ResNet [25] are the most used. By employing this framework, the semantic discrepancies across visual representations can be precisely quantified. The content loss can be expressed as follows:

$$\mathcal{L}_{\text{content}} = \frac{1}{H_l W_l C_l} \sum_{i=1}^{H_l} \sum_{j=1}^{W_l} \sum_{k=1}^{C_l} |\phi_l(I_{SR}) - \phi_l(I_{HR})| \tag{16}$$

where ϕ_l is the feature extractor of the lth layer of the pretrained network. H_l, W_l, and C_l represent the height, width, and the number of channels of the feature map in the lth layer.

Texture loss: Texture loss is also known as the style reconstruction loss. Gatys et al. [83,84] were inspired by the transfer of artistic styles in oil painting and proposed the texture loss for SISR. In SISR, the core goal is to reconstruct images with consistent texture, color, and edge features, which means that the reconstructed image should be highly consistent with the target image in terms of style. Assuming that the target image is It and the generated image is Ig, their Gram matrices are G (It) and G (Ig), respectively. Texture loss is typically calculated using the mean square error (MSE), where C is the dimension of the Gram matrix, i.e., the number of channels in the feature map. The calculation formula is

$$\mathcal{L}_{\text{texture}} = \frac{1}{C^2} \sum_{i=1}^{C} \sum_{j=1}^{C} \left(G(I_t)_{i,j} - G(I_g)_{i,j} \right)^2 \tag{17}$$

With the texture loss, the model-generated image is closer to the target image in terms of texture details, which improves the image quality, reduces artifacts, and enhances the visual fidelity and model generalization [85].

Targeted perceptual loss: The traditional perceptual loss has limited capabilities due to the fact that semantic information is not taken into account in the image reconstruction error. Rad et al. [86] proposed a new approach by optimizing a deep network-based decoder using a targeted objective function that divides the image into three regions, namely the background (gb), the boundary (ge), and the target (go), and using the segmentation from the labels generated from the object, background, and boundary (OBB) labels to estimate the appropriate perceptual loss of the boundary while considering the texture similarity of the background, resulting in more realistic textures and sharper edges. Guo et al. [87] used targeted perceptual loss to guide the network to recover the texture details of the image; compared to SRGAN and other networks, their method can generate more realistic and reliable textures.

Double perception loss: Abnormal feature propagation in the decoder module can exacerbate model instability due to excessive scaling operations. This mechanism will continuously distort and accumulate in the network layer, systematically compromising the authenticity of the output image. Song et al. [88] proposed the dual perceptual loss, which takes into account the benefits of learning two features simultaneously and effectively enhances image reconstruction.

Wavelet domain loss: In the current methods, attempting to reconstruct image details can result in artifacts and illusions. Korkmaz et al. [89] showed that, by training the SR model of the GAN using a wavelet domain loss function, it is possible to better learn the features of real details versus artifacts. Extensive experimental results show that the model achieves a better perception–distortion trade-off based on multiple objective measurements and visual assessments.

3.2.1. Adversarial Training

The generative adversarial network is a new framework for evaluating generative models through an adversarial process, proposed by Goodfellow et al. [20] in 2014. As shown in Figure 11, the framework mainly consists of a generator for capturing the data distribution and a discriminator for judging the "authenticity" of the input data, and the framework trains the network by means of adversarial games.

Ledig et al. [21] first applied generative adversarial networks (GANs) to SR images and proposed the Super-Resolution Generative Adversarial Network (SRGAN). SRGAN not only utilizes a GAN to improve the network structure but also applies perceptual loss to SR and adopts a loss function that combines perceptual loss and adversarial loss, significantly improving the details of reconstructed images in terms of perceptual quality. However, the "deception" of the GAN framework results in relatively low objective evaluation metrics, such as the PSNR and SSIM. Inspired by SRGAN, Wang et al. [90] proposed ESRGAN, which achieves more effective perception reconstruction by improving SRGAN. Wang et al. [91] proposed a new GAN inverse method that utilizes the powerful generation abilities of StyleGAN-XL while improving the quantitative and qualitative results of SISR. In 2024, Yang et al. [92] improved the generator of SRGAN and designed a multi-scale feature fusion module as the backbone feature extraction network of the generator. This module represents the multi-scale features of images at the granularity level and extends the perceptual domain of each network layer, enabling the full extraction of image feature information. This model enhances texture details and achieves better visualization effects. Huang et al. [93] proposed a network architecture, CL-GAN, which extracts features from clustered image regions and constructs them based on the GAN

framework. This architecture enhances the SR reconstruction process and improves the image reconstruction performance.

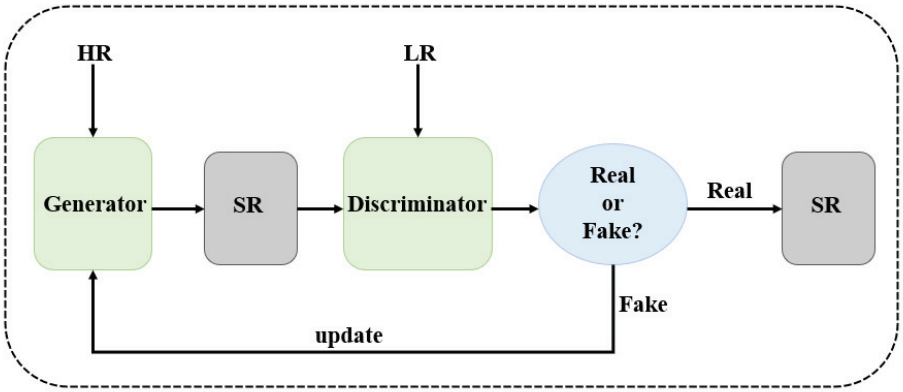


Figure 11. Generative adversarial networks.

3.2.2. Cycle Consistency

The theory of cyclic consistency assumes that there is a mapping relationship between the source domain and the target domain. Zhu et al. [19] proposed a cyclic consistency mechanism. In the absence of paired examples, they propose a learning method that can convert images from the source domain X to the target domain Y. The goal of this method is to use adversarial loss to learn a mapping G: $X \rightarrow Y$, so that the distribution of images generated by $G(X)$ is the same as that of Y. This mapping is extremely underconstrained; thus, the approach relates it to the inverse mapping F: $Y \rightarrow X$ and pushes $F(G(X)) \approx X$ by introducing the cyclic consistency loss. The superiority of this approach has been shown by quantitative comparisons with several earlier approaches.

In SISR, the idea of cyclic consistency has also received a lot of attention. We can learn both the mapping from LR to HR and the inverse procedure given an LR image domain X and an HR image domain Y. It has been demonstrated by researchers that more realistic images can be achieved by learning how to perform image degradation without paired data [94]. A cell cycle network is suggested in CinCGAN [95] that upsamples using a pretrained model after first mapping noisy and ambiguous inputs to the noiseless LR domain. This method can improve the quality and related performance of image reconstruction. In DRN [96], a closed loop that offers extra supervision is formed by learning a mapping from HR images to LR images to estimate the downsampling kernel and reconstruct the LR images. In addition, DRN provides us with a new learning method that can directly learn from LR images, making the images closer to the real world. In SGA [97], the correlation between pixel features is constructed through a parameter-free Gram matrix, and the invariance of the feature layer is maintained by using the consistency loss. This loss normalizes the consistency between the LR input and its corresponding LR output, which enables us to reconstruct more realistic images and improves the correlation performance.

3.2.3. Diffusion-Based Methods

In the field of SISR, diffusion-based methods have received extensive attention in recent years, inspired by the Denoising Diffusion Probabilistic Model (DDPM). To date, a series of representative models have emerged in the field of diffusion-based methods.

SRDiff [98] is the first model to apply diffusion models to SISR. It utilizes the iterative properties of Markov chains and uses LR as input to gradually convert Gaussian noise into HR. This process enables the diversification and realism of SISR, laying an important foundation for subsequent research. Regarding SR3 [99], its uniqueness lies in its training

method, which trains the model at different levels of noise, enabling the model to be optimized for Gaussian noise input. Compared with GANs, SR3 generates images that are more realistic and natural, and it can present more delicate and vivid image details. Meanwhile, it is worth noting that SR3 adopts a multi-stage approach and performs processes such as scaling optimization, reflecting the concept of hierarchical diffusion. Layered diffusion has significant advantages in efficiency. It can ensure the rationality of the global structure while improving the authenticity of local details, effectively enhancing the reconstruction efficiency. In IDM [100], the framework innovatively integrates implicit neural representations with denoising diffusion models in an end-to-end manner. This helps the model to learn continuous image resolution representations, enabling it to generate images of different resolutions and further expanding the application scope of diffusion-based methods. DR2 [101] mainly focuses on processing complex low-quality facial images. The algorithm first uses DDPM to denoise and extract features from low-quality facial images. Then, through a designed enhancement module, the preliminary processed images are optimized and reconstructed, ultimately restoring them to high-resolution and clear facial images. This has important application value in related fields such as facial recognition.

At present, there are still some methods that utilize existing diffusion-based models to assist SISR in improving the accuracy of related reconstruction. StableSR [102] and DiffBIR [103] are two models that utilize the powerful prior knowledge of pretrained text-to-image diffusion models. By fine-tuning these pretrained models, StableSR and DiffBIR can effectively combine text information with the image SR task, achieving good results in real-world SISR and providing new ideas for the resolution of image SR problems in practical scenarios. Similarly, the integration of pretrained models and SISR demonstrates enormous potential. Pretrained models can provide guidance for SISR by learning image semantics and other information from large amounts of data, thereby accelerating the speed and improving the accuracy of image reconstruction. For example, by aligning the features of LR input with the semantic embeddings of pretrained models, SISR models can better maintain consistency in the object shape and scene structure during upsampling. DiffIR [104] utilizes a model pretrained on real images and incorporates rich prior information into the SISR model. This approach enables DiffIR to more accurately capture the features and structures of the image, greatly improving the efficiency and accuracy of the model. Wu et al. [105] proposed the potential diffusion of continuous-scale super-resolution (LDCSR) model. This model adopts a two-stage latent diffusion paradigm to capture the prior differences between HR and LR images and learns in the latent space to predict the true prior differences, thereby improving the quality of image reconstruction.

Although diffusion-based SISR methods have achieved remarkable results in terms of image generation fidelity, effectively reducing artifacts and generating more realistic and natural images, they also face some challenges. On the one hand, this type of method requires a large number of new samples for training to ensure that the model can learn sufficient image features and variation patterns; this may be limited by data acquisition and storage issues in practical applications [106]. On the other hand, the convergence speed of the model is slow, meaning that it takes a relatively long time to process images, making it difficult to meet the requirements of some applications with high real-time requirements. Recently, the question of how to overcome these shortcomings and further improve the efficiency and practicality of diffusion-based methods has become one of the key research areas for researchers in this field.

3.3. Additional Information Utilization Methods

We have already introduced the relevant work on improving the image perception quality in the above. In this section, we will explore how to utilize the additional infor-

mation of images to enhance the model's correlation performance, enabling the model to better reconstruct real images.

To further enhance the performance of models, researchers have explored various additional information utilization methods. These methods leverage various types of information both inside and outside the image, providing new ideas and approaches for the optimization of SISR models. These methods are mainly based on fundamental methods such as image internal statistics, multi-factor learning, prior guidance, reference images, and knowledge distillation.

Methods Based on Image Internal Statistics: Zontak et al. [107] discovered that there are some image patches in images for which it is difficult to find matches in external databases. Moreover, the internal entropy of image patches within a single image is much smaller than the external entropy of patches in a collection of natural images. Based on this characteristic, methods that utilize image internal statistics show unique advantages when dealing with such images. Take the ZSSR [18] method as an example. It fully utilizes the internal statistical properties of the image to train a specific CNN for that image. During the training phase, multiple LR-HR image pairs are generated through data augmentation techniques and used to train the CNN. In the testing phase, the LR images are input into the trained CNN to obtain the reconstructed HR image. In SinGAN [108], a fully convolutional GAN with a pyramid structure is constructed to learn the internal block distributions of images at different scales. When generating HR images, multiple upsampling operations are performed on LR images to obtain an SR output.

Multi-Factor Learning Methods: Traditionally, SISR usually requires the training of specific models for different upsampling factors, which undoubtedly increases the complexity and cost of model training. The multi-factor learning method can overcome this limitation by deeply exploring the scale correlations among various upsampling factors, allowing a single model to adapt to multiple different upsampling factors and thus enhancing the model's performance. In LapSRN [40], the LR image is gradually reconstructed in a pyramid network structure. During this process, not only can the final large-scale reconstruction result be obtained, but also the intermediate results can be directly used as the outputs corresponding to the respective multiple factors. Wang et al. [109] also found that there are correlations among multi-scale tasks. Based on this, they designed specific scale processing modules at the beginning and end of the model to handle different upsampling factors. Li et al. [45] further optimized this strategy in the MDCN model. Compared with MDSR, MDCN can effectively utilize model parameters and learn correlations between different scales to a greater extent. Optimizing this strategy can enhance the effectiveness of the model, thereby improving the related performance.

Prior Guidance Methods: Currently, many SISR methods tend to build end-to-end CNN models to complete SR tasks. However, in the process of image processing, due to the loss or damage of many useful image features, these models are not satisfactory in reconstructing real high-frequency details. To address this issue, the SISR framework based on prior knowledge has emerged. Many experiments have shown that, by utilizing prior image information, models can not only converge faster but also achieve higher reconstruction accuracy. In recent years, various types of image priors have emerged, such as total variation (TV) priors, sparse priors, and edge priors. Yang et al. [110] proposed a Deep Edge-Guided Recursive Residual (DEGREE) network for SR image reconstruction. Fang et al. [111] proposed an efficient and accurate soft edge-assisted network (SeaNet). Compared to DEGREE, SeaNet can learn more accurate image edges. Moreover, image prior methods also play a positive role. For example, in SFTGAN [112], a spatial feature transformation (SFT) technique is used to reconstruct more realistic image details through semantic category priors. In SPSR [113], the model utilizes gradient maps and integrates

them into the SR branch to provide better prior information, which facilitates image reconstruction, improves image authenticity, and enhances texture details. In FeMaSR [114], VQ-GAN [115] is first pretrained and its related discrete features are provided as prior information to the model, which can effectively achieve high-quality image restoration. Huang et al. [116] proposed the Multi-Modal Prior-Guided Diffusion Model (MPGSR), which involves two stages of multi-modal prior-guided extraction and adaptive guided injection and can restore realistic high-resolution images with better performance.

Reference Image Methods: Unlike traditional SR methods, reference image-based SISR (RefSR) uses additional reference images to assist SR in reconstruction. Meanwhile, in order to enhance the texture information of images, researchers have developed various methods, such as image registration and patch matching. Some RefSR methods [117,118] have been proposed, such as those assuming that the reference image and LR image have similar content and aligning these two images. For example, Yue et al. [117] were able to minimize the resource consumption by matching reference images with LR images. In CrossNet [118], it is proposed to use optical flow technology to align the reference image and the LR image at different scales and connect the aligned images to the corresponding layers of the decoder. Zhang et al. [119] applied patch matching between LR images and reference images to achieve the adaptive transfer of the reference image texture to LR images. In TTSR [120], Yang et al. proposed a texture transformation network that can extract corresponding texture details from images and transfer them to LR images, thereby improving the image quality. Cong et al. [121] proposed reference-based iterative interaction (RIISSR) for stereo image super-resolution, which mainly utilizes P2 matching to establish cross-correspondence relationships, demonstrating better performance.

Knowledge Distillation Methods: Knowledge distillation is a technique that transfers the representational power of large models to small models, and it has shown significant effectiveness in many computer vision tasks. Knowledge distillation is mainly divided into two types, namely soft label distillation and feature distillation. Regarding soft label distillation, the softmax output of large models serves as soft labels, providing valuable implicit knowledge for small models. Regarding feature distillation, the intermediate feature maps of large models are transferred to small models for easier extraction. In the field of SISR, some studies have introduced the knowledge distillation technique to further improve the performance of lightweight models. In SRKD [122], a small network is guided by a larger network, and it can achieve similar feature distributions to a large-scale network. JDSR [123] explores joint distillation by combining relevant information, effectively improving the performance of lightweight models. CSD [124] innovatively combines contrastive learning and distillation tasks to further resolve the related problems of lightweight models and optimize model performance. Romina et al. [77] proposed MambaLiteSR, which utilizes the Vision Mamba architecture, integrates state space blocks and reconstruction modules, and employs knowledge distillation to optimize the performance and improve the efficiency.

3.4. Comparison of Method Types for Simulating SISR

Based on the SISR technology classification framework, including efficient network design, perceptual quality optimization, and additional information utilization, we list three types of SISR models in Table 3. From the PSNR and SSIM indicators in the table, it can be seen that partitioned SISR technology has a significant performance improvement. Therefore, the proposed classification framework can effectively classify SR methods and reflect their performance characteristics, achieving progress compared to previous work.

Table 3. Comparison of method types for simulating SISR.

Type	Model	Set5 PSNR/SSIM	Set14 PSNR/SSIM	Urban100 PSNR/SSIM	Training Dataset
Efficient Network/ Mechanism Design Methods	SRCNN [7]	30.48/0.8628	27.50/0.7513	24.52/0.7221	T91+ImageNet
	DSRN [37]	31.40/0.8830	28.07/0.7700	25.08/0.7470	T91
Perceptual Quality Methods	MSRN [29]	32.07/0.8903	28.60/0.7751	26.04/0.7896	DIV2K
	SeaNet [111]	32.33/0.8970	28.81/0.7855	26.32/0.7942	DIV2K
	CRN [38]	32.34/0.8971	28.74/0.7855	26.44/0.7967	DIV2K
	EDSR [27]	32.46/0.8968	28.80/0.7876	26.64/0.8033	DIV2K
	RCAN [48]	32.63/0.9002	28.87/0.7889	26.82/0.8087	DIV2K
Additional Information Utilization Methods	SwinIR [69]	32.92/0.9044	29.09/0.7950	27.45/0.8254	DIV2K+Flickr2K
	GRL-B [125]	33.10/0.9094	29.37/0.8058	28.53/0.8504	DIV2K+Flickr2K

3.5. Real-World Image SR

In real-world scenarios, image degradation patterns are complex and difficult to model, often involving anisotropic blur, downsampling, and the addition of signal-dependent noise. Additionally, images are affected by the in-camera signal processing (ISP) pipeline. These factors result in the poor performance of traditional SISR models when applied to real-world images. Furthermore, most existing models are restricted to specific integer upscaling factors, significantly limiting their practical application and generalizability. To address these challenges, researchers have proposed a series of deep learning-based methods, which can be broadly categorized into two main approaches: blind image SR and arbitrary-scale SR.

3.5.1. Blind Image SR

Blind SISR, which aims to perform SR reconstruction on LR images corrupted by unknown and complex degradation processes, has attracted considerable interest due to its critical role in real-world applications. However, it is important to note that there is no standardized definition for blind SISR. For a comprehensive discussion of its scope and challenges, readers may refer to [126]. In this study, we adopt a simplified taxonomy for blind SISR methodologies, dividing them into two primary categories according to their degradation modeling strategies: explicit degradation modeling and implicit degradation modeling.

Explicit Degradation Modeling: With the advancement of deep learning technologies, significant progress has been made in the field of blind SISR. Explicit degradation modeling, as a critical component, provides essential support for high-quality image restoration. Early methods had certain limitations, such as the internal GAN-based kernel estimation method proposed by Bell Kligler et al. [127], which overly relied on kernel estimation modules and lacked unified modeling. To address this limitation, researchers have begun exploring methods to integrate degradation estimation and SR reconstruction into a unified framework. For instance, Gu et al. [128] proposed an iterative kernel correction algorithm that progressively refines blur kernel estimates through alternating optimization, while Luo et al. [129] introduced the deep alternating network (DAN), which simultaneously estimates blur kernels and reconstructs SR images within a single network. These methods significantly improve model robustness and the reconstruction quality but still face challenges in complex degradation scenarios, where estimation errors could lead to reconstruction failures.

In recent years, deep learning-based integrated approaches for degradation estimation and SR reconstruction have become a research hotspot. Sun et al. [130] proposed the learning correction error (LCE) method, which uses a lightweight corrector to generate corrected LR images, reducing the space required and improving the correlation perfor-

mance. Li et al. [131] introduced BlindDiff, which combines the maximum a posteriori (MAP) optimization framework into the diffusion model to reconstruct SR images and make them clearer. Ohayon et al. [132] approached image restoration from the perspective of fairness, proposed the concept of perceived fairness (PF), and defined a group perception index (GPI) to evaluate the restoration quality between different images. Additional frameworks include IKR-Net [133], which handles complex degradation through iterative kernel reconstruction and noise estimation, and CDFormer [134], which utilizes content-aware degradation-driven Transformers to capture degradation and content representation, further advancing blind image SR technology.

Implicit Degradation Modeling: In blind SISR research, implicit modeling has received a lot of attention, and it mainly guides SR reconstruction through learning image degradation features. In 2021, Wang et al. [135] proposed DASR, a method based on self-supervised contrastive learning that compares the features of different image blocks through contrastive learning. DASR implicitly captures degraded information without explicit parameter estimation, and it uses degraded features to guide the SR network for high-quality reconstruction under unknown degradation. Subsequently, Yang et al. [136] proposed DMSR in 2022, which is a method that utilizes the degradation of BlindSR to guide the network and improve the quality of image restoration.

Furthermore, Chen et al. [137] introduced DIEM, a dynamic degradation intensity estimation module for adaptive BlindSR. DIEM dynamically estimates pixel-level degradation information to guide reconstruction, and its core innovation is that it generates specific degradation feature maps for input changes, effectively solving common problems such as oversmoothing or oversharpening in real-world scenarios. Liu et al. [138] proposed a blind SR model based on Transformer in 2024. The model integrates contrastive learning into the Transformer network and combines the advantages of a CNN and Transformer to extract local features and capture global semantic features. The model has achieved significant performance improvements on multiple benchmark datasets. Jiang et al. [139] proposed a blind image super-resolution framework (CdCL) based on content-decoupled contrastive learning and designed an implicit degradation-adaptive module for detail perception. This framework and module can better adapt degraded representations to LR features, reduce model complexity, and improve model performance. In summary, implicit modeling in blind SR has undergone significant evolution, and future research may focus on optimizing relevant degradation feature learning and exploring more complex degradation scenarios to enhance model generalization and robustness.

3.5.2. Arbitrary-Scale SR

Despite the continuous development of SISR technology, it still faces many challenges in practical applications, such as using a single model to handle arbitrary scaling factors. To address these issues, Wang et al. [140] proposed a multi-scale dilated convolution residual network, which includes residual attention dense blocks (RADBs) and dilated multi-scale residual modules (DMRMs). The RADB is used to learn features from the original low-resolution images, enhancing the network's ability to recognize both high- and low-frequency information. The DMRM, based on dilated convolutions, aims to address the limitations of the convolutional receptive field and potential information loss when extracting features through a single channel. However, despite the excellent performance of these methods at specific scales, they still have limitations when dealing with arbitrary scale factors. To overcome this, Li et al. [141] proposed a novel multi-scale cross-fusion network (MCNet) for efficient SR tasks with arbitrary scale factors. MCNet introduces a scale-wise module (SWM) and a multi-scale cross-fusion module (MSCF), which combine scale information with pixel features and enrich spatial information through multi-scale

feature interaction learning while removing redundant noise. Wei et al. [142] proposed a dynamic implicit network (DINet) for multi-contrast MRI with arbitrary-scale SR, which mainly includes scale-adaptive dynamic convolution, a dual-branch implicit attention module, and the fusion of response modules to improve model performance.

3.6. Domain-Specific Applications

At present, image SR technology has been widely applied in various fields. This section will introduce various application scenarios for SR technology, as shown in Figure 12.

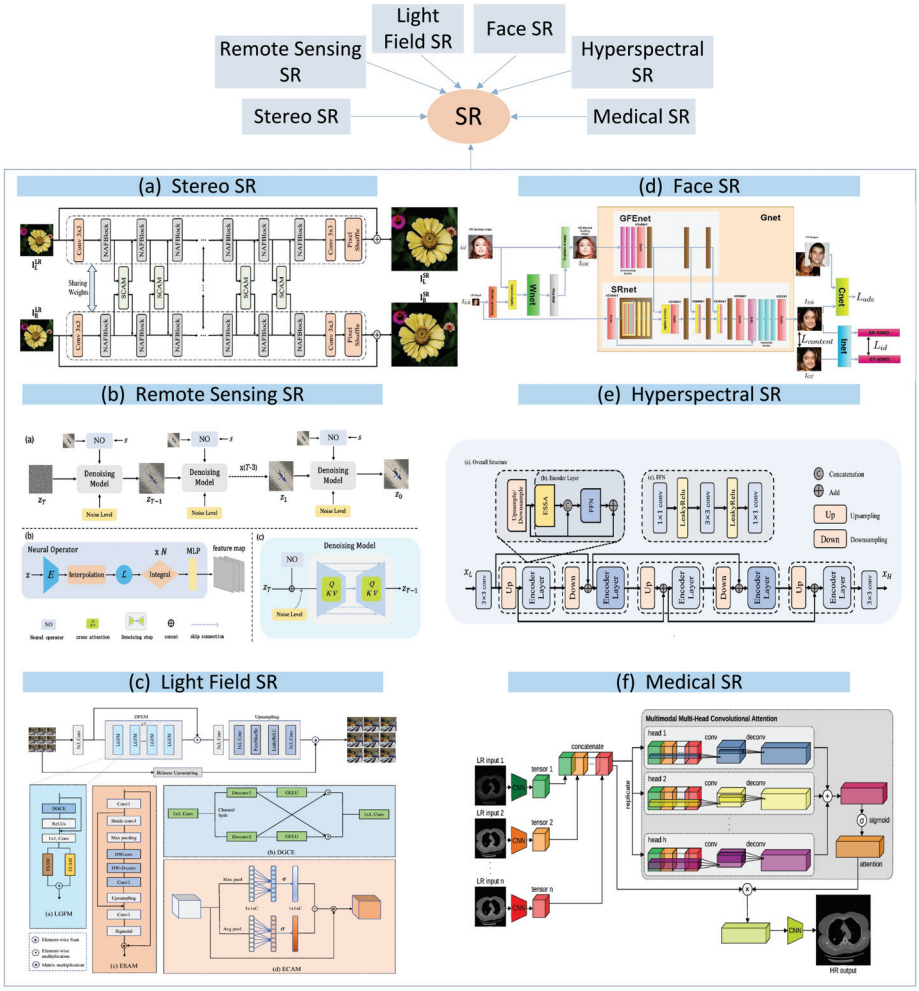


Figure 12. Examples of various popular SR tasks. (**a**) NAFSSR [143] for stereo SR, (**b**) NeurOp-Diff [144] for remote sensing SR, (**c**) LGFN [145] for LF SR, (**d**) GWAInet [137] for face SR, (**e**) Essaformer [146] for hyperspectral SR, and (**f**) MMHCA [147] for medical SR.

3.6.1. Stereo Image SR

As is well known, dual-camera systems are commonly used for depth information estimation, but stereo imaging technology can also play a role in image restoration tasks. In these tasks, it is necessary to process two images with a difference of more than one pixel, and fully utilizing the information of these two images can effectively improve the spatial resolution.

In StereoSR [148], Jeon et al. proposed a method to improve the resolution of stereo images by learning the sub-pixel disparity. However, in this method, there are still certain challenges when dealing with large stereoscopic images with significant disparity variations. To address this issue, Wang et al. [149,150] proposed a disparity attention mechanism with a global receptive field along the epipolar line, which can generate reliable correspon-

dences between stereo image pairs and thus improve the quality of the reconstructed SR images. In [151], the researchers introduced a high-quality stereo image dataset named Flickr1024, which contains 1024 pairs of stereo images, providing a rich resource for stereo image SR research. In [152], the authors designed a stereo attention module that integrates stereo information with a pretrained single-image SR network through bidirectional interaction, further enhancing the performance of stereo image SR. In [109], a symmetric bidirectional disparity attention module was proposed, combined with an inline occlusion handling scheme, to achieve effective cross-view information interaction. In [153], the researchers proposed SSRDE-FNet, a unified framework capable of simultaneously handling stereo image SR and disparity estimation tasks. In [143], in addition to using NAFNet [154] to extract single-image features from the left and right views separately, a stereo cross-attention module was introduced to fuse features from both views. Furthermore, in [155], the researchers proposed Steformer, an efficient Transformer-based stereo image SR model. Through the residual cross Steformer block (RCSB) and cross-to-internal attention mechanism (C2IAM), Steformer demonstrates excellent performance in cross-view information extraction and feature fusion. In [156], the double stereo cross-attention module (DSCAM) was proposed, which enhances the integration of complementary cross-view information through an overlapping window mechanism and refines and emphasizes the combined features through additional multiplication steps.

3.6.2. Remote Sensing Image SR

At present, satellite image processing technology is constantly advancing, and remote sensing is also playing an increasingly important role. However, remote sensing technology has significant shortcomings in terms of spatial resolution, spectral resolution, and radiometric resolution. Therefore, the current remote sensing technology still faces many problems.

At present, there are various innovative remote sensing image SR methods. For example, in [157], the researchers proposed an innovative unsupervised hourglass neural network architecture that can be used to enhance the SR effects of remote sensing images. This network introduces diverse spatial patterns by generating random noise, which can be extended under global reconstruction constraints to improve the image quality. In [158], the researchers designed a deep residual compression and excitation network (DRSEN) to address the complex distribution of remote sensing images. In [159], a mixed high-order attention network (MHAN) was proposed, which consists of two parts: one is a feature extraction network for the efficient extraction of image features, and the other is a feature refinement network with a high-order attention mechanism dedicated to detail recovery to enhance the SR effects of images. In [160], a Dense Sampling Super-Resolution (DSSR) network was proposed to explore large-scale remote sensing image SR reconstruction, significantly improving the image resolution through an optimized network structure and sampling strategy. In [161], a hybrid-scale self-similarity development network (HSENet) was proposed, which can utilize both single-scale and cross-scale similarities to achieve high-quality image reconstruction. In [162], Wang et al. proposed a multi-scale enhancement network (MEN) that fully utilizes the multi-scale features of remote sensing images to significantly enhance the network's reconstruction capabilities. In [163], Liu et al. proposed a dual-learning-based graph neural network (DLGNN), in which the graph neural network (GNN) aggregates adjacent feature patches across scales, fully considering the self-similarity of remote sensing images to further enhance SR effects. In [144], a continuous remote sensing image SR method was proposed, which is based on the neural operator diffusion model (NeurOp-Diff). This method learns arbitrary-scale resolution representations through neural operators and uses diffusion models for denoising, effectively solving the problems of artifacts and oversmoothing in existing SR methods. In [164], a new frame-

work called ESC-MISR was proposed for the multi-image super-resolution (MISR-RS) task in remote sensing, which reconstructs high-resolution images by enhancing the spatial correlations between multiple images.

3.6.3. Light-Field Image SR

A light-field (LF) camera is a device with unique advantages that can capture the light-field information emitted by the scene and has multiple related views. In recent years, its importance has been constantly highlighted. However, LF cameras face a trade-off between spatial and angular resolutions [165]. To overcome this limitation, SR technology was introduced. In [166], the researchers introduced a cascaded convolutional neural network architecture to simultaneously enhance the spatial and angular resolutions of light-field images. They also constructed a new light-field image dataset for training and validation. In order to reduce the dependence on precise depth or disparity information as a prior for LF image SR, Sun et al. [167] proposed a bidirectional recurrent convolutional neural network and an implicit multi-scale fusion strategy for SR image reconstruction. In [168], a Transformer-based LF image SR method was proposed, which integrates complementary information between different views using corner Transformers and captures local and long-range dependencies in each sub-aperture image using spatial Transformers. In [145], a lightweight LF image SR method was proposed, which introduces a local convolutional modulation module to achieve efficient feature extraction and fusion, enabling the model to capture more detailed information in the light-field image while maintaining a lightweight design. Wang et al. [169] proposed a practical SR method for real light-field images. This convolutional neural network, designed by learning to adjust to different degradations and integrating spatial and angular information from the light field, achieves better performance and generalization to real light-field images.

3.6.4. Face Image SR

Among the many directions of SR technology applied to domain-specific images, face image SR is the most well known. Due to its huge application potential in facial recognition systems, such as security and surveillance systems, face image SR has developed into a research hotspot, attracting numerous scholars. In recent years, facial image SR has been continuously improving. In [170], CPGAN was proposed, which optimizes image quality through traditional facial SR loss and a new illumination compensation loss. In [171], the generator architecture adopts a spatial Transformer network to overcome problems related to the misalignment of input images. In [172,173], it is shown that minimizing the distance between SR and HR facial images can preserve identity-related features and improve the image accuracy. In [174], mask occlusion is regarded as image noise, and a joint and collaborative learning network (JDSR-GAN) is constructed to handle the SR task for masked face images. Qiu et al. [175] proposed a dual-path hybrid network (DPHNet) that can solve the SR problem of motion-blurred facial images. The network includes CNN and Transformer branches, which are integrated through convolutional block attention modules to effectively improve the performance. These methods [175–178] are mainly based on the prior generation of a GAN, which can restore the image to a realistic, textured, and high-definition facial image.

3.6.5. Hyperspectral Image SR

Unlike the human eye, which can only perceive visible light, hyperspectral imaging technology has the ability to collect and process information in the full electromagnetic spectrum range [179]. However, the performance of hyperspectral systems has certain limitations, and it is difficult to balance the spatial resolution and spectral resolution. Based on this, researchers have conducted research on hyperspectral image SR to overcome this difficulty.

In [180], multi-scale 3D convolution was used to capture local features in both the spectral and spatial domains. By integrating the advantages of the Mamba model in efficiently modeling long-distance information, good results have been achieved. In [146], Zhang et al. proposed an efficient Transformer model for hyperspectral image SR through a novel and efficient self-attention method based on the SCC kernel. In [181], a spectral and correlation algorithm for hyperspectral remote sensing image SR was introduced, which can fully represent the relevant features, improve the efficiency, reduce the costs, and thus improve the spatial resolution of hyperspectral images. In [182], Dang et al. proposed a dual-branch convolutional neural network for the fusion-based SR task of hyperspectral images. This neural network is based on cross-modal feature attention interaction and can generate hyperspectral images with high-fidelity spectral and spatial information. In [183], a hyperspectral image SR method based on adaptive depth prior regularization is adopted, which can solve the problem of the excessive dependence on priors in current hyperspectral SR methods, resulting in improved parameter interpretability. Jia et al. [184] proposed DiffHSR, which utilizes multi-modal techniques and generative priors and employs low-cost data and fine-tuning methods to more effectively improve the processing of different degraded images and enhance the image quality.

3.6.6. Medical Image SR

Medical imaging SR plays an irreplaceable role in clinical diagnosis and surgical planning, as well as in medical imaging, such as computed tomography (CT) and magnetic resonance imaging (MRI). Clear and accurate images can help medical staff to comprehensively understand the patient's physical condition, provide strong support for subsequent medical decisions, and facilitate accurate disease diagnosis and reasonable surgical plans. The rapid development of deep learning technology has brought new opportunities to the field of medical image SR.

Chen et al. [185] proposed a multi-level dense connected super-resolution network (mDCSRN) trained under GAN guidance, which can not only be used to generate HRMR images but also accelerate the computation speed. In [186], a 3D super-resolution convolutional neural network (3DSRCNN) was proposed, which can improve the resolution of 3D-CT volumetric images and enhance the image quality. In [187], in order to reduce the burden of model representation, Zhao et al. proposed a deep channel segmentation network (CSN). In [188], Peng et al. introduced a spatially aware interpolation network (SAINT) for medical slice synthesis, which can effectively alleviate the storage limitations of voxel data. In [189], Feng et al. proposed a Task Transformer Network (T2Net) that can share representations and transfer features between reconstruction and SR tasks. In [147], Georgescu et al. used multi-modal LR input for medical image SR and proposed a novel multi-modal multi-head convolutional attention mechanism for multi-contrast medical image SR. In [190], implicit neural network representation (INR) was proposed, which is an emerging deep learning technique that overcomes the limitations of discrete voxel grids by representing medical images as continuous functions. It has demonstrated its efficiency and flexibility in medical image reconstruction tasks. Kui et al. [191] proposed the Iterative Collaborative Network (ICONet), which mainly consists of a reconstruction branch, an SR branch, and an SR-Rec fusion module. It can effectively solve the problems of incomplete feature extraction in traditional medical image SR methods and insufficient interaction between tasks in multi-task learning. Li et al. [192] proposed a self-correcting texture supplementation network (STS-SR) for reference-based image super-resolution. This network includes a texture-specified self-correcting feature transfer module and a cross-scale texture complementary network; it performs well and can improve the role of autonomous artificial intelligence in healthcare.

4. Reconstruction Results

In this section, we list some representative reconstruction results obtained from SISR models and compare them, including the training dataset and model parameters, as shown in Table 4. Based on the differences in relevant data, it can be found that (1) for the model, large datasets can produce better reconstruction results and better correlation performance; (2) contrary to traditional belief, a model's performance might not improve simply if it has more parameters; (3) Transformer-based models can demonstrate unique advantages, thereby achieving better model performance; (4) the research on micro-models is still not advanced, and, in the future, we should focus on conducting research in this area to overcome the difficulties faced by micro-models.

Furthermore, it is worth noting that excellent SISR results play a crucial role in improving object detection and classification performance. In object detection scenarios, SISR reconstruction can restore the subtle details of the target object during the detection process, thereby improving the detection accuracy. For example, in YOLO-SASE [193], the algorithm combines the YOLO detection framework with the SRGAN network, which can improve the detection ability for infrared small targets in complex backgrounds. Similarly, in image classification, due to the low resolution and blurry images typically obtained, SISR processing can provide clearer texture and boundary information, thereby improving the accuracy of classification.

Table 4. Model comparison.

Model	Set5 PSNR/SSIM	Set14 PSNR/SSIM	Urban100 PSNR/SSIM	MOS	Training Datasets	Parameters
SRCNN [7]	30.48/0.8628	27.50/0.7513	24.52/0.7221	3.2	T91+ImageNet	57 K
VDSR [26]	31.35/0.8838	28.02/0.7680	25.18/0.7540	3.5	BSD+T91	665 K
LapSRN [40]	31.54/0.8855	28.19/0.7720	25.21/0.7560	3.6	BSD+T91	812 K
MemNet [34]	31.74/0.8893	28.26/0.7723	25.50/0.7630	3.7	BSD+T91	677 K
IDN [59]	31.82/0.8903	28.25/0.7730	25.41/0.7632	3.7	BSD+T91	678 K
RFDN [60]	32.18/0.8948	28.58/0.7812	26.04/0.7848	3.8	DIV2K	441 K
DSRN [37]	31.40/0.8830	28.07/0.7700	25.08/0.7470	3.4	T91	1.2 M
MSRN [29]	32.07/0.8903	28.60/0.7751	26.04/0.7896	3.9	DIV2K	6.3 M
CARN [38]	32.13/0.8937	28.60/0.7806	26.07/0.7837	4.0	BSD+T91+DIV2K	1.6 M
SeaNet [111]	32.33/0.8970	28.81/0.7855	26.32/0.7942	4.1	DIV2K	7.4 M
CRN [38]	32.34/0.8971	28.74/0.7855	26.44/0.7967	4.1	DIV2K	9.5 M
EDSR [27]	32.46/0.8968	28.80/0.7876	26.64/0.8033	4.2	DIV2K	43 M
RDN [194]	32.47/0.8990	28.81/0.7871	26.61/0.8028	4.2	DIV2K	22.6 M
MDCN [45]	32.48/0.8985	28.83/0.7879	26.69/0.8049	4.3	DIV2K	4.5 M
SRRFN [39]	32.56/0.8993	28.86/0.7882	26.78/0.8071	4.3	DIV2K	4.2 M
RCAN [48]	32.63/0.9002	28.87/0.7889	26.82/0.8087	4.4	DIV2K	16 M
SwinIR [69]	32.92/0.9044	29.09/0.7950	27.45/0.8254	4.5	DIV2K+Flickr2K	11.8 M

5. Potential Issues and Future Directions

5.1. Lightweight SISR for Mobile Intelligence

At present, with the increasing demand for smart devices, the requirements for lightweight SISR models are also becoming higher, which makes related research crucial. Although there has been some progress in current lightweight models, due to limitations such as computational costs, these models still cannot meet the requirements of some mobile intelligent devices and edge sensors in terms of size and performance. Therefore,

studying lightweight SISR models not only plays an important role in promoting their development but also facilitates their commercial application in sensor scenarios.

To achieve this goal, further exploration of more efficient network architectures and mechanisms is needed. It is worth noting that evaluating the hardware adaptability of these lightweight models on edge devices is also crucial, as they are commonly used in sensor scenarios, such as the NVIDIA Jetson Nano and Arduino. For classic embedded AI platforms such as the NVIDIA Jetson Nano, the hardware memory limit is 4 GB, and key metrics evaluated include the inference latency and memory usage. For the Arduino, its hardware resources are extremely limited, e.g., the Arduino Uno only has 32 KB of RAM. When evaluating the inference latency, an ultra-lightweight architecture is required. In terms of the Arduino's memory usage, its model parameters should be compressed to the KB level, and large-scale feature maps should be avoided during inference. We calculate the costs on two platforms, as listed in Table 5.

Table 5. Cost analysis of SISR model calculation.

Model	Platform	Latency	Memory Usage	Application Scenario
DSRN [37]	NVIDIA Jetson Nano	23 ms	0.8 GB	General image SR tasks
CARN [38]	NVIDIA Jetson Nano	15 ms	0.5 GB	Lightweight real-time SR
RDN [194]	NVIDIA Jetson Nano	26 ms	1.2 GB	Medium-precision image SR
MDCN [45]	NVIDIA Jetson Nano	19 ms	0.7 GB	Medium-speed image SR
SRRFN [39]	NVIDIA Jetson Nano	24 ms	1.1 GB	Near-real-time SR
RCAN [48]	NVIDIA Jetson Nano	40 ms	2.5 GB	High-precision image SR
TinySR [195]	Arduino Uno	320 ms	28 KB	Environmental cameras
MobileSR [196]	Arduino Uno	450 ms	31 KB	Low-frequency sensors

In addition, when studying SISR models, the model can be combined with compression techniques such as network binarization and network quantization to balance relevant performance aspects while meeting size requirements.

Overall, by integrating hardware evaluation into the optimization of lightweight SISR models, edge sensor devices can be better matched, providing better applicability for lightweight devices.

5.2. SISR Applicable to Multiple Scenarios

Through the existing DL-based SISR models, it can be seen that the relevant research on these models is relatively mature. However, in-depth research on these models reveals certain limitations. For example, during the training and testing phases, these models need to maintain consistent related structures, but this requirement limits their applicability, making it difficult for a single model to be widely applicable to various types of scenarios. This limitation requires the design of specialized models for different scenarios, but this often increases the costs (including manpower and material resources). To address this deficiency, it is essential to design a flexible SISR model. At the same time, when designing such SISR models, it is not only important to make them adjustable and applicable to different application scenarios but also to ensure that they have good performance.

5.3. New Loss Functions and Assessment Methods

Currently, most SISR models still rely on the L1 loss or MSE loss as their core dependencies. Although some new loss functions have been introduced, they still struggle to balance image reconstruction accuracy and perceptual quality and cannot achieve ideal results. Therefore, designing more suitable loss functions to resolve this problem remains a key focus of future research. At the same time, existing image quality evaluation methods also have obvious limitations. Some evaluation methods rely more on subjectivity, and some evaluation methods have obvious unfairness in different scenarios. Therefore, it is of great research significance to obtain a more optimal loss function and design an evaluation method that breaks through these limitations.

5.4. Mutual Promotion with High-Level Tasks

It is well known that high-level computer vision tasks (such as image classification, image segmentation, and image analysis) are highly sensitive to the quality of input images. Therefore, SISR technology is often used in the preprocessing stage to improve the image quality. Meanwhile, the quality of SR images can significantly affect the performance of these high-level tasks. As a result, integrating image SR with high-level tasks has become a major research topic in recent years. To this end, Zangeneh et al. [197] proposed a novel nonlinear coupled mapping architecture that uses two deep convolutional neural networks to map LR and HR face images into a shared space to achieve LR face recognition. Wang et al. [198] developed a dual SR learning method for semantic segmentation, integrating image SR with semantic segmentation into an end-to-end model. In addition, Xiang et al. [199] significantly enhanced the performance of high-level vision tasks by jointly reducing compression artifacts and performing SR reconstruction. Although these methods have successfully combined SR with high-level tasks and achieved good results, they mainly focus on the output results of high-level vision tasks and rarely consider using feedback from other tasks to further optimize the quality of SR images. Therefore, we suggest using the accuracy of high-level computer vision tasks as a metric to measure the quality of SR images.

5.5. Building Real SISR

Although there has been significant progress and improvement in SISR, many algorithm models still fail to achieve the expected results when applied to real-world scenarios. As more and more researchers focus on true SISR, it is expected to become the core direction of this field in the future. The key to achieving true SISR is to build large-scale, high-precision real-image datasets. The application scenarios in the real world are complex and diverse, and various scenarios may be affected by degradation. To construct this realistic dataset, on one hand, a combination of manual and technical methods can be employed, and GANs can be utilized to simulate different degradation patterns and generate finer texture details. On the other hand, blind SISR can be developed for single-image SR reconstruction techniques. Blind SISR can better learn the degradation patterns of images and reconstruct high-quality SR images based on the corresponding degradation patterns. Although blind SISR has achieved good results in image reconstruction, blind SISR methods still need to adapt to different types of scenes and degradation modes. Therefore, the development of blind SISR has important research significance for true SISR. By studying real SISR, the image quality can be improved, image details can be refined, and the image accuracy can be enhanced.

5.6. Efficient and Accurate Arbitrary-Scale SISR

In many scenarios, SISR is often used, but most DL-based SISR models have certain limitations, and existing SISR models are often limited in their size. Therefore, developing an SISR model that can adjust to any scale is of great research significance. Although there are currently some SISR methods with arbitrary scales, they often have problems such as complex usage conditions and a lack of flexibility when used. Therefore, this is a valuable research direction that can be further explored in the future.

5.7. Considering Features of Different Images

At present, there are already some models in specific fields, but these model methods do not consider different image features and are relatively simple. By utilizing different image features in specific fields, constructing accurate SISR models can effectively improve the performance of these models. Therefore, conducting in-depth research on SISR models using domain-specific images is of great significance. In the future, researching SISR models based on specific domain image features will become one of the focuses of many researchers.

5.8. System Security of SISR

In the practical application of SISR technology, system security should be taken into consideration in the future. Kwon et al. [200] proposed an optimal cluster extended intrusion tolerance system, which provides a solution for responding to DoS attacks. The system monitors the response time, CPU usage, and queue length in real time through the cluster controller. Moreover, it adjusts the number of online virtual machines in a timely manner, expands the cluster during the attack, and reduces the cluster after the attack is alleviated, which can reduce resource waste. This mechanism can maintain service quality in both sustained and single-pulse DoS scenarios. SISR systems can draw inspiration from their system security. For example, in monitoring and other SISR-dependent scenarios, similar strategies can be introduced to protect system security and ensure the stability of SR images in the face of attacks.

6. Conclusions

STSR is a key research direction in the field of computer vision, with significant research significance and application value. In this review, we first introduced and provided examples of SISR methods based on deep learning. Secondly, the relevant work on SISR was described, and its applications in specific fields were briefly introduced. Afterwards, in order to clearly demonstrate the differences in performance between different models, we listed the reconstruction results of some classic models and compared their performance. Finally, we analyzed the potential issues and future development trends in SISR. We believe that this review can provide some reference value for future research, thereby promoting the development of SISR.

Author Contributions: Conceptualization, Z.L. and S.J.; software, Z.L. and S.J.; validation, Z.L. and S.J.; formal analysis, Z.L. and S.J.; investigation, S.F. and Q.S.; resources, J.Z.; data curation, S.F. and Q.S.; writing—original draft preparation, S.F. and Q.S.; writing—review and editing, Z.L., S.J. and J.Z.; visualization, S.F.; supervision, J.Z.; project administration, Z.L. and S.J.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Major Science and Technology Projects of Longmen Laboratory (grant no. 231100220400, 231100220300); the Open Fund of the Engineering Research Center of Intelligent Swarm Systems, Ministry of Education (grant no. ZZU-CISS-2024003); the National Natural Science Foundation of China (62201200); the Aeronautical Science Foundation of China (20230001042001, 20240001042002); the China Postdoctoral Science Foundation (2023M730982);

the Scientific and Technological Project of Henan Province (252102211023); and the Key Scientific Research Projects of Universities in Henan Province (25A120005).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Duchon, C.E. Lanczos Filtering in One and Two Dimensions. *J. Appl. Meteorol.* **1979**, *18*, 1016–1022. [CrossRef]
2. Sun, J.; Xu, Z.; Shum, H.Y. Image super-resolution using gradient profile prior. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
3. Kim, K.I.; Kwon, Y. Single-Image Super-Resolution Using Sparse Regression and Natural Image Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1127–1133. [PubMed]
4. Chang, H.; Yeung, D.Y.; Xiong, Y. Super-resolution through neighbor embedding. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; IEEE: Piscataway, NJ, USA, 2004; Volume 1.
5. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873. [CrossRef]
6. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
7. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part IV 13; Springer: Berlin/Heidelberg, Germany, 2014; pp. 184–199.
8. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [CrossRef]
9. Collobert, R.; Weston, J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 5–9 July 2008; pp. 160–167.
10. Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 416–423.
11. Agustsson, E.; Timofte, R. Ntire 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 126–135.
12. Timofte, R.; Agustsson, E.; Gool, E.A. NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1110–1121.
13. Bevilacqua, M.; Roumy, A.; Guillemot, C.; line Alberi Morel, M. Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 791–799.
14. Zeyde, R.; Elad, M.; Protter, M. On Single Image Scale-Up Using Sparse-Representations. In Proceedings of the International Conference on Curves and Surfaces, Avignon, France, 24–30 June 2010; pp. 711–730.
15. Huang, J.B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5197–5206.
16. Matsui, Y.; Ito, K.; Aramaki, Y.; Fujimoto, A.; Ogawa, T.; Yamasaki, T.; Aizawa, K. Sketch-based manga retrieval using manga109 dataset. *Multimed. Tools Appl.* **2017**, *76*, 21811–21838. [CrossRef]
17. Belharbi, S.; Whitford, M.K.; Hoang, P.; Murtaza, S.; McCaffrey, L.; Granger, E. SR-CACO-2: A dataset for confocal fluorescence microscopy image super-resolution. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 59948–59983.
18. Shocher, A.; Cohen, N.; Irani, M. "zero-shot" super-resolution using deep internal learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3118–3126.
19. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
20. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*.

21. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
22. Fuoli, D.; Van Gool, L.; Timofte, R. Fourier space losses for efficient perceptual image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 2360–2369.
23. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
24. He, K.; Sun, J. Convolutional neural networks at constrained time cost. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5353–5360.
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
26. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
27. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
28. Lu, Z.; Tu, Z.; Zhang, Z.; Jiang, C.; Chen, P. Single Image Super-Resolution Reconstruction Based on Multilayer Residual Networks With Attention Mechanisms. *IEEE Access* **2025**, *13*, 73794–73802. [CrossRef]
29. Li, J.; Fang, F.; Mei, K.; Zhang, G. Multi-scale residual network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Munich, Germany, 8–14 September 2018; pp. 517–532.
30. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Munich, Germany, 8–14 September 2018; pp. 286–301.
31. Inderjeet; Sahambi, J.S. GAMSRN: Global Attention Multi-Scale Residual Network for Single-Image Super-Resolution and Low-Light Enhancement. In Proceedings of the 2025 National Conference on Communications (NCC), New Delhi, India, 6–9 March 2025; pp. 1–6.
32. Tong, T.; Li, G.; Liu, X.; Gao, Q. Image super-resolution using dense skip connections. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4799–4807.
33. Purohit, K.; Mandal, S.; Rajagopalan, A. Deep Networks for Image and Video Super-Resolution. *arXiv* **2022**, arXiv:2201.11996. [CrossRef]
34. Tai, Y.; Yang, J.; Liu, X.; Xu, C. Memnet: A persistent memory network for image restoration. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4539–4547.
35. Mei, K.; Jiang, A.; Li, J.; Liu, B.; Ye, J.; Wang, M. Deep residual refining based pseudo-multi-frame network for effective single image super-resolution. *IET Image Process.* **2019**, *13*, 591–599. [CrossRef]
36. Shen, M.; Yu, P.; Wang, R.; Yang, J.; Xue, L.; Hu, M. Multipath feedforward network for single image super-resolution. *Multimed. Tools Appl.* **2019**, *78*, 19621–19640. [CrossRef]
37. Han, W.; Chang, S.; Liu, D.; Yu, M.; Witbrock, M.; Huang, T.S. Image super-resolution via dual-state recurrent networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
38. Ahn, N.; Kang, B.; Sohn, K.A. Fast, accurate, and lightweight super-resolution with cascading residual network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Munich, Germany, 8–14 September 2018; pp. 252–268.
39. Li, J.; Yuan, Y.; Mei, K.; Fang, F. Lightweight and accurate recursive fractal network for image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019; pp. 3814–3823.
40. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 624–632.
41. Wang, Y.; Perazzi, F.; McWilliams, B.; Sorkine-Hornung, A.; Sorkine-Hornung, O.; Schroers, C. A fully progressive approach to single-image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 864–873.
42. Wong, L.; Zhao, D.; Wan, S.; Zhang, B. Perceptual image super-resolution with progressive adversarial network. *arXiv* **2020**, arXiv:2003.03756. [CrossRef]
43. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
44. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

45. Li, J.; Fang, F.; Li, J.; Mei, K.; Zhang, G. MDCN: Multi-scale dense cross network for image super-resolution. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 2547–2561. [CrossRef]
46. Qin, J.; Huang, Y.; Wen, W. Multi-scale feature fusion residual network for single image super-resolution. *Neurocomputing* **2020**, *379*, 334–342. [CrossRef]
47. Chang, C.Y.; Chien, S.Y. Multi-scale dense network for single-image super-resolution. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1742–1746.
48. Cao, F.; Liu, H. Single image super-resolution via multi-scale residual channel attention network. *Neurocomputing* **2019**, *358*, 424–436. [CrossRef]
49. Yazıcı, Z.A.; Öksüz, İ.; Ekenel, H.K. GLIMS: Attention-guided lightweight multi-scale hybrid network for volumetric semantic segmentation. *Image Vis. Comput.* **2024**, *146*, 105055. [CrossRef]
50. Li, Y.; Zhao, X.; Zhang, X.; Yang, Y.; Li, T. Long-Range Multi-Scale Fusion for Efficient Single Image Super-Resolution. In Proceedings of the ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 6–11 April 2025; pp. 1–5.
51. Mei, K.; Jiang, A.; Li, J.; Ye, J.; Wang, M. An effective single-image super-resolution model using squeeze-and-excitation networks. In Proceedings of the Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, 13–16 December 2018; Proceedings, Part VI 25; Springer: Berlin/Heidelberg, Germany, 2018; pp. 542–553.
52. Wang, Y.; Li, Y.; Wang, G.; Liu, X. Multi-scale attention network for single image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–18 June 2024; pp. 5950–5960.
53. Wan, C.; Yu, H.; Li, Z.; Chen, Y.; Zou, Y.; Liu, Y.; Yin, X.; Zuo, K. Swift parameter-free attention network for efficient super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–18 June 2024; pp. 6246–6256.
54. Ahmadi, M.; Raghavan, J. Novel Spatial-Edge Residual Attention Model for Face Super-Resolution Enhancement Nouveau modèle d'attention résiduelle spatiale et sur les contours pour l'amélioration de la super-résolution des visages. *IEEE Can. J. Electr. Comput. Eng.* **2025**, *48*, 294–304. [CrossRef]
55. Yu, C.; Pei, H. Super-Resolution Reconstruction Method of Face Image Based on Attention Mechanism. *IEEE Access* **2025**, *13*, 121250–121260. [CrossRef]
56. Li, Z.; Yang, J.; Liu, Z.; Yang, X.; Jeon, G.; Wu, W. Feedback network for image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3867–3876.
57. Barman, T.; Deka, B. A Deep Learning-Based Joint Image Super-Resolution and Deblurring Framework. *IEEE Trans. Artif. Intell.* **2023**, *5*, 3160–3173. [CrossRef]
58. Fan, Y.; Jiang, M.; Liu, Z.; Zhao, Z.; Zheng, L. Lightweight Image Super-Resolution Reconstruction Guided by Gated Feedback and LatticeFormer. *IEEE Access* **2025**, *13*, 53214–53226. [CrossRef]
59. Hui, Z.; Wang, X.; Gao, X. Fast and accurate single image super-resolution via information distillation network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 723–731.
60. Liu, J.; Tang, J.; Wu, G. Residual feature distillation network for lightweight image super-resolution. In Proceedings of the Computer Vision–ECCV 2020 Workshops, Glasgow, UK, 23–28 August 2020; Proceedings, Part III 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 41–55.
61. Zhou, L.; Cai, H.; Gu, J.; Li, Z.; Liu, Y.; Chen, X.; Qiao, Y.; Dong, C. Efficient image super-resolution using vast-receptive-field attention. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 256–272.
62. Du, Z.; Liu, D.; Liu, J.; Tang, J.; Wu, G.; Fu, L. Fast and memory-efficient network towards efficient image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 853–862.
63. Lee, W.; Lee, J.; Kim, D.; Ham, B. Learning with privileged information for efficient image super-resolution. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part XXIV 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 465–482.
64. Lin, C.; Chen, Y.; Zou, X.; Deng, X.; Dai, F.; You, J.; Xiao, J. An unconstrained palmprint region of interest extraction method based on lightweight networks. *PLoS ONE* **2024**, *19*, e0307822. [CrossRef] [PubMed]
65. Sun, L.; Pan, J.; Tang, J. Shufflemixer: An efficient convnet for image super-resolution. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 17314–17326.
66. Chen, Z.; Zhang, Y.; Gu, J.; Kong, L.; Yang, X. Recursive generalization transformer for image super-resolution. *arXiv* **2023**, arXiv:2303.06373.

67. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
68. Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Xu, C.; Gao, W. Pre-trained image processing transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12299–12310.
69. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. SwinIR: Image restoration using swin transformer. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 1833–1844.
70. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739.
71. Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; Dong, C. Activating more pixels in image super-resolution transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 22367–22377.
72. Jiang, N.; Zhao, W.; Wang, H.; Luo, H.; Chen, Z.; Zhu, J. Lightweight super-resolution generative adversarial network for SAR images. *Remote Sens.* **2024**, *16*, 1788. [CrossRef]
73. Guo, F.; Feng, Q.; Yang, S.; Yang, W. CMTNet: Convolutional Meets Transformer Network for Hyperspectral Images Classification. *arXiv* **2024**, arXiv:2406.14080. [CrossRef]
74. Zhang, D.; Liang, S.; He, T.; Shao, J.; Qin, K. CVIformer: Cross-View Interactive Transformer for Efficient Stereoscopic Image Super-Resolution. *IEEE Trans. Emerg. Top. Comput. Intell.* **2025**, *9*, 1107–1118. [CrossRef]
75. He, Y.; He, Y. MPSI: Mamba enhancement model for pixel-wise sequential interaction Image Super-Resolution. *arXiv* **2024**, arXiv:2412.07222. [CrossRef]
76. Wang, X.; Li, J.; Li, J.; Wang, S.; Yan, L.; Xu, Y. A Collaborative Network of Mamba and CNN for Lightweight Image Super-Resolution. *IEEE Trans. Consum. Electron.* **2025**, *71*, 3591–3604. [CrossRef]
77. Aalishah, R.; Navardi, M.; Mohsenin, T. MambaLiteSR: Image Super-Resolution with Low-Rank Mamba Using Knowledge Distillation. In Proceedings of the 2025 26th International Symposium on Quality Electronic Design (ISQED), San Francisco, CA, USA, 23–25 April 2025; pp. 1–8.
78. Jiang, K.; Yang, M.; Xiao, Y.; Wu, J.; Wang, G.; Feng, X.; Jiang, J. Rep-Mamba: Re-Parameterization in Vision Mamba for Lightweight Remote Sensing Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5637012. [CrossRef]
79. Hsu, C.C.; Lee, C.M.; Chou, Y.S. Drct: Saving image super-resolution away from information bottleneck. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–18 June 2024; pp. 6133–6142.
80. Blau, Y.; Michaeli, T. The perception-distortion tradeoff. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6228–6237.
81. Xue, D.; Herranz, L.; Corral, J.V.; Zhang, Y. Burst perception-distortion tradeoff: Analysis and evaluation. In Proceedings of the ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5.
82. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
83. Gatys, L.; Ecker, A.S.; Bethge, M. Texture synthesis using convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*. [CrossRef]
84. Gatys, L.A.; Ecker, A.S.; Bethge, M. A neural algorithm of artistic style. *arXiv* **2015**, arXiv:1508.06576. [CrossRef]
85. Ople, J.J.M.; Tan, D.S.; Azcarraga, A.; Yang, C.L.; Hua, K.L. Super-resolution by image enhancement using texture transfer. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 953–957.
86. Rad, M.S.; Bozorgtabar, B.; Marti, U.V.; Basler, M.; Ekenel, H.K.; Thiran, J.P. Srobb: Targeted perceptual loss for single image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2710–2719.
87. Guo, J.; Lv, F.; Shen, J.; Liu, J.; Wang, M. An improved generative adversarial network for remote sensing image super-resolution. *IET Image Process.* **2023**, *17*, 1852–1863. [CrossRef]
88. Song, J.; Yi, H.; Xu, W.; Li, X.; Li, B.; Liu, Y. Dual perceptual loss for single image super-resolution using esrgan. *arXiv* **2022**, arXiv:2201.06383. [CrossRef]
89. Korkmaz, C.; Tekalp, A.M.; Dogan, Z. Training generative image super-resolution models by wavelet-domain losses enables better control of artifacts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 5926–5936.

90. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Munich, Germany, 8–14 September 2018; pp. 63–79.
91. Wang, Y.; Lin, C.; Luo, D.; Tai, Y.; Zhang, Z.; Xie, Y. High-resolution gan inversion for degraded images in large diverse datasets. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 2716–2723.
92. Yang, S.; Yuan, J.; Tang, W.; Liu, R.; Yuan, M.; Huang, J. Super-Resolution Reconstruction of Medical Images Based on Improved Generative Adversarial Networks. In Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Computer Information Technology (AICIT), Yichang, China, 20–22 September 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–4.
93. Huang, J.; Li, K.; Jia, J.; Wang, X. Single Image Super-Resolution Through Image Pixel Information Clustering and Generative Adversarial Network. *Big Data Min. Anal.* **2025**, *8*, 1044–1059. [CrossRef]
94. Bulat, A.; Yang, J.; Tzimiropoulos, G. To learn image super-resolution, use a gan to learn how to do image degradation first. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Munich, Germany, 8–14 September 2018; pp. 185–200.
95. Yuan, Y.; Liu, S.; Zhang, J.; Zhang, Y.; Dong, C.; Lin, L. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 701–710.
96. Guo, Y.; Chen, J.; Wang, J.; Chen, Q.; Cao, J.; Deng, Z.; Xu, Y.; Tan, M. Closed-loop matters: Dual regression networks for single image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5407–5416.
97. Yang, Y.; Qi, Y.; Qi, S. Relation-consistency graph convolutional network for image super-resolution. *Vis. Comput.* **2024**, *40*, 619–635. [CrossRef]
98. Li, H.; Yang, Y.; Chang, M.; Chen, S.; Feng, H.; Xu, Z.; Li, Q.; Chen, Y. Srdiff: Single image super-resolution with diffusion probabilistic models. *Neurocomputing* **2022**, *479*, 47–59. [CrossRef]
99. Saharia, C.; Ho, J.; Chan, W.; Salimans, T.; Fleet, D.J.; Norouzi, M. Image super-resolution via iterative refinement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4713–4726. [CrossRef]
100. Gao, S.; Liu, X.; Zeng, B.; Xu, S.; Li, Y.; Luo, X.; Liu, J.; Zhen, X.; Zhang, B. Implicit diffusion models for continuous super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 10021–10030.
101. Wang, Z.; Zhang, Z.; Zhang, X.; Zheng, H.; Zhou, M.; Zhang, Y.; Wang, Y. Dr2: Diffusion-based robust degradation remover for blind face restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 1704–1713.
102. Wang, J.; Yue, Z.; Zhou, S.; Chan, K.C.; Loy, C.C. Exploiting diffusion prior for real-world image super-resolution. *Int. J. Comput. Vis.* **2024**, *132*, 5929–5949. [CrossRef]
103. Lin, X.; He, J.; Chen, Z.; Lyu, Z.; Dai, B.; Yu, F.; Qiao, Y.; Ouyang, W.; Dong, C. Diffbir: Toward blind image restoration with generative diffusion prior. In Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024; Springer: Cham, Switzerland, 2024; pp. 430–448.
104. Xia, B.; Zhang, Y.; Wang, S.; Wang, Y.; Wu, X.; Tian, Y.; Yang, W.; Van Gool, L. Diffir: Efficient diffusion model for image restoration. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 13095–13105.
105. Wu, H.; Mo, J.; Sun, X.; Ma, J. Latent Diffusion, Implicit Amplification: Efficient Continuous-Scale Super-Resolution for Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1–17. [CrossRef]
106. Gendy, G.; He, G.; Sabor, N. Diffusion models for image super-resolution: State-of-the-art and future directions. *Neurocomputing* **2025**, *617*, 128911. [CrossRef]
107. Zontak, M.; Irani, M. Internal statistics of a single natural image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 977–984.
108. Shaham, T.R.; Dekel, T.; Michaeli, T. Singan: Learning a generative model from a single natural image. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4570–4580.
109. Wang, Y.; Ying, X.; Wang, L.; Yang, J.; An, W.; Guo, Y. Symmetric parallax attention for stereo image super-resolution. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; pp. 766–775.
110. Yang, W.; Feng, J.; Yang, J.; Zhao, F.; Liu, J.; Guo, Z.; Yan, S. Deep edge guided recurrent residual learning for image super-resolution. *IEEE Trans. Image Process.* **2017**, *26*, 5895–5907. [CrossRef]
111. Fang, F.; Li, J.; Zeng, T. Soft-edge assisted network for single image super-resolution. *IEEE Trans. Image Process.* **2020**, *29*, 4656–4668. [CrossRef] [PubMed]

112. Wang, X.; Yu, K.; Dong, C.; Loy, C.C. Recovering realistic texture in image super-resolution by deep spatial feature transform. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 606–615.
113. Ma, C.; Rao, Y.; Cheng, Y.; Chen, C.; Lu, J.; Zhou, J. Structure-preserving super resolution with gradient guidance. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 7769–7778.
114. Chen, C.; Shi, X.; Qin, Y.; Li, X.; Han, X.; Yang, T.; Guo, S. Real-world blind super-resolution via feature matching with implicit high-resolution priors. In Proceedings of the 30th ACM International Conference on Multimedia, Lisboa, Portugal, 10–14 October 2022; pp. 1329–1338.
115. Yu, J.; Li, X.; Koh, J.Y.; Zhang, H.; Pang, R.; Qin, J.; Ku, A.; Xu, Y.; Baldridge, J.; Wu, Y. Vector-quantized image modeling with improved vqgan. *arXiv* **2021**, arXiv:2110.04627.
116. Huang, D.; Song, J.; Huang, X.; Hu, Z.; Zeng, H. Multi-Modal Prior-Guided Diffusion Model for Blind Image Super-Resolution. *IEEE Signal Process. Lett.* **2025**, *32*, 316–320. [CrossRef]
117. Yue, H.; Sun, X.; Yang, J.; Wu, F. Landmark image super-resolution by retrieving web images. *IEEE Trans. Image Process.* **2013**, *22*, 4865–4878. [CrossRef] [PubMed]
118. Zheng, H.; Ji, M.; Wang, H.; Liu, Y.; Fang, L. Crossnet: An end-to-end reference-based super resolution network using cross-scale warping. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Munich, Germany, 8–14 September 2018; pp. 88–104.
119. Zhang, Z.; Wang, Z.; Lin, Z.; Qi, H. Image super-resolution by neural texture transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7982–7991.
120. Yang, F.; Yang, H.; Fu, J.; Lu, H.; Guo, B. Learning texture transformer network for image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5791–5800.
121. Cong, R.; Liao, R.; Li, F.; Sheng, R.; Bai, H.; Wan, R.; Kwong, S.; Zhang, W. Reference-Based Iterative Interaction with P2-Matching for Stereo Image Super-Resolution. *IEEE Trans. Image Process.* **2025**, *34*, 3779–3789. [CrossRef]
122. Gao, Q.; Zhao, Y.; Li, G.; Tong, T. Image super-resolution using knowledge distillation. In Proceedings of the Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 527–541.
123. Luo, X.; Liang, Q.; Liu, D.; Qu, Y. Boosting lightweight single image super-resolution via joint-distillation. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual Event China, 20–24 October 2021; pp. 1535–1543.
124. Wang, X.; Li, Y.; Zhang, H.; Shan, Y. Towards real-world blind face restoration with generative facial prior. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 9168–9178.
125. Li, Y.; Fan, Y.; Xiang, X.; Demandolx, D.; Ranjan, R.; Timofte, R.; Van Gool, L. Efficient and explicit modelling of image hierarchies for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 18278–18289.
126. Liu, A.; Liu, Y.; Gu, J.; Qiao, Y.; Dong, C. Blind image super-resolution: A survey and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 5461–5480. [CrossRef]
127. Bell-Kligler, S.; Shocher, A.; Irani, M. Blind super-resolution kernel estimation using an internal-gan. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
128. Gu, J.; Lu, H.; Zuo, W.; Dong, C. Blind super-resolution with iterative kernel correction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1604–1613.
129. Luo, Z.; Huang, Y.; Li, S.; Wang, L.; Tan, T. Unfolding the alternating optimization for blind super resolution. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 5632–5643.
130. Sun, H.; Yuan, Y.; Su, L.; Shao, H. Learning correction errors via frequency-self attention for blind image super-resolution. In Proceedings of the 2024 9th International Conference on Image, Vision and Computing (ICIVC), Suzhou, China, 15–17 July 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 384–392.
131. Li, F.; Wu, Y.; Liang, Z.; Cong, R.; Bai, H.; Zhao, Y.; Wang, M. Blinddiff: Empowering degradation modelling in diffusion models for blind image super-resolution. *arXiv* **2024**, arXiv:2403.10211. [CrossRef]
132. Ohayon, G.; Elad, M.; Michaeli, T. Perceptual fairness in image restoration. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 70259–70312.
133. Ates, H.F.; Yildirim, S.; Gunturk, B.K. Deep learning-based blind image super-resolution with iterative kernel reconstruction and noise estimation. *Comput. Vis. Image Underst.* **2023**, *233*, 103718. [CrossRef]
134. Liu, Q.; Zhuang, C.; Gao, P.; Qin, J. Cdformer: When degradation prediction embraces diffusion model for blind image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 7455–7464.
135. Wang, L.; Wang, Y.; Dong, X.; Xu, Q.; Yang, J.; An, W.; Guo, Y. Unsupervised degradation representation learning for blind super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10581–10590.

136. Yang, F.; Yang, H.; Zeng, Y.; Fu, J.; Lu, H. Degradation-guided meta-restoration network for blind super-resolution. *arXiv* **2022**, arXiv:2207.00943.
137. Chen, G.Y.; Weng, W.D.; Su, J.N.; Gan, M.; Chen, C.P. Dynamic degradation intensity estimation for adaptive blind super-resolution: A novel approach and benchmark dataset. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *34*, 4762–4772. [CrossRef]
138. Liu, Q.; Gao, P.; Han, K.; Liu, N.; Xiang, W. Degradation-aware self-attention based transformer for blind image super-resolution. *IEEE Trans. Multimed.* **2024**, *26*, 7516–7528. [CrossRef]
139. Yuan, J.; Ma, J.; Wang, B.; Hu, W. Content-Decoupled Contrastive Learning-Based Implicit Degradation Modeling for Blind Image Super-Resolution. *IEEE Trans. Image Process.* **2025**, *34*, 4751–4766. [CrossRef]
140. Wang, S.; Zhang, M.; Miao, M. The super-resolution reconstruction algorithm of multi-scale dilated convolution residual network. *Front. Neurorobotics* **2024**, *18*, 1436052. [CrossRef]
141. Li, G.; Xiao, H.; Liang, D.; Ling, B.W.K. Multi-scale cross-fusion for arbitrary scale image super resolution. *Multimed. Tools Appl.* **2024**, *83*, 79805–79814. [CrossRef]
142. Wei, J.; Yang, G.; Wei, W.; Liu, A.; Chen, X. Multi-Contrast MRI Arbitrary-Scale Super-Resolution via Dynamic Implicit Network. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 8973–8988. [CrossRef]
143. Chu, X.; Chen, L.; Yu, W. Nafssr: Stereo image super-resolution using nafnet. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 19–20 June 2022; pp. 1239–1248.
144. Xu, Z.; Tang, Y.; Xu, B.; Li, Q. NeurOp-Diff: Continuous Remote Sensing Image Super-Resolution via Neural Operator Diffusion. *arXiv* **2025**, arXiv:2501.09054.
145. Yu, Z.; Chen, L.; Zeng, Z.; Yang, K.; Luo, S.; Chen, S.; Zhong, C. LGFN: Lightweight Light Field Image Super-Resolution using Local Convolution Modulation and Global Attention Feature Extraction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–18 June 2024; pp. 6712–6721.
146. Zhang, M.; Zhang, C.; Zhang, Q.; Guo, J.; Gao, X.; Zhang, J. Essaformer: Efficient transformer for hyperspectral image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 23073–23084.
147. Georgescu, M.I.; Ionescu, R.T.; Miron, A.I.; Savencu, O.; Ristea, N.C.; Verga, N.; Khan, F.S. Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 2195–2205.
148. Jeon, D.S.; Baek, S.H.; Choi, I.; Kim, M.H. Enhancing the spatial resolution of stereo images using a parallax prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1721–1730.
149. Wang, L.; Guo, Y.; Wang, Y.; Liang, Z.; Lin, Z.; Yang, J.; An, W. Parallax attention for unsupervised stereo correspondence learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 2108–2125. [CrossRef] [PubMed]
150. Wang, L.; Wang, Y.; Liang, Z.; Lin, Z.; Yang, J.; An, W.; Guo, Y. Learning parallax attention for stereo image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12250–12259.
151. Wang, Y.; Wang, L.; Yang, J.; An, W.; Guo, Y. Flickr1024: A large-scale dataset for stereo image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019; pp. 3852–3857.
152. Ying, X.; Wang, Y.; Wang, L.; Sheng, W.; An, W.; Guo, Y. A stereo attention module for stereo image super-resolution. *IEEE Signal Process. Lett.* **2020**, *27*, 496–500. [CrossRef]
153. Dai, Q.; Li, J.; Yi, Q.; Fang, F.; Zhang, G. Feedback network for mutually boosted stereo image super-resolution and disparity estimation. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual Event China, 20–24 October 2021; pp. 1985–1993.
154. Chen, L.; Chu, X.; Zhang, X.; Sun, J. Simple baselines for image restoration. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Cham, Switzerland, 2022; pp. 17–33.
155. Lin, J.; Yin, L.; Wang, Y. Steformer: Efficient stereo image super-resolution with transformer. *IEEE Trans. Multimed.* **2023**, *25*, 8396–8407. [CrossRef]
156. Cao, W.; Lei, X.; Jiang, Y.; Bai, Z.; Qian, X. Leveraging more information for blind stereo super-resolution via large-window cross-attention. *Appl. Soft Comput.* **2025**, *168*, 112492. [CrossRef]
157. Haut Hurtado, J.M.; Fernández Beltrán, R.; Paoletti Ávila, M.E.; Plaza Miguel, J.; Plaza, A.; Pla, F. A new deep generative network for unsupervised remote sensing single-image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6792–6810. [CrossRef]
158. Gu, J.; Sun, X.; Zhang, Y.; Fu, K.; Wang, L. Deep residual squeeze and excitation network for remote sensing image super-resolution. *Remote Sens.* **2019**, *11*, 1817. [CrossRef]
159. Zhang, D.; Shao, J.; Li, X.; Shen, H.T. Remote sensing image super-resolution via mixed high-order attention network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5183–5196. [CrossRef]

160. Dong, X.; Wang, L.; Sun, X.; Jia, X.; Gao, L.; Zhang, B. Remote sensing image super-resolution using second-order multi-scale networks. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 3473–3485. [CrossRef]
161. Lei, S.; Shi, Z. Hybrid-scale self-similarity exploitation for remote sensing image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5401410. [CrossRef]
162. Wang, Y.; Shao, Z.; Lu, T.; Wu, C.; Wang, J. Remote sensing image super-resolution via multiscale enhancement network. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 5000905. [CrossRef]
163. Liu, Z.; Feng, R.; Wang, L.; Han, W.; Zeng, T. Dual learning-based graph neural network for remote sensing image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5628614. [CrossRef]
164. Zhang, Z.; Hao, X.; Li, J.; Pang, J. ESC-MISR: Enhancing Spatial Correlations for Multi-image Super-Resolution in Remote Sensing. In Proceedings of the International Conference on Multimedia Modeling, Nara, Japan, 8–10 January 2025; Springer: Berlin/Heidelberg, Germany, 2025; pp. 373–387.
165. Wang, Y.; Wang, L.; Yang, J.; An, W.; Yu, J.; Guo, Y. Spatial-angular interaction for light field image super-resolution. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part XXIII 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 290–308.
166. Yoon, Y.; Jeon, H.G.; Yoo, D.; Lee, J.Y.; Kweon, I.S. Light-field image super-resolution using convolutional neural network. *IEEE Signal Process. Lett.* **2017**, *24*, 848–852. [CrossRef]
167. Wang, Y.; Liu, F.; Zhang, K.; Hou, G.; Sun, Z.; Tan, T. LFNet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution. *IEEE Trans. Image Process.* **2018**, *27*, 4274–4286. [CrossRef]
168. Liang, Z.; Wang, Y.; Wang, L.; Yang, J.; Zhou, S. Light field image super-resolution with transformers. *IEEE Signal Process. Lett.* **2022**, *29*, 563–567. [CrossRef]
169. Wang, Y.; Liang, Z.; Wang, L.; Yang, J.; An, W.; Guo, Y. Real-World Light Field Image Super-Resolution Via Degradation Modulation. *IEEE Trans. Neural Netw. Learn. Syst.* **2025**, *36*, 5559–5573. [CrossRef]
170. Zhou, E.; Fan, H.; Cao, Z.; Jiang, Y.; Yin, Q. Learning face hallucination in the wild. In Proceedings of the AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; Volume 29.
171. Yu, X.; Porikli, F. Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3760–3768.
172. Dogan, B.; Gu, S.; Timofte, R. Exemplar guided face image super-resolution without facial landmarks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 1814–1823.
173. Zhang, K.; Zhang, Z.; Cheng, C.W.; Hsu, W.H.; Qiao, Y.; Liu, W.; Zhang, T. Super-identity convolutional neural network for face hallucination. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Austin, TX, USA, 25–30 January 2018; pp. 183–198.
174. Gao, G.; Tang, L.; Wu, F.; Lu, H.; Yang, J. JDSR-GAN: Constructing an efficient joint learning network for masked face super-resolution. *IEEE Trans. Multimed.* **2023**, *25*, 1505–1512. [CrossRef]
175. Qiu, T.; Yan, Y. DPHNet: Dual-Path Hybrid Network for Blurry Face Image Super-Resolution. *IEEE Access* **2025**, *13*, 56607–56615. [CrossRef]
176. Gu, Y.; Wang, X.; Xie, L.; Dong, C.; Li, G.; Shan, Y.; Cheng, M.M. Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Cham, Switzerland, 2022; pp. 126–143.
177. Zhou, S.; Chan, K.; Li, C.; Loy, C.C. Towards robust blind face restoration with codebook lookup transformer. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 30599–30611.
178. Zhu, F.; Zhu, J.; Chu, W.; Zhang, X.; Ji, X.; Wang, C.; Tai, Y. Blind face restoration via integrating face shape and generative priors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 7662–7671.
179. Rickard, L.J.; Basedow, R.W.; Zalewski, E.F.; Silverglate, P.R.; Landers, M. HYDICE: An airborne system for hyperspectral imaging. In Proceedings of the Imaging Spectrometry of the Terrestrial Environment, Orlando, FL, USA, 11–16 April 1993; SPIE: Bellingham, WA, USA, 1993; Volume 1937, pp. 173–179.
180. Wu, L.; Wang, H.; Zhang, T. A multiscale 3D convolution with context attention network for hyperspectral image classification. *Earth Sci. Inform.* **2022**, *15*, 2553–2569. [CrossRef]
181. Wang, B.; Chen, J.; Wang, H.; Tang, Y.; Chen, J.; Jiang, Y. A spectral and spatial transformer for hyperspectral remote sensing image super-resolution. *Int. J. Digit. Earth* **2024**, *17*, 2313102. [CrossRef]
182. Dang, L.; Weng, L.; Hou, Y.; Zuo, X.; Liu, Y. Double-branch feature fusion transformer for hyperspectral image classification. *Sci. Rep.* **2023**, *13*, 272. [CrossRef] [PubMed]
183. Fei, M.; Fang, W.; Shuai, H. Adaptive deep prior for hyperspectral image super-resolution. *Laser Technol.* **2024**, *48*, 491–498.

184. Jia, Y.; Xie, Y.; An, P.; Tian, Z.; Hua, X. DiffHSR: Unleashing Diffusion Priors in Hyperspectral Image Super-Resolution. *IEEE Signal Process. Lett.* **2025**, *32*, 236–240. [CrossRef]
185. Chen, Y.; Shi, F.; Christodoulou, A.G.; Xie, Y.; Zhou, Z.; Li, D. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network. In Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention, Granada, Spain, 16–20 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 91–99.
186. Wang, Y.; Teng, Q.; He, X.; Feng, J.; Zhang, T. CT-image of rock samples super resolution using 3D convolutional neural network. *Comput. Geosci.* **2019**, *133*, 104314. [CrossRef]
187. Zhao, X.; Zhang, Y.; Zhang, T.; Zou, X. Channel splitting network for single MR image super-resolution. *IEEE Trans. Image Process.* **2019**, *28*, 5649–5662. [CrossRef]
188. Peng, C.; Lin, W.A.; Liao, H.; Chellappa, R.; Zhou, S.K. Saint: Spatially aware interpolation network for medical slice synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 7750–7759.
189. Feng, C.M.; Yan, Y.; Fu, H.; Chen, L.; Xu, Y. Task transformer network for joint MRI reconstruction and super-resolution. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021; Proceedings, Part VI 24; Springer: Berlin/Heidelberg, Germany, 2021; pp. 307–317.
190. Sheibanifard, A.; Yu, H. A novel implicit neural representation for volume data. *Appl. Sci.* **2023**, *13*, 3242. [CrossRef]
191. Kui, X.; Ji, Z.; Zou, B.; Li, Y.; Dai, Y.; Chen, L.; Vera, P.; Ruan, S. Iterative Collaboration Network Guided by Reconstruction Prior for Medical Image Super-Resolution. *IEEE Trans. Comput. Imaging* **2025**, *11*, 827–838. [CrossRef]
192. Li, Y.; Hao, W.; Zeng, H.; Wang, L.; Xu, J.; Routray, S.; Jhaveri, R.H.; Gadekallu, T.R. Cross-Scale Texture Supplementation for Reference-based Medical Image Super-Resolution. *IEEE J. Biomed. Health Inform.* **2025**, 1–15. [CrossRef] [PubMed]
193. Zhou, X.; Jiang, L.; Hu, C.; Lei, S.; Zhang, T.; Mou, X. YOLO-SASE: An Improved YOLO Algorithm for the Small Targets Detection in Complex Backgrounds. *Sensors* **2022**, *22*, 4600. [CrossRef]
194. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2472–2481.
195. Dong, L.; Fan, Q.; Yu, Y.; Zhang, Q.; Chen, J.; Luo, Y.; Zou, C. TinySR: Pruning Diffusion for Real-World Image Super-Resolution. *arXiv* **2025**, arXiv:2508.17434.
196. Zhang, L.; Li, H.; Liu, X.; Niu, J.; Wu, J. MobileSR: Efficient Convolutional Neural Network for Super-resolution. In Proceedings of the GLOBECOM 2020-2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
197. Zangeneh, E.; Rahmati, M.; Mohsenzadeh, Y. Low resolution face recognition using a two-branch deep convolutional neural network architecture. *Expert Syst. Appl.* **2020**, *139*, 112854. [CrossRef]
198. Wang, L.; Li, D.; Zhu, Y.; Tian, L.; Shan, Y. Dual super-resolution learning for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3774–3783.
199. Xiang, X.; Lin, Q.; Allebach, J.P. Boosting high-level vision with joint compression artifacts reduction and super-resolution. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 2390–2397.
200. Kwon, H.; Kim, Y.; Yoon, H.; Choi, D. Optimal Cluster Expansion-Based Intrusion Tolerant System to Prevent Denial of Service Attacks. *Appl. Sci.* **2017**, *7*, 1186. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Sensors Editorial Office
E-mail: sensors@mdpi.com
www.mdpi.com/journal/sensors



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editors. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editors and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.





Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-9224-2