

Article

Fine-Scale *Spartina alterniflora* Mapping Using Advanced Deep Learning and High-Resolution UAV Imagery

Zegang Chen ^{1,2}, Yixing Tang ², Yulong Hu ³, Bin Chu ^{4,5} , Zhong Long ^{1,2}, Dong Xie ², Yunfei Zhang ² 
and Zhipan Wang ^{1,2,*} 

¹ Ecology Geological Survey and Monitoring Institute of Hunan Province, Changsha 410119, China; chenzegang@hnsdzysts.com (Z.C.); longzhong@hnsdzysts.com (Z.L.)

² School of Aerospace Engineering, Changsha University of Science and Technology, Changsha 410114, China; 202301160106@csust.edu.cn (Y.T.); 1294187544@csust.edu.cn (D.X.); zhang.yunfei@csust.edu.cn (Y.Z.)

³ China Transport Telecommunications & Information Center, Beijing 100010, China; hyulong1987@163.com

⁴ School of Electronic Information, Wuhan University, Wuhan 430072, China; chubin@whu.edu.cn

⁵ Hunan Institute of Geomatics Sciences and Technology, Changsha 410007, China

* Correspondence: wangzhp25@csust.edu.cn

Highlights

What are the main findings?

- We construct UAV-SaSeg, the first high-resolution UAV semantic segmentation dataset tailored for complicated coastal invasive species scenarios.
- DINOsegNext achieves state-of-the-art performance for *Spartina alterniflora* mapping, with an IoU of 0.7420 and an F1-Score of 0.8519.

What are the implications of the main findings?

- The model effectively leverages generalized knowledge and spatial-frequency feature decoupling to suppress complex intertidal background noise.
- With a low inference latency of merely 8.5 ms per image, the proposed model is highly efficient and well-suited for onboard edge computing.

Abstract

Accurate mapping of *Spartina alterniflora* (*S. alterniflora*) is critical for coastal conservation. While deep learning has improved mapping, current algorithms that rely on moderate-resolution satellite imagery struggle to detect minute, early-stage invasive patches amid complex backgrounds with similar textures (e.g., water ripples, native vegetation). To overcome this resolution bottleneck, we construct UAV-SaSeg, the first high-resolution UAV semantic segmentation dataset targeting these complicated scenarios. Furthermore, we propose DINOsegNext, a novel mapping model integrating the DINOv3 foundation model with a Global Filter (GF) multi-frequency convolution. This architecture leverages generalized knowledge and spatial-frequency feature decoupling to filter out confusing background noise effectively. Validated in real-world intertidal zones, DINOsegNext outperforms state-of-the-art models, achieving an IoU of 0.7420 and an F1-Score of 0.8519. Crucially, it maintains high computational efficiency, with an inference latency of merely 8.5 ms per image, making it well-suited for onboard edge computing. This work provides urgently needed high-resolution data and a robust technical path for the precise, efficient monitoring of invasive species.



Academic Editor: Farid Melgani

Received: 19 May 2026

Revised: 5 July 2026

Accepted: 7 July 2026

Published: 16 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: *Spartina alterniflora*; semantic segmentation; Unmanned Aerial Vehicle (UAV); vision foundation model; lightweight network

1. Introduction

Since its introduction to China, *Spartina alterniflora* (*S. alterniflora*) has caused severe ecological damage to coastal wetland ecosystems. Its explosive expansion has rapidly encroached upon the habitats of native flora, such as reeds and mangroves, leading to the decline in native plant communities [1]. Achieving accurate dynamic monitoring and rapid spatial mapping of *S. alterniflora* is a crucial data foundation for regional ecological conservation and environmental management. For a long time, long-time-series dynamic monitoring using satellite remote sensing—favored for its broad coverage and low cost—has been the mainstream research approach [2–4]. However, these studies primarily rely on medium-resolution satellite data, where the prevalent mixed-pixel problem blurs invasion boundaries, making it difficult to accurately extract early, sparsely distributed invasion patches [5]. Furthermore, the strong phenotypic plasticity of *S. alterniflora* and its highly similar spectral characteristics to those of native vegetation compound the difficulty of accurate identification [6].

To overcome the aforementioned spatial resolution bottlenecks, low-altitude Unmanned Aerial Vehicle (UAV) remote sensing has emerged as an ideal tool for fine-scale ecological monitoring, owing to its high-resolution, centimeter-level imagery and flexible, on-demand acquisition capabilities. With the widespread adoption of UAV monitoring, feature extraction technologies have transitioned from traditional methods to deep learning [7,8]. Early research predominantly relied on traditional machine learning algorithms, such as Support Vector Machines (SVM) and Random Forests (RF) [9]. These methods are highly dependent on expert prior knowledge for feature design [10,11] and lack built-in convolutional feature extraction mechanisms. Consequently, they struggle to capture the complex leaf textures and lodging morphology of plants in UAV imagery, often resulting in fragmented classification results [9,12,13]. Additionally, their low computational efficiency fails to meet the demands of real-time monitoring. In recent years, deep learning models, such as U-Net and the Swin Transformer, have demonstrated remarkable performance. They effectively alleviate the challenges posed by intra-class spectral variability and inter-class spectral similarity, thereby significantly improving extraction accuracy [14,15]. However, mainstream high-accuracy segmentation models (e.g., DeepLabv3+) suffer from parameter redundancy and dense matrix computations [16,17], preventing their direct deployment on onboard UAV-embedded platforms with limited computing resources. Although some studies have mitigated memory constraints through patch-based inference [18], this approach often introduces distortion along object boundaries, failing to meet the rigorous requirements of fine-scale ecological monitoring.

To address deployment challenges caused by insufficient onboard computing power, lightweight semantic segmentation networks have become a core research focus, including in refs. [19–21]. Nevertheless, in the specific scenarios of *S. alterniflora* mapping—characterized by variable morphology and high background confusion—general lightweight models (e.g., lightweight PSPNet [22]) often lose rich semantic context due to over-compression of channels and network layers. This leads to underfitting in complex boundary regions and a drastic decline in segmentation accuracy [23,24]. Moreover, as a data-driven approach, deep learning relies heavily on high-quality annotated samples. Currently, there is a severe scarcity of data resources in this domain: publicly available datasets are predominantly satellite-based and lack the fine-grained textures captured from a UAV

perspective, while the few proprietary UAV datasets rarely account for field interferences such as drastic illumination changes, tidal dynamics, and phenological variations.

In summary, balancing pixel-level segmentation accuracy with lightweight deployment and achieving high-precision extraction in complex field scenarios under a severe shortage of high-quality datasets remain core scientific problems that urgently need resolution in the field of UAV-based *S. alterniflora* monitoring. Rather than building a lightweight network from scratch, introducing large-scale vision foundation models with exceptionally strong generalized visual representation capabilities (e.g., the DINO series [25,26]) and adapting them for lightweight deployment offers a novel paradigm to break through this technical bottleneck.

To address the aforementioned challenges, this paper proposes a highly robust and systematic research framework for UAV-based *S. alterniflora* monitoring, with its primary contributions spanning three dimensions: data, algorithm, and application generalization. First, at the data level, we construct and open-source China's first 0.02 m high-resolution UAV semantic segmentation dataset for *S. alterniflora*. This dataset was rigorously collected in the field and finely annotated, encompassing diverse and complex microhabitats, thereby breaking the data barriers for high-precision ecological monitoring research. Second, at the algorithmic level, we propose DINOsegNext, a lightweight network built on vision foundation models. It deeply integrates the global semantic representation capability of DINOv3, multi-scale contextual attention, and the multi-frequency detail extraction capabilities of global filtering layers, achieving an optimal balance between segmentation accuracy and lightweight inference under constrained computational resources. Finally, at the application generalization level, we conduct large-scale, cross-regional generalization performance validations. The results confirm that the proposed method exhibits exceptional anti-interference capabilities across various complex field environments, demonstrating significant value in practical large-scale ecological remote sensing monitoring applications.

2. Methods

2.1. Study Area

Sanmen County, Taizhou City, Zhejiang Province, China, was selected as the core study area, as shown in Figure 1. Located in a subtropical monsoon climate zone, the region features mild, humid conditions and significant tidal interactions, creating a suitable habitat for coastal wetland vegetation. Sanmen County possesses abundant tidal flat resources and serves as a vital base for marine aquaculture in China. However, in recent years, the explosive expansion of *Spartina alterniflora* has posed a severe challenge to the local ecosystem. The invasion of this species has not only led to waterway siltation and the loss of native ecological niches but has also severely occupied key aquaculture spaces, resulting in substantial economic and ecological losses. Consequently, to enhance monitoring efficacy in this complex coastal environment, this study employs high-resolution Unmanned Aerial Vehicle (UAV) remote sensing technology to provide robust data support for the formulation of scientific governmental control strategies.

2.2. Data

During its maturity period in October, *Spartina alterniflora* features tall, erect culms (1–3 m in height) and densely arranged linear-lanceolate leaves, with its high-density canopy exhibiting a transitional color from deep green to yellowish-green. In the 0.02-m resolution orthophotos, its visual distinguishing features are highly prominent: its unique yellowish-green tone forms a sharp contrast with the dark gray mudflat background; continuous communities exhibit an uneven “granular” or “wavy” rough texture, while scattered patches appear as round or oval “carpet-like” clumps with well-defined bound-

aries. Figure 1 includes field photographs showing the typical growth status of *Spartina alterniflora*.

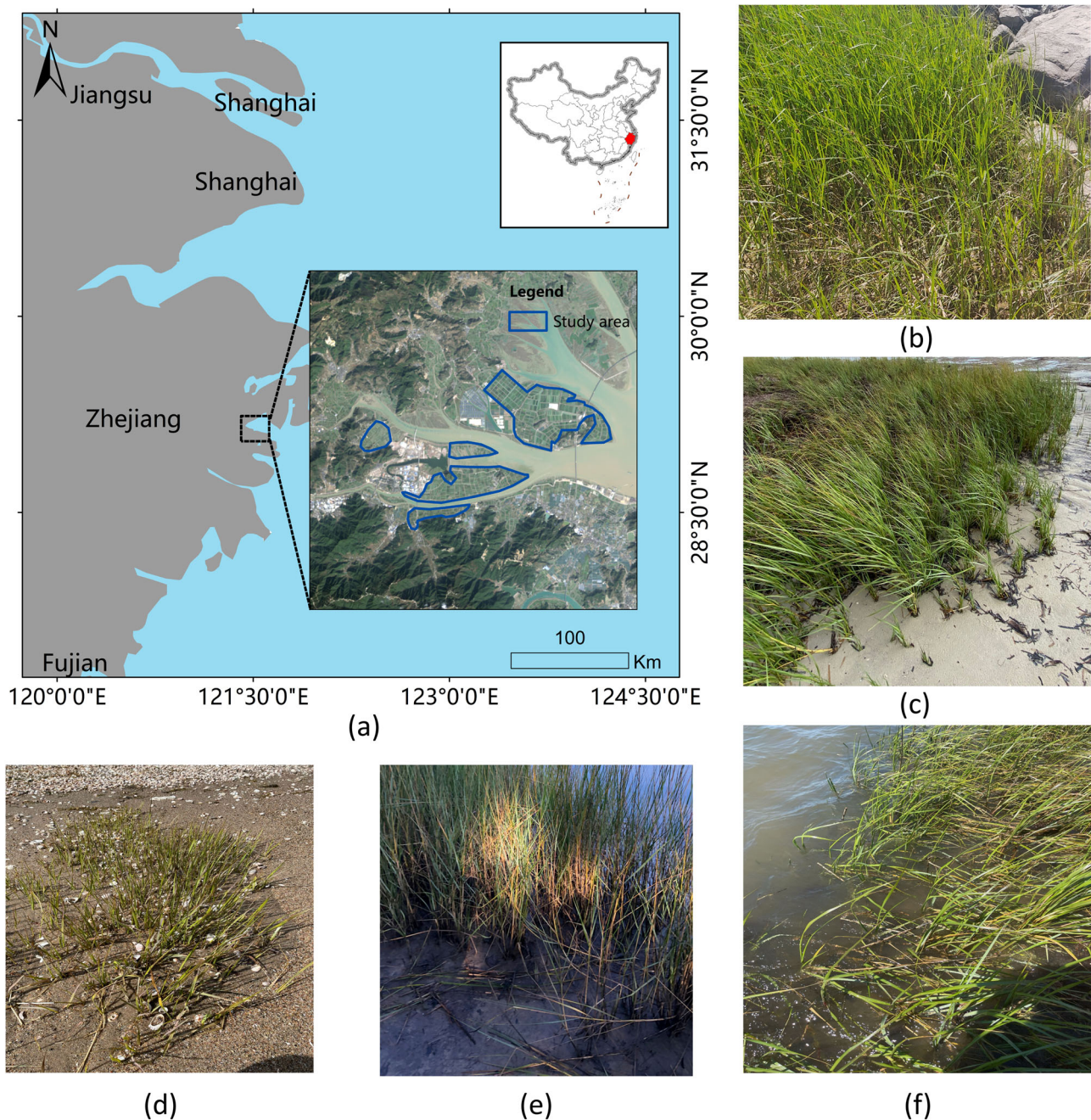


Figure 1. Overview of the study area and field photographs of *Spartina alterniflora*. (a) Schematic map of the study area location; (b–f) Field photographs showing the typical growth status of *Spartina alterniflora* (commonly known as smooth cordgrass).

Data acquisition was conducted in October 2025 in Shaliu Subdistrict, Sanmen County, Zhejiang Province, China, a typical coastal wetland ecosystem. This timing coincided with the mature stage of *S. alterniflora*, ensuring maximal spectral separability from background vegetation. We utilized a DJI multi-rotor UAV system equipped with a high-performance RGB sensor and a centimeter-level RTK module. Flights were conducted under optimal meteorological conditions to mitigate atmospheric and lighting artifacts. The raw imagery underwent orthorectification and mosaicking to produce a Digital Orthophoto Map (DOM) at 0.02 m resolution, capturing sufficient detail to identify leaf textures and fragmented

patches. A specific large-extent DOM ($116,135 \times 68,685$) was withheld to test the model's generalization in real-world scenarios.

To establish the UAV-SaSeg benchmark for deep learning, we cropped the DOM into 1024×1024 patches. Ground Truth generation adhered to a strict quality-control workflow: initial pixel-level labeling using LabelMe and ArcGIS 10.8 was followed by expert cross-validation to ensure boundary precision. The dataset employs binary masks (255 for *S. alterniflora* and 0 for background) and comprises 358 sample pairs. These were randomly split into training, validation, and test sets (6:2:2 ratio) for optimization, tuning, and evaluation, respectively (see Table 1).

Table 1. Detailed information about the dataset.

Spatial Resolution	Spectral Bands	Total Number	Open Access
0.02 m	R, G, B	358	Yes

Statistically, the UAV-SaSeg dataset exhibits distinct characteristics that underscore its immense value for both coastal ecology and computer vision communities. First, with a Ground Sample Distance (GSD) of 0.02 m, the dataset possesses an exceptionally high information density. The 358 training, validation, and testing patches (1024×1024) collectively provide over 375 million fine-grained, precisely annotated pixels, as shown in Figure 2. This high-resolution image captures significant intra-class variance and morphological diversity, ranging from dense, homogeneous canopies to highly fragmented, sparse patches that are often confused with complex mudflat backgrounds.

Figure 3 contrasts the imaging capabilities of high-resolution UAV imagery with those of conventional commercial satellite data. Although the Maxar WorldView imagery used for comparison boasts a spatial resolution of 0.6 m—representing a top-tier precision in civilian remote sensing—it remains insufficient to capture the micro-spectral characteristics and fine-grained spatial textures within *Spartina alterniflora* communities. Therefore, this dataset relies on UAV imagery to provide enriched textural and spectral representations, ultimately facilitating robust training and precise identification by semantic segmentation algorithms.

The dataset is publicly available at: [<https://github.com/wzp8023391/Spartina-alterniflora-Monitoring-Using-UAV>] Accessed on 5 May 2026.

2.3. Methods

To address the dual challenges of constrained computational resources on unmanned aerial vehicle (UAV) platforms and the difficulty in capturing the intricate edge details of *Spartina alterniflora* (*S. alterniflora*) [27,28], we propose DINOsegNext, a lightweight segmentation network based on spatial-frequency dual-domain perception. As illustrated in Figure 4, the model adopts a compact encoder–decoder architecture: first, a fine-tuned Vision Foundation Model (DINOv3) [25] is utilized to extract robust spatial semantic features; second, a Global Filter (GF) operating in the frequency domain is introduced at the front end of the decoder to capture multi-frequency representations; and finally, a lightweight decoder head derived from SegNeXt [29] is employed to accomplish spatial context aggregation and mask generation.

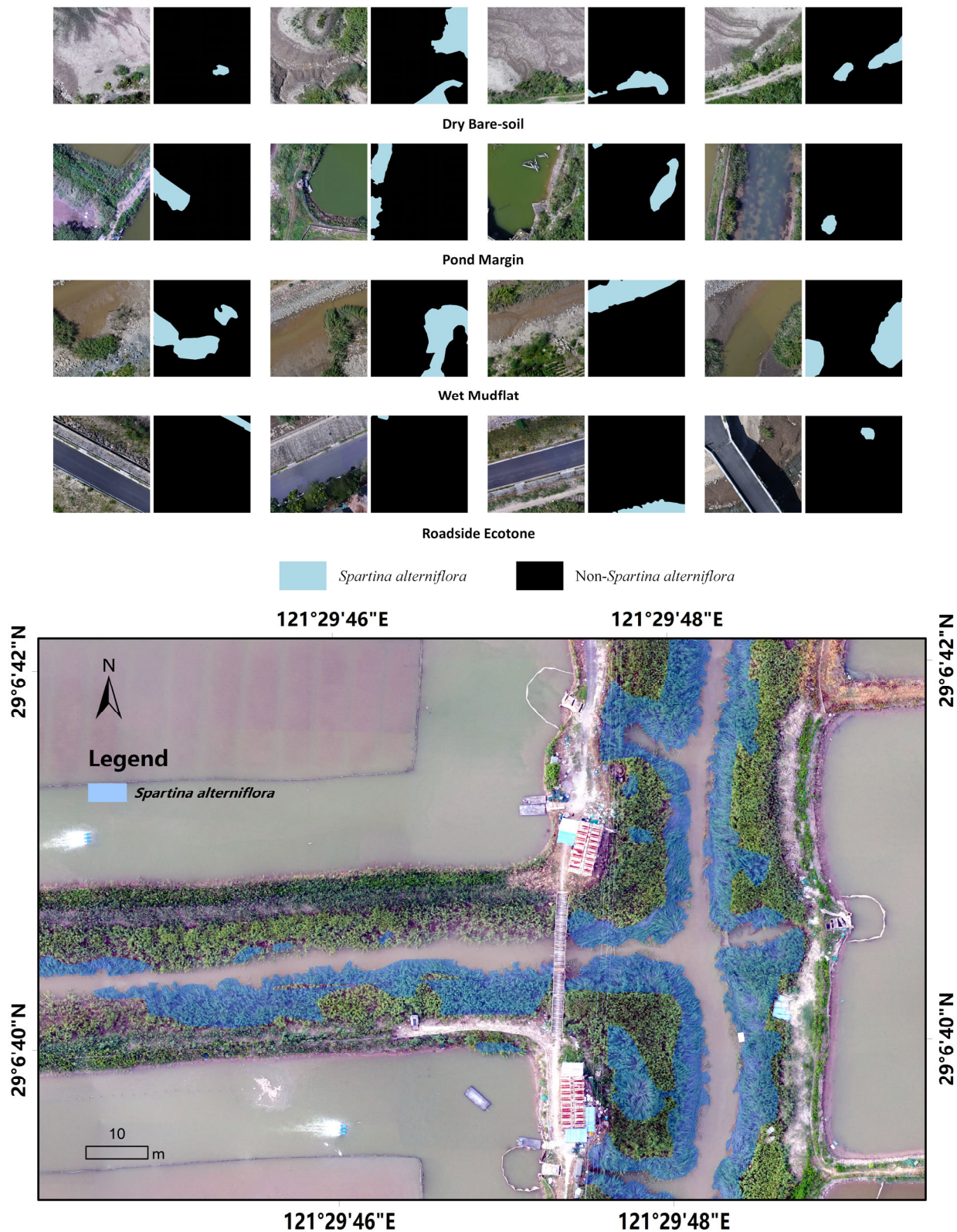


Figure 2. Representative samples from the collected *Spartina alterniflora* UAV dataset.



Figure 3. Overview of the experimental test set area and a resolution benchmarking between satellite and UAV data. The close-up magnifications clearly illustrate the significant disparity in spatial fidelity, underscoring the necessity of UAV-based high-resolution imagery for complex canopy discrimination.

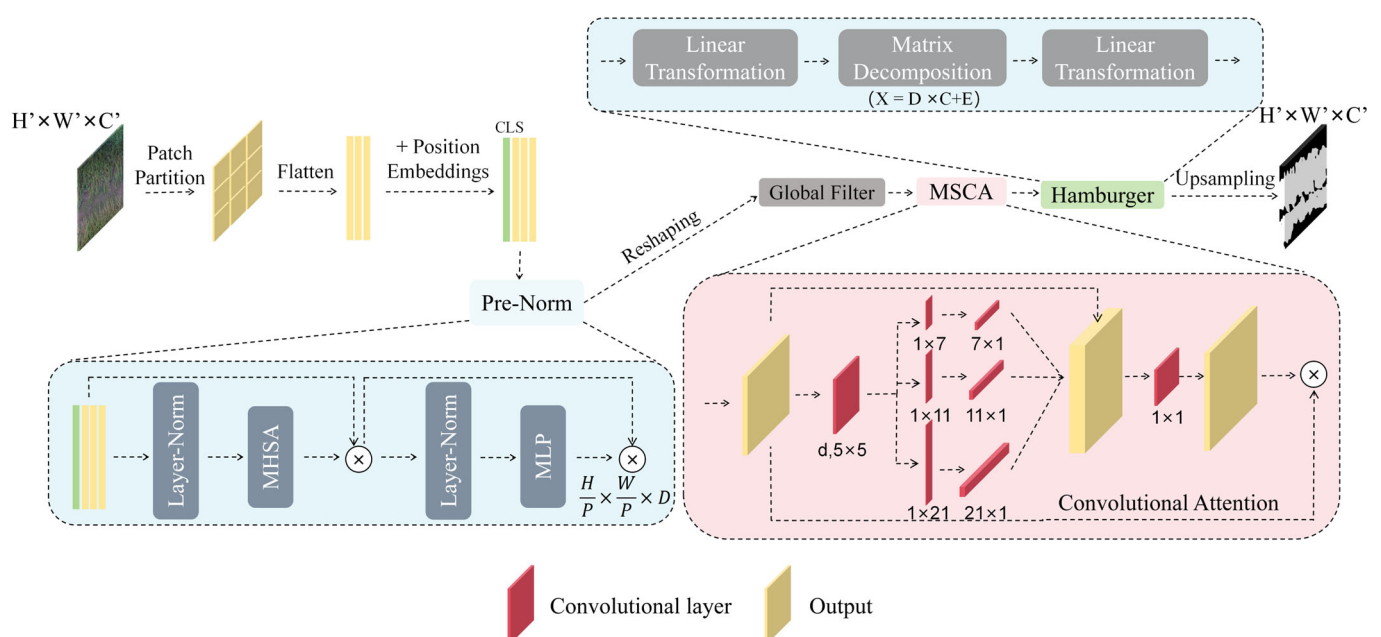


Figure 4. Schematic diagram of the overall architecture of the proposed DINOsegNext model.

(1) Fine-tuned Vision Foundation Model Encoder

To acquire highly generalizable features under limited sample conditions, we adopt the pre-trained DINOv3 as the backbone network. Leveraging its asymmetric teacher–student architecture and Masked Image Modeling (MIM) mechanism, DINOv3 inherently possesses formidable capabilities for generalized visual representation. Building upon this, we fine-tune DINOv3 specifically on our *S. alterniflora* dataset. This fine-tuning strategy circumvents the prohibitive computational costs of training from scratch while enabling the model to rapidly adapt to the complex coastal wetland scenarios captured from UAV perspectives. Consequently, it establishes a high-quality feature foundation that is highly robust to variations in illumination and morphological appearance for subsequent segmentation tasks.

(2) Multi-frequency Domain Convolution Mechanism

Accurately extracting *S. alterniflora* from high-resolution (0.02 m) imagery requires not only understanding the macroscopic community distribution but also delineating fine-scale leaf textures and jagged boundaries. Achieving this with traditional spatial convolutions typically requires stacking numerous convolutional layers to expand the receptive field, which increases computational overhead. To mitigate this, we innovatively introduce a Global Filter (GF) module [30] before the decoding head. Specifically, this module incorporates convolutional kernels at three distinct resolutions. This multi-resolution frequency-domain mapping mechanism theoretically enables the simultaneous capture of high-, mid-, and low-frequency image components. The low-frequency component helps anchor the continuous main regions of *S. alterniflora* and smooth the background; the mid-frequency component characterizes the lodging textures within the canopy; and the high-frequency component precisely sharpens the vegetation edge details. More importantly, as demonstrated by [30], the GF module transforms large-scale spatial contextual interactions into element-wise multiplication within the frequency domain. This transformation significantly expands the receptive field and enriches feature dimensions while substantially reducing the model’s Floating Point Operations (FLOPs), perfectly aligning with the stringent constraints of onboard edge computing. Figure 5 illustrates the frequency-domain convolution kernels of the GF layer at different resolutions.

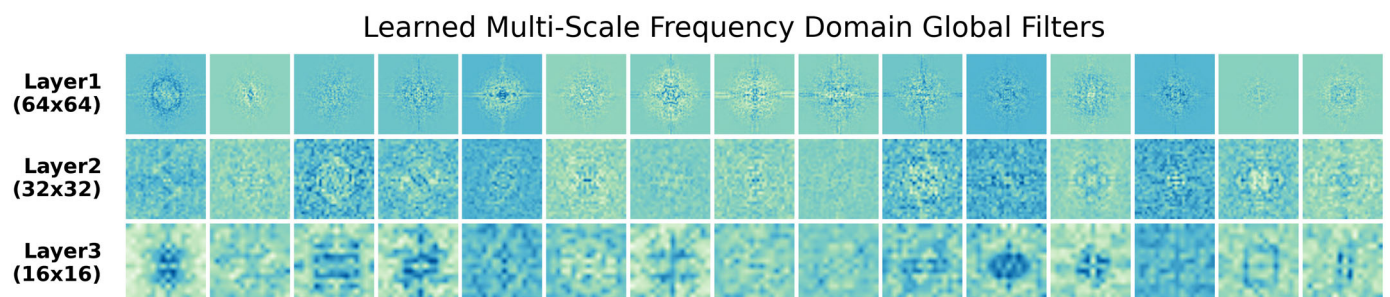


Figure 5. Visualization of the learned multi-scale global filters in the frequency domain across different network layers. The rows, from top to bottom, illustrate the spectral weights initialized/learned at resolutions of 64×64 , 32×32 , and 16×16 , respectively, demonstrating the model’s capacity to capture multi-frequency components (ranging from high-frequency edges to low-frequency global contexts).

(3) Lightweight Decoder Head based on SegNeXt

Following feature enhancement via frequency-domain convolutions, we employ the lightweight decoder head from the SegNeXt architecture for the final spatial context aggregation. This decoder synergistically integrates Multi-Scale Convolutional Attention (MSCA) [31] and the “Hamburger” module. Initially, MSCA employs strip convolutions to

approximate the receptive field of large-kernel convolutions, thereby efficiently capturing multi-scale local spatial context at a minimal computational cost. Subsequently, the features are fed into the core Hamburger module, which discards computationally intensive self-attention mechanisms in favor of Matrix Decomposition for global modeling. In practice, the Hamburger module decomposes the input high-dimensional feature matrix into low-rank subspaces, effectively filtering out redundant low-level background noise (e.g., tidal water reflections or sediment interference). This cascaded mechanism—combining MSCA’s local multi-scale perception with the Hamburger module’s global matrix decomposition—not only significantly enhances the spatial consistency of the foreground target (*S. alterniflora*) but also maintains parameter counts and computational complexity far below those of traditional pyramid pooling modules (such as ASPP). Consequently, it achieves an optimal equilibrium between pixel-level segmentation accuracy and model inference efficiency.

2.4. Accuracy Evaluation and Parameter Settings

To address the dual challenges of the fragmented spatial distribution of *S. alterniflora* patches and severe class imbalance in coastal wetland scenes, this study establishes a comprehensive evaluation framework that balances segmentation accuracy with computational efficiency.

Regarding accuracy evaluation, given that Overall Accuracy (OA) can easily yield misleadingly inflated scores on highly imbalanced datasets [32], we selected Intersection over Union (IoU) and the F1-score as the core evaluation metrics to rigorously assess the model’s robustness and its capability to capture minute, early-stage invasive patches. Furthermore, Recall and Precision were introduced as auxiliary metrics: the former quantifies the completeness of sporadic invasion detection, thereby minimizing the ecological risks associated with missed detections. At the same time, the latter evaluates the model’s discriminative power against co-occurring native vegetation (e.g., reeds) to ensure the precision of ecological management decisions.

For efficiency evaluation, to meet the rapid processing demands of large-scale, high-resolution digital orthophoto maps (DOMs) and to ensure compatibility with future on-board edge deployment, we used Floating Point Operations (FLOPs) and parameter counts (Params) to quantify the static structural complexity and computational workload of the models [33,34]. Crucially, we also introduced inference time as a metric for dynamic execution efficiency. This combination of static and dynamic metrics provides a more comprehensive benchmark for assessing the operational throughput and engineering feasibility of the models in real-time monitoring workflows.

All computational experiments were executed on a high-performance workstation configured with an Intel Core i5-12600KF CPU, an NVIDIA GeForce RTX 4070 GPU, and 32 GB of RAM, running on Windows 11. The experimental architecture was implemented using Python 3.13 and the PyTorch 2.7 deep learning framework.

During the training phase, the AdamW optimizer was employed to enhance the model’s generalization capability [35]. The initial learning rate was set to 0.00001, with a batch size of 4, and the total number of training epochs was capped at 45. To strike an optimal balance between leveraging the robust prior features of the DINO foundation model and adapting effectively to the specific coastal wetland dataset, a two-stage optimization strategy was implemented: during the initial 20 epochs, the parameters of the backbone network were frozen to warm up and train the decoder exclusively; subsequently, the backbone was unfrozen to permit global, end-to-end fine-tuning across the remaining epochs. The resulting convergence profile of the loss curve during model training is illustrated in Figure 6.

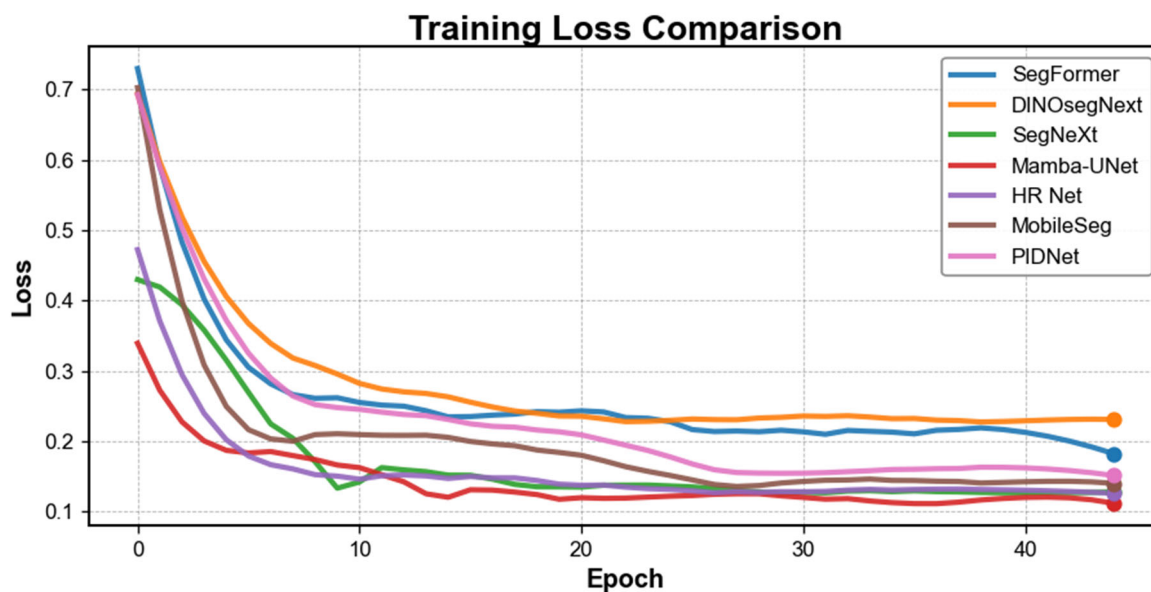


Figure 6. Visualization comparison of training loss curves across different models.

2.5. Inference Implementation and Optimization Strategies

To evaluate the model's actual deployment performance in resource-constrained environments, we conducted inference tests on a mobile platform equipped with an NVIDIA GeForce RTX 4050 Laptop GPU. To maximize throughput and reduce the memory footprint, we adopted half-precision floating-point (FP16) acceleration technology [36]. When processing large-scale UAV orthoimages, we designed a Dynamic Sliding Window Algorithm. This algorithm integrates a background filtering mechanism that automatically identifies and skips No-data regions and invalid imaging areas, thereby significantly reducing redundant computations. The final inference results are automatically encapsulated as georeferenced polygon vector (Shapefile) and raster (GeoTIFF) files to support subsequent GIS spatial analysis directly.

3. Results

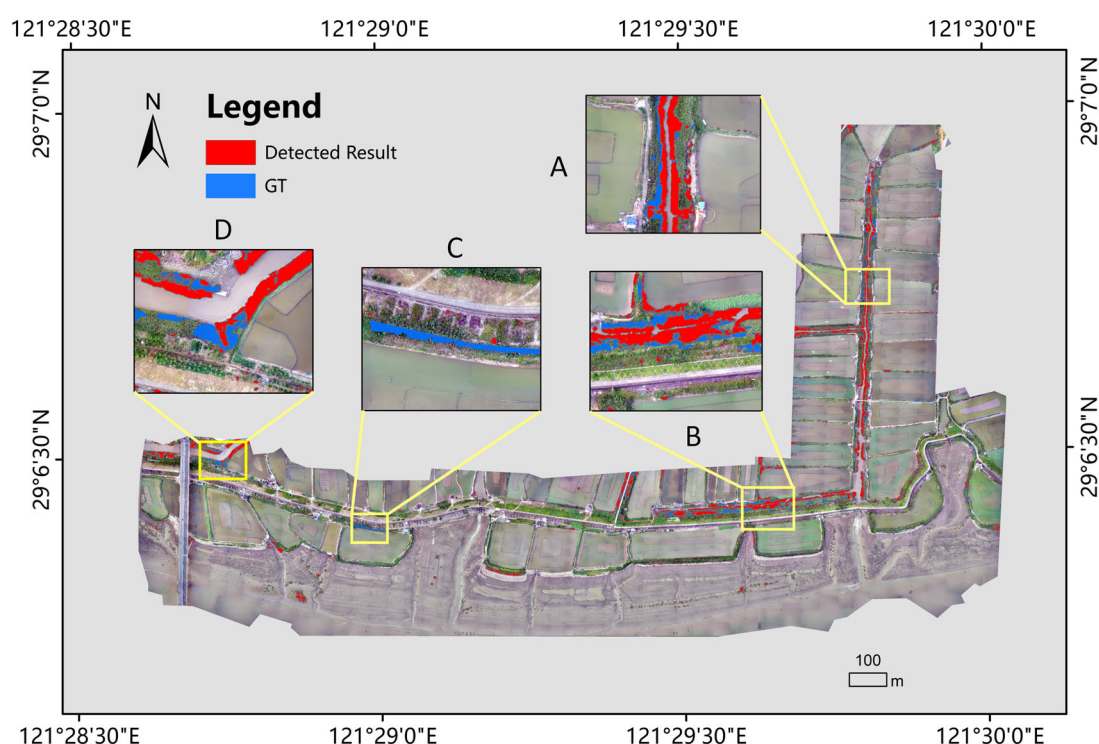
3.1. Accuracy Evaluation and Efficiency Analysis

We conducted a comprehensive comparative test on seven mainstream segmentation models, including DINOsegNext. Experimental results show a significant discrepancy in inference latency between CPU and GPU endpoints. This phenomenon is mainly attributed to the parallel acceleration optimization of CUDA (Compute Unified Device Architecture) libraries for dense convolution operations and attention mechanisms on GPUs [37].

Nonetheless, the CPU inference scheme still demonstrates extremely high practical value. Compared to traditional manual visual interpretation—where a skilled interpreter requires approximately 16 person-hours (two working days) to complete the *S. alterniflora* extraction for a full image—even when using only a CPU, the automated model proposed in this paper can reduce processing time to the minute level. This indicates that the method has immense potential to replace manual labor and enable efficient, automated monitoring in field operations that lack high-end computing power. Specific performance metrics for each model are in Table 2, and Figure 7 demonstrates the performance of our proposed model on the validation set.

Table 2. Quantitative comparison of segmentation performance on the test set.

Model Name	Precision	Recall	F1-Score	IoU	OA	Latency (ms)	FLOPs (G)	Params (M)
SegFormer	0.4762	0.8768	0.6172	0.4463	0.9793	11.0	62.4	27.5
HRNet	0.6272	0.6137	0.6204	0.4496	0.9855	14.0	40	29.5
PIDNet	0.5287	0.9031	0.667	0.5003	0.9827	9.0	78.5	12.4
SegNeXt	0.7635	0.8644	0.8108	0.6818	0.9895	12.5	34.9	27.6
MobileSeg	0.7688	0.8680	0.8154	0.6883	0.9898	9.5	13	5.6
Mamba-UNet	0.7915	0.8433	0.8167	0.6902	0.9905	25.0	70	18
DINOsegNext (ours)	0.8245	0.8812	0.8519	0.7420	0.9931	8.5	28.5	22.5

**Figure 7.** Visualization of semantic segmentation inference results in typical scenes. A–D are the subregions of the study area.

3.2. Comparison with State-of-the-Art Lightweight Models

To comprehensively evaluate the competitive edge of DINOsegNext in resource-constrained scenarios, we conducted a rigorous benchmarking analysis against the most representative lightweight paradigms in semantic segmentation. The comparative baselines include classic Convolutional Neural Network (CNN)-based architectures (HRNet [38], PIDNet [39], MobileSeg), mainstream Transformer-based models (SegFormer [19], SegNeXt [29]), and the emerging State Space Model (SSM)-based Mamba-UNet [40]. Detailed quantitative metrics for these models are summarized in Table 3.

Quantitative results indicate that while these state-of-the-art (SOTA) models perform exceptionally well on generic datasets, they exhibit significant performance divergence in the fine-grained segmentation of *Spartina alterniflora*. DINOsegNext achieved superior performance across all key metrics. Notably, SegFormer, a benchmark in the Transformer domain, delivered suboptimal results on this task, with segmentation accuracy at the boundaries of fragmented patches falling significantly short of expectations. This under-performance may stem from the pure Transformer architecture's insufficient inductive

bias, which limits its ability to effectively capture subtle local textural features under data-constrained conditions [41].

Table 3. Overview of the comparative models used in this study.

Model Name	Publication Journal	Principle
SegFormer	NeurIPS (2021)	SegFormer abandons positional encoding and employs a hierarchical Transformer encoder (MiT) to output multi-scale features. It introduces a lightweight “All-MLP” decoder that aggregates information across layers, striking a trade-off between efficiency and performance.
HRNet	IEEE TPAMI (2021)	Unlike traditional encoders that downsample images to recover semantics, HRNet maintains high-resolution representations throughout the entire process. It connects high- to low-resolution convolutions in parallel and repeatedly performs multi-scale fusion to preserve rich spatial details.
PIDNet	CVPR (2023)	Inspired by PID controllers in control theory, PIDNet utilizes a three-branch architecture: a Proportional (P) branch for high-resolution details, an Integral (I) branch for long-range context, and a Derivative (D) branch to model high-frequency boundary features, effectively reducing the “overshoot” effect at object boundaries.
SegNeXt	NeurIPS (2022)	SegNeXt revisits the advantages of Convolutional Neural Networks (CNNs). It proposes Multi-Scale Convolutional Attention (MSCA) to simulate the large receptive field of Transformers using cheap convolutions, proving that a well-designed CNN can outperform Transformers in efficiency.
MobileSeg	arXiv (2023)	Designed specifically for mobile devices, it adopts a Coupled-Branch structure to decouple semantic and detailed information. It utilizes a Stride-16 aggregation strategy and pre-fusion mechanisms to maximize inference speed while maintaining competitive accuracy on edge hardware.
Mamba-UNet	arXiv (2024)	Incorporating the State Space Model (SSM) into a U-shaped architecture, Mamba-UNet leverages the Mamba block’s linear computational complexity to model long-range dependencies. It overcomes the quadratic complexity bottleneck of Transformers and is well-suited to processing high-resolution biomedical or remote-sensing images.

Further qualitative analysis highlights the differences in model robustness within complex environments. We specifically selected typical coastal scenes characterized by high spectral confusion—such as sandy transition zones with complex textures and aquaculture ponds—for visual verification. The experiments revealed that models such as HRNet and PIDNet were susceptible to background noise in these areas, resulting in a high False Positive Rate as they frequently misclassified mudflat textures as *S. alterniflora*. In contrast, leveraging the powerful feature representation capabilities of the foundation model, DINOsegNext effectively suppressed background noise, maintaining exceptional identification precision even in these challenging regions prone to the “same spectrum, different objects” phenomenon. Figure 8 presents a grid plot visualizing the quantitative performance of each model. Figure 9 visualizes the true positives, false positives, and false negatives of all compared models.

3.3. Ablation Study

To quantitatively validate the effectiveness of introducing the Global Filter (GF) multi-frequency convolution module for feature decoupling and noise suppression, we conducted rigorous comparative ablation experiments under identical environmental configurations. The experiments were implemented using Python 3.13 and the PyTorch framework, running on an NVIDIA GeForce RTX 4050 Laptop GPU.

We divided the evaluation into two groups: the Baseline group, which used a standard spatial-domain architecture that fed encoder features directly into the SegNeXt decoder head; and the Experimental group (Ours), which integrated the proposed multi-frequency GF module before the decoding head.

Table 4 presents the quantitative evaluation results of the ablation study regarding the GF module. A comparison of the data reveals that incorporating the GF module led to substantial improvements across all key metrics: notably, model Precision improved from 73.15% in the baseline model to 82.45% (an increase of 9.3%), while IoU increased from 67.36% to 74.2%. More importantly, rather than introducing architectural overhead, integrating the GF module significantly improved operational efficiency, reducing inference latency from 11.2 ms to 8.5 ms. This dual-advancement powerfully demonstrates the superiority of frequency-domain convolutions in simultaneously mitigating complex background interference and boosting execution speed (Figure 10). Figure 11 visualizes the true positives, false positives, and false negatives of the proposed and baseline models.

Accuracy and Latency Trade-off



Figure 8. Comparative visualization of the trade-off between segmentation accuracy (IoU) and inference efficiency (ms) among different contrastive models for *Spartina alterniflora* extraction. The x-axis denotes the average inference latency (ms), and the y-axis denotes the Intersection over Union (IoU) metric.

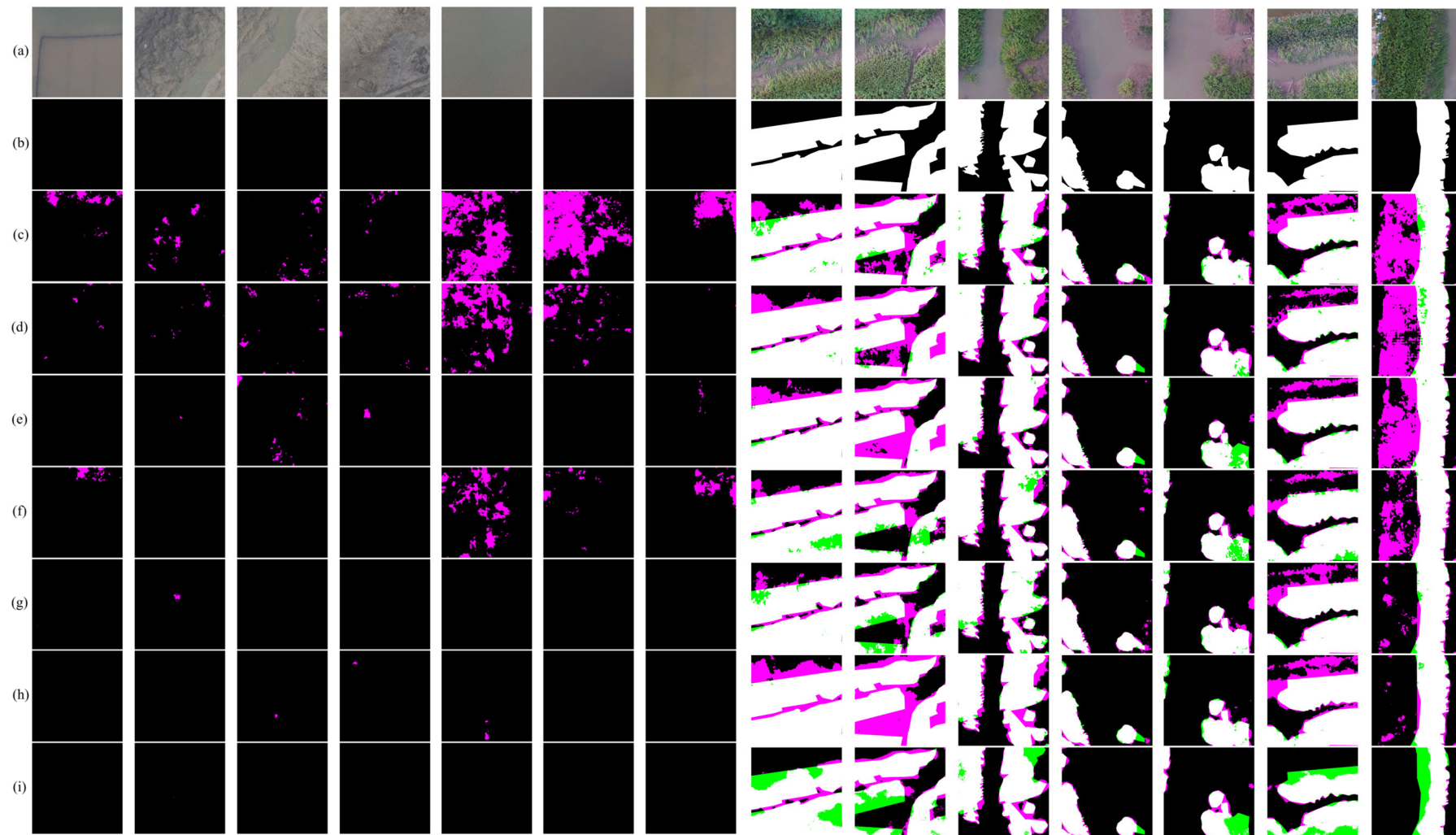


Figure 9. Visual comparison of segmentation results among different models, with white, green, and purple colors indicating correct detections (True Positives), missed detections (False Negatives), and false alarms (False Positives), respectively. The sub-panels correspond to: (a) Original RGB image, (b) Ground Truth, (c) SegFormer, (d) HRNet, (e) PIDNet, (f) SegNeXt, (g) MobileSeg, (h) Mamba-UNet, and (i) DINOsegNext (Ours).

Table 4. Quantitative results of the ablation study on different model components.

Model Name	Precision	Recall	F1-Score	IoU	OA	Latency (ms)
baseline	0.7315	0.8950	0.8050	0.6736	0.9865	11.2
proposed	0.8245	0.8812	0.8519	0.7420	0.9931	8.5



Figure 10. Visualization of training loss curves for the ablation study.

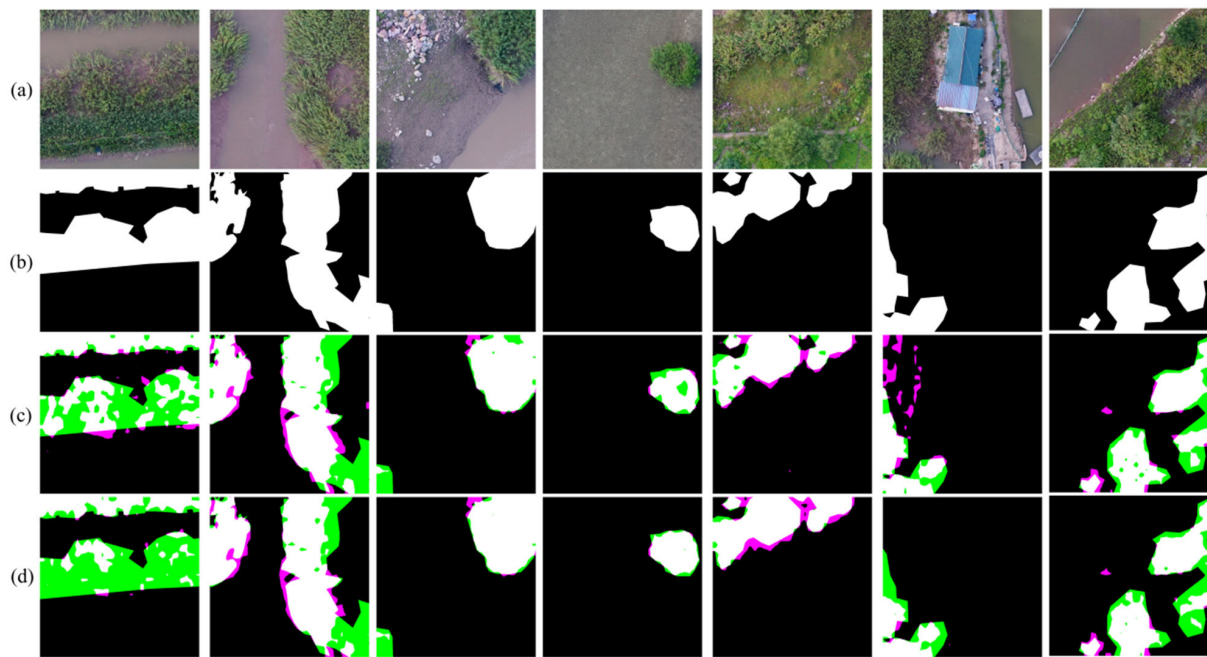


Figure 11. Visual comparison of segmentation results for the ablation study, with white, green, and purple colors indicating correct detections (True Positives), missed detections (False Negatives), and false alarms (False Positives), respectively. (a) Original RGB image, (b) Ground Truth, (c) Baseline model, and (d) Proposed model.

Specifically, purely spatial decoding mechanisms that rely on conventional sliding windows are highly susceptible to interference from background noise that shares similar textures with *S. alterniflora* (e.g., water ripples or mixed native vegetation), frequently resulting in a high False Positive Rate [42]. Conversely, by mapping spatial features into the frequency domain, the GF module leverages multi-resolution spectral element-wise multiplication to aggregate global context. This mechanism not only captures macroscopic distribution patterns but also effectively filters out redundant feature components that primarily respond to background interference. By converting dense spatial contextual interactions into efficient frequency-domain multiplications, this strategy shields the model from the deceptive induction of locally similar textures while eliminating heavy pixel-level computational redundancy, thereby achieving both higher precision in determining positive samples and faster operational throughput.

4. Discussion

4.1. Mitigation of Over-Prediction and Optimization of Feature Learning

Experimental results indicate that several baseline models exhibited high false-positive rates during the detection phase. These models tended to over-predict, which, despite yielding an excessively high Recall, resulted in a drastic degradation of Precision. This phenomenon is primarily attributed to the massive parameter spaces of these networks, coupled with the relatively limited number of training samples in our study. Consequently, the models inevitably suffered from overfitting due to insufficient data-driven optimization [43]. In contrast, the DINOv3 model utilized in our study, built upon a large-scale self-supervised pre-training paradigm [44], inherently possesses robust feature extraction and prior representation capabilities. By merely fine-tuning its internal parameters, the model successfully overcomes overfitting and achieves high-precision detection even with limited data. Furthermore, optimizing the loss function with Focal Loss [45] dynamically calibrates the weights assigned to hard and easy examples. This mechanism effectively suppresses interference from numerous simple background samples, thereby significantly enhancing the model's resilience to false positives and improving overall detection accuracy.

4.2. Inference Efficiency and Frequency-Domain Optimization via Global Filter

Regarding inference efficiency, experiments demonstrate that our proposed DINOseg-Next model achieves the lowest inference time among all evaluated models. This superior efficiency is fundamentally credited to the integration of the Global Filter (GF) layer [30]. The GF layer transforms the feature extraction process from the spatial domain to the frequency domain using the Fast Fourier Transform (FFT). According to the convolution theorem, complex and dense convolution operations, or self-attention mechanisms, in the spatial domain can be replaced by computationally lightweight element-wise multiplications in the frequency domain. This transformation drastically reduces the computational complexity—shifting from spatial quadratic complexity to the logarithmic complexity $O(N \log N)$ of FFT—without sacrificing the original information capacity, thereby substantially accelerating the model's overall computational efficiency.

4.3. Limitations and Future Work

Despite DINOsegNext demonstrating exemplary performance in suppressing complex background noise and enhancing the Precision of *Spartina alterniflora* identification, it remains constrained by the limited geographical scope of the current dataset. Consequently, the high phenotypic heterogeneity of *S. alterniflora* across different latitudes and tidal zones has not been fully encompassed [46], causing the fine-tuning of this foundation model to

remain somewhat restricted by “data hunger” and limiting its generalization potential in unseen or extreme habitats [47].

In light of this, future work will focus on expanding the dataset’s scale and diversity while conducting rigorous cross-domain validation. First, we intend to expand the geographic scope of data acquisition to construct a comprehensive *S. alterniflora* atlas across various latitudes along China’s coast (from subtropical to temperate zones), thereby encompassing more diverse intertidal zone typologies and growth stages to address phenotypic monotony. Second, building on this extended dataset, we will conduct in-depth cross-domain testing to systematically verify the model’s robustness under varying illumination levels, water turbidity, and co-occurring vegetation communities. Our ultimate focus is to develop an automated operational system with broad, wide-area applicability and all-weather monitoring capabilities, providing robust technical support for national-scale dynamic monitoring and ecological governance of *S. alterniflora*.

5. Conclusions

This study aimed to address the dual bottlenecks of “scarcity of high-resolution data” and “difficulties in edge deployment” in the fine-grained monitoring of *Spartina alterniflora* (*S. alterniflora*). To mitigate the lack of high-resolution samples in existing public datasets, we independently collected and constructed the first UAV-based semantic segmentation dataset for *S. alterniflora* (UAV-SaSeg). This dataset fills a critical data void in the field and provides a high-quality benchmark for identifying minute, early-stage invasive patches.

At the algorithmic level, this study systematically validated the effectiveness of the “Vision Foundation Model + Lightweight Decoder” design paradigm for specific ecological monitoring tasks. Experimental results demonstrate that employing the fine-tuned DINOv3 as the backbone network, combined with a lightweight decoding mechanism based on matrix decomposition, achieves superior feature extraction capabilities while maintaining a minimal computational footprint. Notably, as corroborated by the ablation studies, the integration of the multi-frequency Global Filter (GF) convolution not only significantly improves the segmentation IoU in the presence of complex textural backgrounds but also substantially reduces computational overhead by shifting dense spatial contextual interactions to efficient frequency-domain multiplications. This makes the architecture highly conducive to the onboarding of edge computing.

In summary, this study demonstrates that lightweight architectures based on foundation models represent a highly feasible technical path for the efficient and precise monitoring of *S. alterniflora*. Future efforts will focus on expanding the construction of wide-area datasets and cross-domain validation to realize an automated operational monitoring system with national-scale applicability.

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

Funding: This work was supported by the Hunan Institute of Geology Scientific Research Grant Project under Grant HNGYSTP20261, the Open Research Fund of the Science and Technology Innovation Platform of Changsha University of Science and Technology under Grant 2025ZKPT057, and the Hunan Provincial Natural Science Foundation of China under Grants 2024JJ8367 and 2026JJ60618, and

the the Hunan Engineering Technology Research Center of Natural Resources Survey and Monitoring (No. 2018TP2040).

Data Availability Statement: <https://github.com/wzp8023391/Spartina-alterniflora-Monitoring-Using-UAV>, accessed on 5 May 2026.

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

References

1. Liu, M.; Mao, D.; Wang, Z.; Li, L.; Man, W.; Jia, M.; Ren, C.; Zhang, Y. Rapid Invasion of *Spartina alterniflora* in the Coastal Zone of Mainland China: New Observations from Landsat OLI Images. *Remote Sens.* **2018**, *10*, 1933. [CrossRef]
2. Zhang, X.; Xiao, X.; Qiu, S.; Xu, X.; Wang, X.; Chang, Q.; Wu, J.; Li, B. Quantifying latitudinal variation in land surface phenology of *Spartina alterniflora* saltmarshes across coastal wetlands in China by Landsat 7/8 and Sentinel-2 images. *Remote Sens. Environ.* **2022**, *269*, 112810. [CrossRef]
3. Han, X.; Wang, Y.; Ke, Y.; Liu, T.; Zhou, D. Phenological heterogeneities of invasive *Spartina alterniflora* salt marshes revealed by high-spatial-resolution satellite imagery. *Ecol. Indic.* **2022**, *144*, 109492. [CrossRef]
4. Min, Y.; Cui, L.; Li, J.; Han, Y.; Zhuo, Z.; Yin, X.; Zhou, D.; Ke, Y. Detection of large-scale *Spartina alterniflora* removal in coastal wetlands based on Sentinel-2 and Landsat 8 imagery on Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *125*, 103567. [CrossRef]
5. Zhang, X.; Xiao, X.; Wang, X.; Xu, X.; Qiu, S.; Pan, L.; Ma, J.; Ju, R.; Wu, J.; Li, B. Continual expansion of *Spartina alterniflora* in the temperate and subtropical coastal zones of China during 1985–2020. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *117*, 103192. [CrossRef]
6. Tian, J.; Wang, L.; Yin, D.; Li, X.; Diao, C.; Gong, H.; Shi, C.; Menenti, M.; Ge, Y.; Nie, S.; et al. Development of spectral-phenological features for deep learning to understand *Spartina alterniflora* invasion. *Remote Sens. Environ.* **2020**, *242*, 111745. [CrossRef]
7. Li, Y.; Qin, F.; He, Y.; Liu, B.; Liu, C.; Pu, X.; Wan, F.; Qiao, X.; Qian, W. The effect of season on *Spartina alterniflora* identification and monitoring. *Front. Environ. Sci.* **2022**, *10*, 1044839. [CrossRef]
8. Lv, Q.; Zhou, P.; Yang, S.; Shi, Y.; Ma, J.; Yang, J.; Chen, G. Provincial Scale Monitoring of Mangrove Area and Smooth Cordgrass Evasion in Subtropical China Using UAV Imagery and Machine-Learning Methods. *Remote Sens.* **2026**, *18*, 345. [CrossRef]
9. Anderson, C.J.; Heins, D.; Pelletier, K.C.; Knight, J.F. Improving Machine Learning Classifications of *Phragmites australis* Using Object-Based Image Analysis. *Remote Sens.* **2023**, *15*, 989. [CrossRef]
10. Tian, Y.; Jia, M.; Wang, Z.; Mao, D.; Du, B.; Wang, C. Monitoring Invasion Process of *Spartina alterniflora* by Seasonal Sentinel-2 Imagery and an Object-Based Random Forest Classification. *Remote Sens.* **2020**, *12*, 1383. [CrossRef]
11. Zhuo, W.; Wu, N.; Shi, R.; Liu, P.; Zhang, C.; Fu, X.; Cui, Y. Aboveground biomass retrieval of wetland vegetation at the species level using UAV hyperspectral imagery and machine learning. *Ecol. Indic.* **2024**, *166*, 112365. [CrossRef]
12. Abeysinghe, T.; Simic Milas, A.; Arend, K.; Hohman, B.; Reil, P.; Gregory, A.; Vázquez-Ortega, A. Mapping Invasive *Phragmites australis* in the Old Woman Creek Estuary Using UAV Remote Sensing and Machine Learning Classifiers. *Remote Sens.* **2019**, *11*, 1380. [CrossRef]
13. Baibagyssov, A.; Magiera, A.; Thevs, N.; Waldhardt, R. Resource Characteristics of Common Reed (*Phragmites australis*) in the Syr Darya Delta, Kazakhstan, by Means of Remote Sensing and Random Forest. *Plants* **2025**, *14*, 933. [CrossRef] [PubMed]
14. Li, H.; Cui, G.; Liu, H.; Wang, Q.; Zhao, S.; Huang, X.; Zhang, R.; Jia, M.; Mao, D.; Yu, H.; et al. Dynamic Analysis of *Spartina alterniflora* in Yellow River Delta Based on U-Net Model and Zhuhai-1 Satellite. *Remote Sens.* **2025**, *17*, 226. [CrossRef]
15. Wang, Z.; Li, J.; Tan, Z.; Liu, X.; Li, M. Swin-UperNet: A Semantic Segmentation Model for Mangroves and *Spartina alterniflora* Loisel Based on UperNet. *Electronics* **2023**, *12*, 1111. [CrossRef]
16. Li, J.; Xiu, J.; Yang, Z.; Liu, C. Dual Path Attention Net for Remote Sensing Semantic Image Segmentation. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 571. [CrossRef]
17. Wang, Y.; Shi, H.; Dong, S.; Zhuang, Y.; Chen, L. Dual-Path Sparse Hierarchical Network for Semantic Segmentation of Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
18. Higgsison, W.; Cobb, A.; Tschierschke, A.; Dyer, F. Estimating the cover of *Phragmites australis* using unmanned aerial vehicles and neural networks in a semi-arid wetland. *River Res. Appl.* **2021**, *37*, 1312–1322. [CrossRef]
19. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.
20. Yu, C.; Gao, C.; Wang, J.; Yu, G.; Shen, C.; Sang, N. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. *Int. J. Comput. Vis.* **2021**, *129*, 3051–3068. [CrossRef]
21. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.

22. Li, Z.; Wang, Z.; Zhang, H.; Zuo, Y.; Liu, X.; Zhang, B.; Chen, Z.; Cheng, S. Light-Adapspnet: A Lightweight Adaptive Weighted Semantic Segmentation Network Based on Pspnet for Extracting Spartina Alterniflora in the Yellow River Delta Wetland. *Comput. Electron. Agric.* **2026**, *242*, 111338. [\[CrossRef\]](#)
23. Hu, Y.; Zhu, W.; Ren, G.; Wu, P.; Wang, J.; Zhu, W.; Xin, H.; Fu, S.; Xu, H.; Ma, Y. Remote Sensing Identification and Mapping of Spartina Alterniflora in Chinese Mainland Coastal at 2 m Spatial Resolution. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2026**, *19*, 399–420. [\[CrossRef\]](#)
24. Zhou, B.; Xu, M.; Tian, J.; Jia, M.; Mao, D.; Cheng, K.; Zhu, X.; Jiang, H.; Song, J.; Ke, Y.; et al. National-scale sub-meter mapping of *Spartina alterniflora* in mainland China 2020. *Earth Syst. Sci. Data* **2025**, *17*, 6601–6620. [\[CrossRef\]](#)
25. Siméoni, O.; Vo, H.V.; Seitzer, M.; Baldassarre, F.; Oquab, M.; Jose, C.; Khalidov, V.; Szafraniec, M.; Yi, S.; Ramamonjisoa, M. Dinov3. *arXiv* **2025**, arXiv:2508.10104.
26. Luo, M.; Zan, Y.; Khoshelham, K.; Ji, S. Domain generalization for semantic segmentation of remote sensing images via vision foundation model fine-tuning. *ISPRS J. Photogramm. Remote Sens.* **2025**, *230*, 126–146. [\[CrossRef\]](#)
27. Ma, Z.; Li, Y.; Ma, R.; Liang, C. Applying Unsupervised Semantic Segmentation to High-Resolution UAV Imagery for Enhanced Road Scene Parsing. *arXiv* **2024**, arXiv:2402.02985.
28. Jia, P.; Gao, Y.; Li, W.; Gao, Q.; Pan, F. AerialCLIP: Lightweight Open-Vocabulary Segmentation for UAV-Based Aerial Images. In Proceedings of the 2025 44th Chinese Control Conference (CCC), Chongqing, China, 28–30 July 2025; pp. 8193–8198.
29. Guo, M.-H.; Lu, C.-Z.; Hou, Q.; Liu, Z.; Cheng, M.-M.; Hu, S.-M. Segnext: Rethinking convolutional attention design for semantic segmentation. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 1140–1156. [\[CrossRef\]](#)
30. Rao, Y.; Zhao, W.; Zhu, Z.; Lu, J.; Zhou, J. Global filter networks for image classification. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 980–993.
31. Rahman, M.M.; Munir, M.; Marculescu, R. EMCAD: Efficient Multi-Scale Convolutional Attention Decoding for Medical Image Segmentation. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 11769–11779.
32. Li, J.; Zhang, H.; Chen, L.; He, B.; Chen, H. CSNet: A Remote Sensing Image Semantic Segmentation Network Based on Coordinate Attention and Skip Connections. *Remote Sens.* **2025**, *17*, 2048. [\[CrossRef\]](#)
33. Wu, X.; Li, W.; Hong, D.; Tao, R.; Du, Q. Deep Learning for Unmanned Aerial Vehicle-Based Object Detection and Tracking: A survey. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 91–124. [\[CrossRef\]](#)
34. Jo, W.-K.; Go, S.-H.; Park, J.-H. Optimizing Onboard Deep Learning and Hybrid Models for Resource-Constrained Aerial Operations: A UAV-Based Adaptive Monitoring Framework for Heterogeneous Urban Forest Environments. *Preprints* **2025**. [\[CrossRef\]](#)
35. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
36. Micikevicius, P.; Narang, S.; Alben, J.; Diamos, G.; Elsen, E.; Garcia, D.; Ginsburg, B.; Houston, M.; Kuchaev, O.; Venkatesh, G.; et al. Mixed Precision Training. *arXiv* **2017**, arXiv:1710.03740. [\[CrossRef\]](#)
37. Shi, S.; Wang, Q.; Xu, P.; Chu, X. Benchmarking State-of-the-Art Deep Learning Software Tools. In Proceedings of the 2016 7th International Conference on Cloud Computing and Big Data (CCBD), Macau, China, 16–18 November 2016; pp. 99–104.
38. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 3349–3364.
39. Xu, J.; Xiong, Z.; Bhattacharyya, S.P. PIDNet: A real-time semantic segmentation network inspired by PID controllers. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 19529–19539.
40. Wang, Z.; Zheng, J.-Q.; Zhang, Y.; Cui, G.; Li, L. Mamba-unet: Unet-like pure visual mamba for medical image segmentation. *arXiv* **2024**, arXiv:2402.05079.
41. Aleissae, A.A.; Kumar, A.; Anwer, R.M.; Khan, S.; Cholakkal, H.; Xia, G.-S.; Khan, F.S. Transformers in Remote Sensing: A Survey. *Remote Sens.* **2023**, *15*, 1860. [\[CrossRef\]](#)
42. Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F.; Brendel, W. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv* **2018**, arXiv:1811.12231.
43. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
44. Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; Joulin, A. Emerging Properties in Self-Supervised Vision Transformers. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 9630–9640.
45. Lin, T.-Y.; Goyal, P.; Girshick, R.B.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *42*, 318–327.

46. Liu, W.; Strong, D.R.; Pennings, S.C.; Zhang, Y. Provenance-by-environment interaction of reproductive traits in the invasion of *Spartina alterniflora* in China. *Ecology* **2017**, *98*, 1591–1599. [[CrossRef](#)] [[PubMed](#)]
47. Huang, Z.; Yan, H.; Zhan, Q.; Yang, S.; Zhang, M.; Zhang, C.; Lei, Y.; Liu, Z.; Liu, Q.; Wang, Y. A Survey on Remote Sensing Foundation Models: From Vision to Multimodality. *arXiv* **2025**, arXiv:2503.22081.

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.