remote sensing

MDPI

Article

Keypoint Detection-Based Aircraft Fine-Grained Recognition for High-Resolution Optical Remote Sensing Images

Qiantong Wang [1,2,3], Xiurui Geng [1,2,3,*], Peifeng Li [1,2,3], Lei Zhang [1,2,3], Ben Niu [1,2,3], Feng Wang [1,2,3], Guangyao Zhou [1,2,3] and Yuxin Hu [1,2,3]

[1] The Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; wangqt@aircas.ac.cn (Q.W.); lipf@aircas.ac.cn (P.L.); zhanglei@aircas.ac.cn (L.Z.); niuben@aircas.ac.cn (B.N.); wangf@aircas.ac.cn (F.W.); zhougy@aircas.ac.cn (G.Z.); yxhu@mail.ie.ac.cn (Y.H.)
[2] The Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100190, China
[3] The School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: gengxr@sina.com.cn

Highlights

What are the main findings?

- Keypoint detection proves effective in fine-grained aircraft recognition tasks in optical remote sensing images.
- Extracting stable and representative keypoints contributes to improving matching accuracy.

What are the implications of the main findings?

- Keypoint detection facilitates the quantitative description of targets.
- Keypoint provides a foundational basis for analysis by LLMs.

Abstract

Humans are capable of identifying aircraft based on quantitative features such as aspect ratio, engine count, wingspan, and structural configuration. Inspired by this, a keypoint-based aircraft identification approach is proposed to address the challenge of fine-grained aircraft recognition in high-resolution remote sensing images. First, a dataset of aircraft labeled with keypoints is built, in which aircraft are reclassified into types according to the similarity of keypoint distributions to improve extraction stability and versatility. Then, a keypoint extraction method with topological constraints is proposed, leveraging the nadir imaging characteristics of remote sensing and accounting for the relationships among keypoints. Subsequently, distinctive quantitative features for identification are selected through representativeness and effectiveness analyses for the following matching algorithm. Finally, a comprehensive template matching-based identification strategy is proposed to recognize targets based on quantitative descriptions derived from keypoints. This novel solution achieves significantly more accurate identification than traditional regression–classification approaches, improving recognition accuracy by over 3% on average. Moreover, the method extends aircraft identification capabilities from closed-set to open-set recognition, demonstrating substantial value for the precise interpretation of aircraft targets in high-resolution optical imagery.

Keywords: optical remote sensing; keypoint detection; quantitative features; template matching; fine-grained aircraft recognition

Academic Editor: Rongjun Qin

Received: 25 September 2025
Revised: 20 October 2025
Accepted: 22 October 2025
Published: 29 October 2025

Citation: Wang, Q.; Geng, X.; Li, P.; Zhang, L.; Niu, B.; Wang, F.; Zhou, G.; Hu, Y. Keypoint Detection-Based Aircraft Fine-Grained Recognition for High-Resolution Optical Remote Sensing Images. *Remote Sens.* **2025**, *17*, 3577. https://doi.org/10.3390/rs17213577

1. Introduction

With the development of optical satellites, visible light imaging has reached sub-meter, providing a stronger foundation for precise interpretation of aircraft. Fine-grained datasets are built with increasingly large sizes, such as DOTA [1], MAR20 [2], and FAIR1M [3]. High-resolution optical remote sensing images intuitively depict the appearance of aircraft targets. In recent years, deep learning algorithms based on convolutional neural networks (CNNs), such as SSD [4], YOLO series [5–9], R-CNN [10–12], and ViT and its variants [13–16], have been applied to aircraft detection and recognition tasks in optical imagery, and have achieved promising results.

However, due to the similar design principles of aircraft, which result in significant appearance consistency—particularly among aircraft of the same type or aircraft performing similar missions, target classification based purely on appearance features remains challenging, as shown in Figure 1. In contrast, comprehensive utilization of these quantitative properties by humans enables more refined aircraft identification. Consequently, researchers have gradually proposed utilizing structural information, keypoints, or templates to achieve fine-grained aircraft recognition, which has demonstrated improved performance in optical remote sensing applications, as shown in Figure 2.

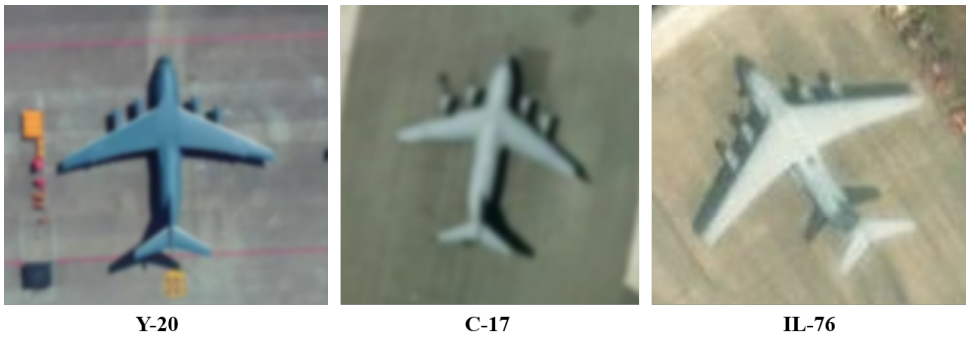


Figure 1. Three types of transport aircraft exhibit similar appearance characteristics.

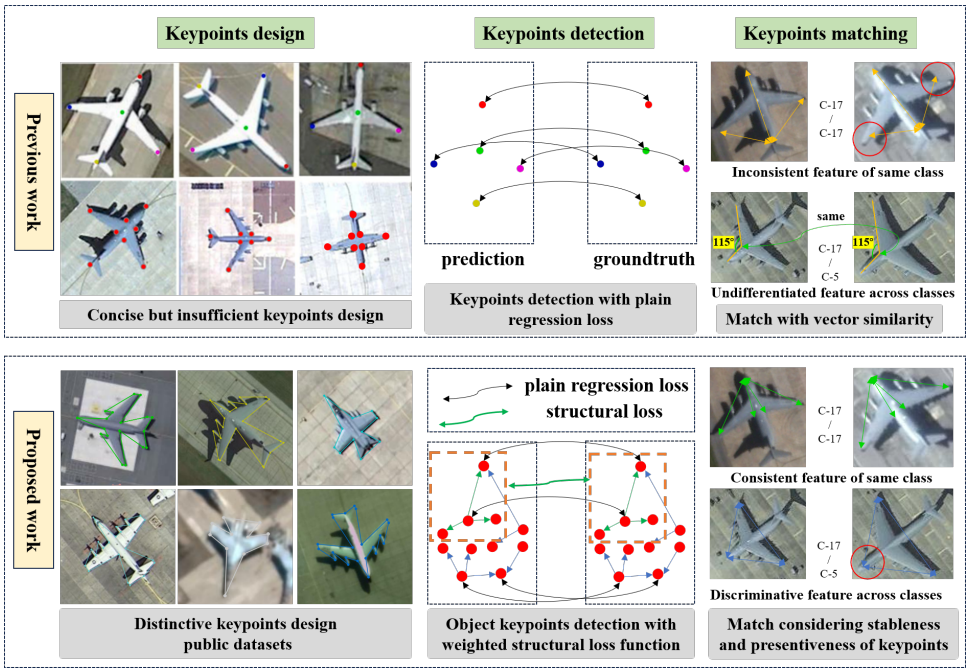


Figure 2. Main differences in keypoints design, detection, and matching, between previous work and the proposed. The colorful dots in the figure represent keypoints of aircraft.

To address challenges such as background noise, occlusion, and weather variations, researchers developed an aircraft identification algorithm employing Harris–Laplace corners, Zernike moments, and color-invariant moments [17]. This work utilized corners as discriminative features for aircraft characterization, although traditional feature extraction methods such as these are susceptible to data variations. Nevertheless, we established corners as viable distinguishing features for aircraft recognition in remote sensing imagery. During the same period, with the advancement of CNNs, keypoint-based facial recognition [18–21] achieved remarkable success, enabling both efficient and accurate facial recognition with recognition accuracy exceeding 99%.

Aircraft recognition shares similar problem characteristics with facial recognition. Several deep learning-based approaches have incorporated keypoints as integral components of their aircraft recognition solutions since 2017. In 2017, researchers proposed an accurate and efficient landmark-based aircraft recognition method [17]. They proposed an 8-point landmark design to characterize aircraft, coupled with a vanilla network architecture for comprehensive landmark regression to localize aircraft keypoints. Based on these network-extracted landmarks, they introduced a template vector-based matching algorithm to measure similarity between candidate aircraft and templates. In 2018, they further proposed an aircraft segmentation network to obtain refined segmentation results that provide critical details for distinguishing different aircraft types [22]. By integrating keypoint results from another processing branch, they performed comprehensive matching with templates using Intersection over Union (IoU) as the similarity metric between segmentation results and reference templates. Similarly, a Conditional Generative Adversarial Networks-based recognition algorithm was proposed in [23], which adopts the keypoints as a condition of Generative Adversarial Networks. Apart from keypoints extraction, an ROI feature extraction method is carefully designed to extract multi-scale features from the GAN in the regions of aircraft. After that, a linear support vector machine (SVM) classifier was adopted to classify each sample using their features. In 2021, a novel aircraft detection and recognition framework integrating component-level analysis [24] was proposed. The method comprises three core stages, data preprocessing, a foundational detection network, and dedicated structural component detection. Common yet distinguishable aircraft structures are explicitly detected as independent targets to support subsequent category inference. By leveraging component classification alongside structural component detection, the approach capitalizes on discriminative differences to enhance overall classification performance. In 2024, an integrated approach for aircraft model recognition was proposed, combining target segmentation and keypoint detection [25]. The methodology organically integrates multi-task deep neural networks with conditional random fields and template matching algorithms. The implementation involves three core phases, multi-task feature extraction and geometric refinement and template-based recognition.

Components recognition and part-to-whole reasoning methods show another possible solution for aircraft recognition. The Aircraft Reasoning Network (ARNet) [26], designed for aircraft detection and fine-grained recognition in remote sensing images (RSIs), incorporates prior knowledge employed in expert interpretation. With an Aircraft Component Discrimination Module (ACDM) that recognizes aircraft based on component features, classification performance is improved for both few-shot and easily confused categories. In 2024, a knowledge-driven deep learning method [27], called the Explainable Aircraft Recognition Framework Based on Part Parsing Prior (APPEAR), was introduced. It explicitly models the rigid structure of an aircraft as a pixel-level part parsing prior, dividing it into the nose, left wing, right wing, fuselage, and tail. This fine-grained prior provides reliable part locations to delineate aircraft architecture and imposes spatial constraints among parts, effectively reducing the search space for model optimization while identify-

ing subtle inter-class differences. Furthermore, the Multiscale Rotation Invariant Prototype Network (MSRIP-Net) [28] simulates the intuitive human reasoning process of identifying objects by segmenting them into multiple components. It automatically recognizes rigid components of aircraft targets without relying on additional part annotations, using only image-level class labels. In [29], the detector DFDet simultaneously focuses on mining contextual knowledge and mitigating angle sensitivity, constructing features containing multi-range contexts with low computational cost and aggregating them into a compact yet informative representation, thereby enhancing the model's robust inference capabilities.

Recently, the advancement of DeepSeek-like algorithms [30–32] has demonstrated that object recognition is fundamentally grounded in reasoning. Purely end-to-end approaches remain challenging for many tasks, especially fine-grained aircraft recognition; thus, these systems must draw inspiration from human cognitive processes involving part-based recognition, keypoint detection, and advanced reasoning with domain knowledge. This necessitates systematic community efforts to analyze aircraft-specific characteristics and develop their practical applications.

Current methodologies primarily face the following limitations.

1. Undefined part definitions. Part-based approaches face generalization challenges due to evolving aircraft design trends—blended wing–body designs and flying wing configurations are increasingly replacing traditional fuselage–wing separations, making explicit part definitions increasingly untenable.

2. Underutilized priors. Existing keypoint extraction algorithms predominantly rely on regression or classification paradigms without incorporating domain-specific constraints from remote sensing data and aircraft properties. Prior knowledge regarding structural invariances and imaging geometry characteristics remains underutilized.

3. Insufficient fine-grained matching. Current aircraft keypoint matching algorithms neglect the impact of aircraft roll angles on the relative positional variations among keypoints, which compromises accuracy. This necessitates analyzing stable and representative keypoints to enhance matching precision.

4. Limitations of monolithic approaches. Many aircraft possess highly analogous platform designs, and inherent errors in keypoint extraction processes collectively render sole reliance on keypoint matching inadequate for precise target differentiation. This necessitates the development of more comprehensive low-dimensional features to enhance discriminability.

These limitations highlight the urgent demand for a unified interpretation framework that incorporates remote sensing imaging characteristics, reasoning-based cognitive requirements, and human-inspired cognition with both qualitative and quantitative analytical capabilities.

Keypoints serve as a fundamental set of descriptive features applicable to aircraft target recognition algorithms based on attributes such as components, colors, or aspect ratios. They also represent a universal depiction method that captures aircraft outlines regardless of configuration. However, keypoint design is susceptible to evolving aircraft configurations and cannot maintain fixed definitions like facial keypoints [33]. Therefore, while keypoints effectively describe basic target features, further research on adaptive extraction algorithms remains essential.

In common keypoint detection and recognition algorithms, the task of predicting target keypoints is often treated as a regression task [34–38]. This approach is versatile for applications like human pose estimation and gesture recognition where viewpoints vary significantly. However, considering optical remote sensing tasks where the observation perspective is predominantly nadir imaging, and the presence of roll angles generally does not disrupt the target's relative structure, the prediction of keypoints for aircraft targets

should additionally incorporate constraints related to topological structure. This can help reduce the occurrence of outliers during keypoint prediction and further improve keypoint detection accuracy.

Furthermore, deficiencies remain in the current field of optical remote sensing aircraft target recognition, although many publicly available datasets for aircraft detection, recognition, or fine-grained recognition have been published and some studies have approached the problem from a keypoint perspective. On the one hand, keypoint design often lacks systematic methodology. On the other hand, there remains a scarcity of publicly available optical remote sensing image datasets specifically annotated with aircraft keypoints.

Since 2017, several types of keypoint deployment schemes have been proposed in different works primarily for aircraft pose estimation, as shown in Figure 2. In [25], the five keypoints defining aircraft geometry are sequentially designated as the aircraft nose, fuselage center, tail section, port-side wing extremity, and starboard-side wing extremity. This ordered set establishes the fundamental reference frame for aerodynamic analysis. Furthermore, eight landmarks for an aircraft are designed in [22], numbered from 0 to 7 in an anticlockwise direction with the aircraft head designated as point 0. In [23], a similar 8-point design scheme is utilized to generate polygon masks for aircraft targets. Additional keypoints generally improve matching accuracy.

In [22], a vanilla network is proposed to detect keypoints by implicitly encoding geometric constraints among landmarks through simultaneous regression of all landmarks. The Euclidean distance between ground truth and predictions is normalized by wingspan to formulate the loss function. During inference, target crops are rotated three times, generating four aircraft crops with different poses. Final keypoint detection results are generated by averaging the four landmark sets. In [23], a coarse segmentation network is proposed to segment aircraft from backgrounds. Furthermore, a fully connected CRF module is used to refine the coarse segmentation results. Finally, keypoints are extracted from the binary segmentation masks. In [25], Mask-RCNN is used to construct a multi-task network with three branches: a detection head, segmentation head, and keypoint detection head. In the detection prediction branch, a one-hot binary mask of size $M \times M$ is treated as ground truth, with cross-entropy loss regulating the prediction. In summary, keypoint detection research for aircraft targets and the success of facial keypoint recognition demonstrate that aircraft keypoint detection is worth developing and represents an important solution.

In [17], DT Nets are adopted to describe the distribution of feature points. By extracting multi-scale Harris–Laplace corners from the image, similarity is calculated based on triangle correspondences within the DT network. In [25], based on keypoint detection results, the normalized Sum of Squared Differences (SSD) Matching Method is adopted to calculate similarity between mask templates and predictions. Differences between the candidate target and all templates in the template library are computed and compared. The template model with the best matching performance is identified as the model of the target aircraft under test.

In the field of face matching, a feature vector is constructed from extracted keypoints. Finally, Euclidean distance is used to calculate the distance between facial features and template features.

The nadir imaging perspective in remote sensing provides inherent stability for aircraft targets, analogous to facial recognition conditions. To enhance keypoint recognition accuracy, algorithms must be adaptively optimized to leverage remote sensing advantages while accommodating its constraints.

Operational factors—including orbital mechanics and emergency observation requirements—frequently produce non-zero side-looking angles. This oblique imaging perspective alters topological relationships between aircraft keypoints. Consequently, viewpoint-

invariant keypoints must be identified that maintain consistent structural relationships across imaging conditions to ensure robust representation. Simultaneously, shared aircraft engineering principles create high visual similarity—especially in top-down views. Effective keypoint selection strategies must therefore identify distinctive keypoint groups that maximize inter-class differentiation.

In this manuscript, a novel comprehensive solution for aircraft keypoint detection in optical remote sensing imagery is proposed, as illustrated in Figure 3. The main contributions of this work are fourfold:

1. A large-scale aircraft keypoint dataset with thousands of aircraft is built, where 21 types of aircraft are carefully labeled. This provides a common foundation for all aircraft keypoint detection research in remote sensing.

2. Considering the characteristics of remote sensing imagery, a keypoint extraction algorithm for aircraft targets incorporating structural prior constraints is proposed, further enhancing the accuracy of aircraft keypoint extraction.

3. A simple yet effective selection method is proposed to identify representative and stable keypoints, laying a solid foundation for the design of target recognition algorithms.

4. Based on the keypoint set for aircraft targets, an algorithm measuring keypoint set consistency between candidates and templates is proposed, ensuring the precision of target matching and recognition by comprehensively utilizing point-to-point topological relationships.

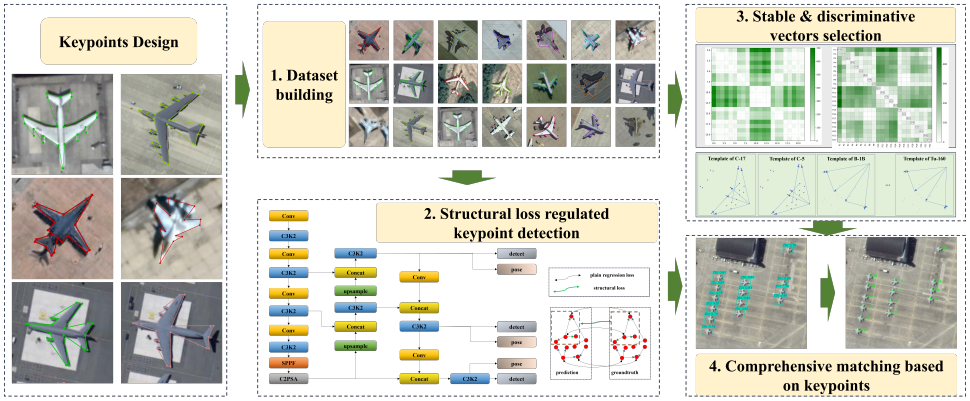


Figure 3. The proposed aircraft fine-grained recognition framework.

2. Materials and Methods

2.1. Basic Information of the Dataset

MAR20 is a remote sensing dataset for military aircraft recognition proposed in [2]. It contains 3842 images with 22,341 annotated instances across more than 20 distinct military aircraft types. Each instance is annotated with both horizontal and oriented bounding boxes. Owing to fine-grained categories belonging to the same type (e.g., transport planes), different aircraft often exhibit similar visual characteristics, resulting in significant inter-class similarity.

Building upon MAR20, a new keypoint aircraft dataset MAR20-KP is proposed in this work. Each aircraft is annotated with polygons using LabelMe according to predefined labeling protocols as shown in Figure 4. The annotation results are formatted as JSON files, with each target containing dual annotation modalities: the original rectangular bounding box and a polygonal boundary where each vertex represents a keypoint. Annotations commence at the aircraft nose, proceed clockwise along the airframe contour, and terminate at the starting point to form a closed polygon as shown in Figure 5.

Figure 4. Samples in the keypoint dataset. Five colors indicates five types of aircraft, each type of which shares similar design. Different colored lines are polygon annotation of different aircraft.

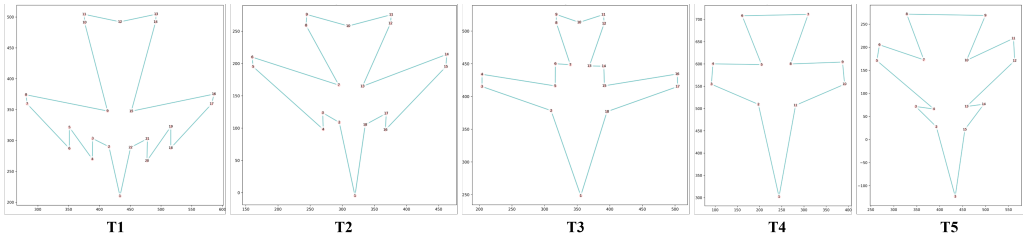


Figure 5. Five types of aircraft in the keypoint dataset. The numbers at the polygon vertices represent the indices of the key points.

To accommodate YOLO-Pose, these JSON files are converted into text format. Each line contains: target category index, coordinates of the top-left $(x_{\text{lu}}, y_{\text{lu}})$ and bottom-right $(x_{\text{rb}}, y_{\text{rb}})$ corners of the axis-aligned bounding box, plus 22 keypoints. Each keypoint is represented by normalized coordinates (x_i, y_i) and visibility flag $v_i \in \{0, 1, 2\}$ (denoting hidden, visible, and occluded states, respectively), structured as

$$\text{id}\ x_{\text{lu}}\quad y_{\text{lu}}\quad x_{\text{rb}}\quad y_{\text{rb}}\quad x_1\quad y_1\quad v_1\quad \cdots\quad x_{22}\quad y_{22}\quad v_{22} \tag{1}$$

For aircraft with fewer than 22 keypoints (e.g., X points), the visibility flags of the last $22 - X$ keypoints are set to 0 (hidden).

Based on MAR20, 20 fine-grained aircraft targets are further identified and reclassified into five types as shown in Table 1 and Figure 4.

Table 1. Target types in MAR20-KP.

Type	MAR20-Class	Class	Counts	Keypoints	Counts
T1	A2	C-130	1729	22	9939
	A3	C-17	1176		
	A4	C-5	642		
	A7	E-3	680		
	A8	B-52H	944		
	A9	P3-C	1086		
	A11	E-8	507		
	A14	KC-135	1778		
	A17	Tu-95	1397		
T2	A18	KC-10	308	18	308
T3	A6	Tu-160	441	18	1365
	A10	B-1B	924		
T4	A15	F-22/35	618	11	9083
	A16	F-18	2632		
	A13	F-15	1652		
	A5	F-16/2	1262		
	A19	Su-27/35	1236		
	A20	Su-24	981		
	A12	Tu-22	702		
T5	A1	Su-30/34	1646	15	1646

2.2. Reorganization of Aircraft Classes

Different from facial recognition keypoints, different aircraft exhibit distinct structural configurations. To ensure feature universality in keypoint settings, aircraft with similar structural layouts are categorized as the same type, sharing identical keypoint designs. Based on mainstream global aircraft designs, five keypoint configuration schemes are developed as shown in Figure 5.

Taking four-engine passenger/cargo aircraft as an example, these aircraft adhere to fundamentally consistent design layouts with engines pylon-mounted beneath the wings, featuring distinct dihedral angles between wings and fuselage. The aft fuselage incorporates either conventional tail assemblies or T-tail horizontal stabilizers. Both stabilizer configurations produce similar overall silhouettes when viewed from above. Consequently, targets with these characteristics are classified as a unified type employing a 22-keypoint scheme, including 1 nose point, 4 wing root points, 8 engine positioning points, 4 wingtip points, and 5 empennage reference points.

Conversely, two-engine passenger/cargo aircraft feature two fewer engines than their four-engine counterparts, resulting in an 18-keypoint scheme. The primary distinction lies in the reduction to 4 engine keypoints.

Due to their distinctive aerodynamic profiles, swept-wing aircraft utilize a dedicated 18-keypoint scheme configured along primary boundaries to comprehensively define their external contours.

Conventional single-engine fighters incorporate two main wing surfaces and two compact horizontal stabilizers that, combined with fuselage contours, form an 11-keypoint design. The canard configuration adds two delta wings to this baseline fighter layout, requiring 4 additional keypoints to position these canard surfaces, thereby establishing a 15-keypoint scheme.

2.3. Method

In this section, a keypoint detection algorithm and similarity-based aircraft recognition method are detailed. The pipeline of the proposed algorithm is illustrated in Figure 3 (Parts 2–4).

2.4. Aircraft Keypoint Detection with Topological Constraints

2.4.1. Basic Keypoint Detection Method

The YOLOv11-Pose model is adopted as the foundational keypoint detection framework, which shares core structural principles with CSPNet. This architecture integrates a keypoint prediction branch alongside the detection backbone.

In YOLO-Pose, the box-head and keypoint-head are two critical components. The box-head detects aircraft bounding boxes, while the keypoint-head estimates aircraft keypoint positions. Its loss function employs OKS (Object Keypoint Similarity) loss [39], a metric based on target keypoint similarity that optimizes pose estimation accuracy. Each bounding box corresponds to a keypoint set where each keypoint includes three parameters: coordinates and confidence. By jointly optimizing box-head and keypoint-head loss functions, YOLO-Pose accomplishes multi-aircraft pose estimation.

The backbone and neck for keypoint detection are identical to object detection architectures, with distinctions emerging only in the head layer. Within this head, apart from bounding box prediction tensors and classification prediction tensors, three distinct feature vectors of $3 \times K$ dimensions predict K aircraft keypoint positions and their confidence scores.

Fundamentally, Euclidean distance formulates the loss function.

$$d_{ij} = \sqrt{(x_{\text{pred},ij} - x_{\text{gt},ij})^2 + (y_{\text{pred},ij} - y_{\text{gt},ij})^2} \tag{2}$$

To handle the invisible keypoints, klf_i and e_{ij} are utilized to normalize the loss from the perspective of keypoint number and size of target.

$$\text{klf}_i = \frac{N}{\sum_{j=1}^{N} \mathbf{1}_{\{\text{mask}_{ij} \neq 0\}} + \epsilon}, e_{ij} = \frac{d_{ij}}{8 \cdot \sigma_j^2 \cdot (A_i + \epsilon)} \tag{3}$$

where ϵ is a small constant that prevents division by zero.

The keypoints loss function is finally formed as follows.

$$\mathcal{L}_{\text{kpt}} = \frac{1}{M} \sum_{i=1}^{M} \left[\text{klf}_i \cdot \sum_{j=1}^{N} \left(\left(1 - e^{-e_{ij}}\right) \cdot \mathbf{1}_{\text{mask}_{ij}} \right) \right] \tag{4}$$

2.4.2. Weighted Keypoint Location Loss

While considering that some keypoints are not localized easily due to blur, shadows, or occlusion, it is noted that humans typically locate primary keypoints first as references for harder-to-detect points. Accordingly, σ_j are set to different values by designating nose and

fuselage keypoints as control points. During training, higher weights are assigned to the loss terms of these keypoints. This encourages the network to prioritize aircraft structural pose estimation, thereby improving localization accuracy for non-salient keypoints. Keypoint error weights are configured as shown in Table 2.

Table 2. σ settings for different keypoints of different types of aircraft.

Weights	T1	T2	T3	T4	T5
$\sigma = 5$	1, 2, 12, 22	1, 2, 10, 18	1, 10, 18	1, 7, 8	1, 6, 9
$\sigma = 1$	else	else	else	else	else

2.4.3. Structural Constrain Loss

Aircraft exhibit highly stable geometric topology, with keypoints inherently connected to airframe components. Integrating these structural priors significantly enhances detection robustness against viewpoint variations and partial occlusion. During training, structural constraints regulate the generated keypoints to suppress outliers and ensure geometric consistency. Therefore, aircraft structural constraints are integrated into the keypoint detection framework as illustrated in Figure 6.

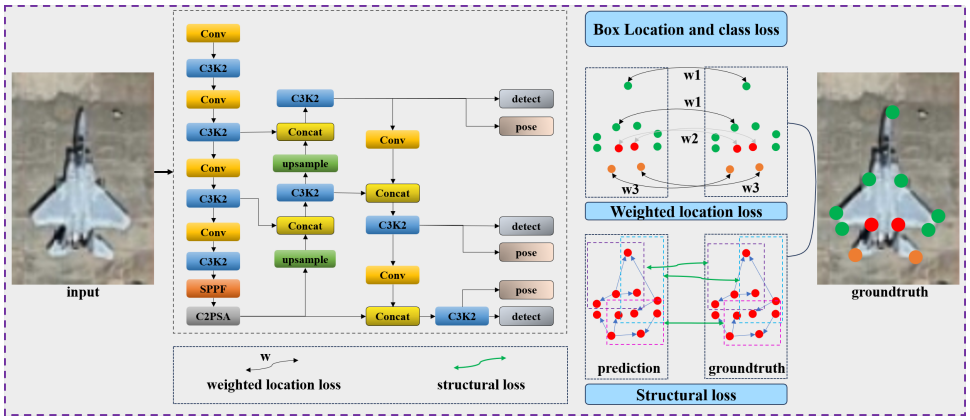


Figure 6. Keypoints detection network with weighted OKS loss and structural loss. The colorful dots are keypoints.

The predicted keypoints $\mathbf{P}$ and ground truth keypoints $\mathbf{G}$ are defined as $[\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n]^T \in \mathbb{R}^{n\times 2}$ and $[\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n]^T \in \mathbb{R}^{n\times 2}$, respectively.

Structural features of the predicted keypoints $\mathbf{S}^{\text{pred}} \in \mathbb{R}^{n\times n}$ and ground truth keypoints $\mathbf{S}^{\text{target}} \in \mathbb{R}^{n\times n}$ are defined as

$$\mathbf{S}^{\text{pred}} = \begin{bmatrix} d(\mathbf{p}_1, \mathbf{p}_1) & \cdots & d(\mathbf{p}_1, \mathbf{p}_n) \\ \vdots & \ddots & \vdots \\ d(\mathbf{p}_n, \mathbf{p}_1) & \cdots & d(\mathbf{p}_n, \mathbf{p}_n) \end{bmatrix} \tag{5}$$

where $d(\mathbf{p}_i, \mathbf{p}_j) = \|\mathbf{p}_i - \mathbf{p}_j\|_2$. $\mathbf{S}^{\text{GT}}$ is generated similarly.

The structural loss $\mathcal{L}_{\text{structure}}$ is computed as follows:

$$\mathcal{L}\text{structure} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\mathbf{S}^{\text{pred}}ij - \mathbf{S}^{\text{GT}}ij\right|_1 \tag{6}$$

where n is the number of keypoints, $\mathbf{p}_i = (x_i^{\text{pred}}, y_i^{\text{pred}})$ and $\mathbf{g}_i = (x_i^{\text{gt}}, y_i^{\text{gt}})$ are coordinates of keypoints, and $\|\cdot\|_2$ denotes the Euclidean distance.

Finally, the loss function for keypoints prediction incorperates weighted $\mathcal{L}_{\text{kpt}}$ and structural loss $\mathcal{L}_{\text{structure}}$.

$$\mathcal{L}_{\text{keypoints}} = \alpha * \mathcal{L}_{\text{kpt}} + (1 - \alpha) * \mathcal{L}_{\text{structure}} \tag{7}$$

where α is set as 0.7 in this work.

2.5. Consistent and Distinctive Keypoints Identification

From an intuitive perspective, keypoint matching between targets and database entries could employ distance-based, graph similarity, or vector distance methods. However, in remote sensing imagery, keypoint distributions exhibit slight instability under varying imaging conditions due to side-view angles. Additionally, certain aircraft targets possess highly similar visual characteristics, necessitating identification of distinctive keypoint combinations. To address these challenges, a variance-based analytical method for evaluating keypoint stability and representativeness is introduced. This methodology determines optimal keypoint sets for subsequent matching stages through quantitative assessment.

Samples capturing diverse side-view angles of 20 common aircraft targets were first collected. Building upon this, geometric topology graphs of keypoints were extracted and normalized by vertically aligning fuselage directions for each aircraft type. Following orientation correction, topology graphs of identical aircraft models were spatially scaled to uniform dimensions.

2.5.1. Consistent Keypoint Identification

For optical remote sensing imagery, the primary factor affecting keypoint distributions and relative positional relationships is side-looking angle. Assuming specific distribution patterns exist under pure nadir imaging conditions, minor-to-significant variations may occur in partial keypoints when imaged at oblique angles or varying incidence angles. This section proposes an efficient metric to quantitatively characterize viewing angle impacts on spatial distributions of target keypoints.

By collecting images of the same target from multiple viewpoints and keypoints on these images, as illustrated in Figure 7, a keypoint set is constructed as $\mathcal{S}_C = \{S_C^1, S_C^2, \ldots, S_C^N\}$ for each class C, where N denotes the number of distinct viewing angles. Each $S_C^i = \{p_1^i, p_2^i, \ldots, p_K^i\}$ represents the keypoint set, including K keypoints, under the i-th observation condition.

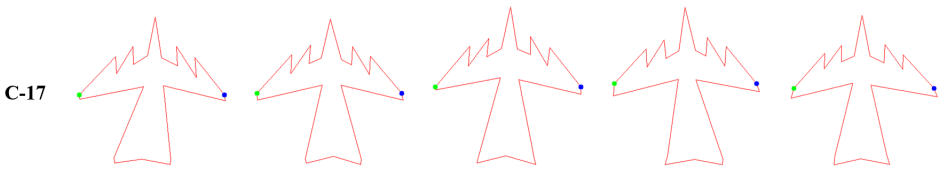
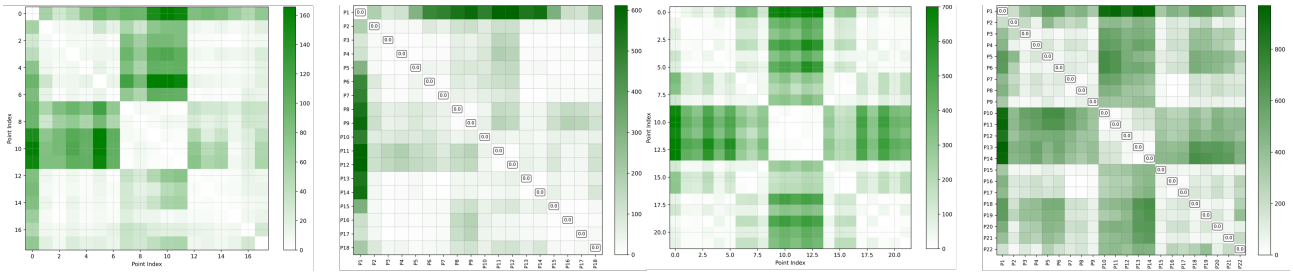


Figure 7. Different keypoint geometric topology graphs of C-17 extracted and normalized by aligning fuselage directions vertically.

A vector matrix $V = [V_1, V_2, \ldots, V_K]$ is established using each point as the datum point.

$$V_i = \begin{bmatrix} p_1^1 - p_i^1 & p_1^2 - p_i^2 & \cdots & p_1^N - p_i^N \\ p_2^1 - p_i^1 & p_2^2 - p_i^2 & \cdots & p_2^N - p_i^N \\ \vdots & \vdots & \ddots & \vdots \\ p_K^1 - p_i^1 & p_K^2 - p_i^2 & \cdots & p_K^N - p_i^N \end{bmatrix} \tag{8}$$

As shown in Figure 8, the fluctuation within each row of matrix V_i indicates the stability of the spatial relationship between the datum keypoints i and other keypoints. Smaller fluctuation corresponds to higher stability, measured with variance. When the fluctuation between p_K and p_i falls below a predetermined threshold τ_1, keypoints p_i and p_K are considered to form a stable vector sv_j. If it exceeds the threshold, the group is unstable, which connot be adopted in the following matching algorithm. Through multi-scenario stability assessment matrices constructed for all keypoints, the stable keypoint vector set for class C is formed as $SKV_C = \{sv_1, sv_2, \cdots, sv_x\}$.



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

Figure 8. (**a**) Consistent map of Tu-160 and distinctive map for Tu160 and B1-B, with 18 keypoints. (**b**) Consistent map of C-17 and distinctive map for all aircraft in T1, with 22 keypoints. In the consistent map, greener color indicates that the keypoint is less stable under varying conditions. In the distinctive map, greener color signifies that the keypoint has stronger representativeness across different categories.

2.5.2. Distinctive Keypoints Identification

Aircraft of the same type share similar design; thus, some features of them are alike. To find out the representative keypoints and vectors, a quantitative method is proposed to characterize representativeness of keypoints across different types.

By collecting images of the same target type from identical perspectives and annotating keypoints, as shown in Figure 9, a keypoint set $\mathcal{S}_T = \{S_T^1, S_T^2, \ldots, S_T^M\}$ is constructed for each target type T, where M denotes distinct subtypes, and $S_T^i = \{p_1^i, p_2^i, \ldots, p_K^i\}$ represents the keypoint set for the i-th class.

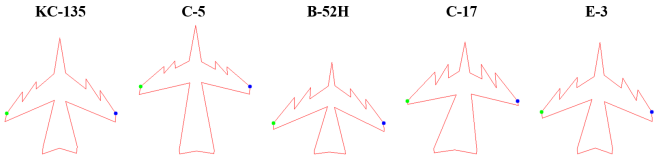


Figure 9. Keypoint geometric topology graphs within type $T1$ were extracted and normalized by aligning fuselage directions vertically.

Similarly to last section, a vector matrix $V = [V_1, V_2, \ldots, V_K]$ with each point as the datum point is established to analysis the representive keypoints.

$$V_i = \begin{bmatrix} p_1^1 - p_1^1 & p_1^2 - p_1^2 & \cdots & p_1^M - p_1^M \\ p_2^1 - p_1^1 & p_2^2 - p_1^2 & \cdots & p_2^M - p_1^M \\ \vdots & \vdots & \ddots & \vdots \\ p_K^1 - p_1^1 & p_K^2 - p_1^2 & \cdots & p_K^M - p_1^M \end{bmatrix} \tag{9}$$

The fluctuation within each row of V_i indicates the stability between corresponding keypoint pairs $p_K - p_i$ across subtypes. Greater fluctuation corresponds to higher value as

representative keypoints, measured with variance. When it exceeds a predetermined threshold τ_2, the keypoint pair is considered as representative vector rv_j. Conversely, if below τ_2, the pair is deemed non-representative and excluded from target similarity measurements.

By constructing multi-type stability assessment matrices for all keypoints, the representative keypoint vector set for target type T is formed as $RKV_T = \{rv_1, rv_2, \cdots, rv_y\}$.

2.5.3. Vector Set for Matching Algorithm

Finally, the stable keypoint vector set is sorted by variance in descending order from the stability map, selecting points with lower variances and eliminating those with higher variances to obtain stable keypoints. Subsequently, on the distinctiveness map, the representative keypoint set is sorted by variation in descending order, selecting points with greater variation to obtain distinctive points that effectively differentiate target classes. The intersection of representative and stable keypoints yields the optimal keypoint vector S_{ma} for match-based aircraft identification.

$$S_{Ma} = SKV_T \cap RKV_C \tag{10}$$

As shown in Figure 8, the greener the color indicates a greater degree of change. On the left image, lighter-colored points should be selected as stability keypoints for each class of targets, while on the right image, darker-colored points are typically chosen as representative keypoints for the same type.

2.6. Comprehensive Matching Based on Keypoints

Given predicted keypoints set P output by an algorithm and corresponding type T_p, the method automatically associates P with the candidate objects $\{Tp_1, Tp_2, \ldots, Tp_M\}$ representing keypoint sets of individual aircraft of that type, where i denotes i_{th} class aircraft in type T_p.

To establish comparable topological representations, the keypoint set P is first normalized to align with the scale and orientation of the keypoints template Tp_i, minimizing the effects of size and angular variations on target structural consistency. The following four principles are applied to accurately match candidate P and template Tp_i.

The final matching result is determined through a comprehensive evaluation of the following five attributes, including Aspect Ratio (AR), Angular Deviation (AD), Fuselage Segment Proportion (FSP), Aligned Topological Structure Distance ($ATSD$) and Color Consistency CC. The overall score SC_i^A of class i is defined as

$$SC_i^A = Score_i^{AR} + Score_i^{Ang} + Score_i^{FSP} + Score_i^{ATSD} \tag{11}$$

The class i^* with the highest confidence score is identified as the target's class.

2.6.1. Aspect Ratio

Since remote sensing images provide top-down views of aircraft targets, the aspect ratio of targets can be accurately characterized through keypoints, facilitating candidate set reduction. Aspect ratio is one of the simplest yet most effective metrics for target classification. For certain aircraft types, distinctive aspect ratios often enable direct identification. Aspect ratio is formulated as AR = Length/Width. For different target types, length and width calculations are performed as shown in Table 3.

Absolute difference between the AR of candidates P and keypoints template Tp_C is calculated and sorted in descending order. The three categories d_1, d_2, d_3 with the highest rankings are considered as candidate types. The score of each type $Score_i^{AR}, i \in \{C_1, C_2, \ldots, C_K\}$ is defined as $5, 3, 1, 0$ for d_1, d_2, d_3, and others.

Table 3. Calculation method for length and width.

	T1	T2	T3	T4	T5
Length	$\|\overrightarrow{P_1P_{12}}\|$	$\|\overrightarrow{P_1P_{10}}\|$	$\|\overrightarrow{P_1P_{10}}\|$	$\|\overrightarrow{P_1P_{6,7}}\|$	$\|\overrightarrow{P_1P_{8,9}}\|$ [1]
Width	$\|\overrightarrow{P_8P_{16}}\|$	$\|\overrightarrow{P_6P_{14}}\|$	$\|\overrightarrow{P_4P_{16}}\|$	$\|\overrightarrow{P_4P_9}\|$	$\|\overrightarrow{P_5P_{11}}\|$

[1] $P_{8,9}$ indicates the middle point between P_8, and P_9.

2.6.2. Angular Deviation

Due to errors in keypoint extraction and side-view imaging affecting aspect ratio accuracy, it is necessary to further incorporate angular factors to calculate similarity between candidates and templates. Sweep angles and other angular features are representative characteristics that can be calculated through vectors between keypoints. Absolute differences between angular vectors $Ang = [A_1, A_2, \ldots, A_m]$ of candidate P, as shown in Table 4, and template Tp_C are calculated and sorted in descending order. The three types d_1, d_2, d_3 with highest rankings are considered candidate types. The angular score for each type $Score_i^{\text{Ang}}$ is defined analogously to $Score_i^{\text{AR}}$.

Table 4. Formualtion of angular vectors for different types.

Ang	T1	T2	T3	T4	T5
A_1	$\theta_{1,2,7}$	$\theta_{1,2,5}$	$\theta_{1,2,3}$	$\theta_{1,2,3}$	$\theta_{1,4,5}$
A_2	$\theta_{8,9,10}$	$\theta_{6,7,8}$	$\theta_{4,5,6}$	$\theta_{4,5,6}$	$\theta_{6,7,8}$
A_3	$\theta_{13,15,16}$	$\theta_{12,13,14}$	$\theta_{14,15,16}$	$\theta_{7,8,9}$	$\theta_{9,10,11}$
A_4	$\theta_{17,22,1}$	$\theta_{15,18,1}$	$\theta_{17,18,1}$	$\theta_{10,11,1}$	$\theta_{12,13,1}$

2.6.3. Fuselage Segment Proportion

Fuselage are divided finto three segments Sec_1, Sec_2, Sec_3 as Table 5.

Table 5. Formualtion of fuselage segment proportion for different Types.

	T1	T2	T3	T4	T5
Sec_1	$\|\overrightarrow{P_1P_{12}}\|$	$\|\overrightarrow{P_1P_{10}}\|$	$\|\overrightarrow{P_1P_{10}}\|$	$\|\overrightarrow{P_1P_{6,7}}\|$	$\|\overrightarrow{P_1P_{8,9}}\|$ [1]
Sec_2	$\|\overrightarrow{P_8P_{16}}\|$	$\|\overrightarrow{P_6P_{14}}\|$	$\|\overrightarrow{P_4P_{16}}\|$	$\|\overrightarrow{P_4P_9}\|$	$\|\overrightarrow{P_5P_{11}}\|$
Sec_3	$\|\overrightarrow{P_8P_{16}}\|$	$\|\overrightarrow{P_6P_{14}}\|$	$\|\overrightarrow{P_4P_{16}}\|$	$\|\overrightarrow{P_4P_9}\|$	$\|\overrightarrow{P_5P_{11}}\|$

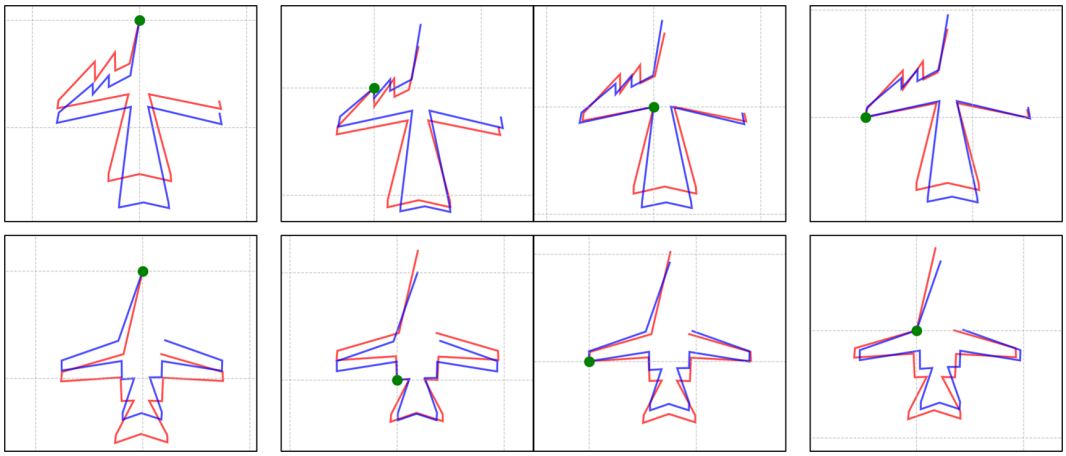
[1] $P_{8,9}$ indicates the middle point between P_8, and P_9.

The ratio $Sec_1 : Sec_2 : Sec_3$ is then calculated and compared with that of templates. The three categories d_1, d_2, d_3 with the highest rankings are considered as candidate types. The score of each type $Score_i^{FSP}$ is defined similarly as $Score_i^{AR}$.

2.6.4. Aligned Topological Structure Distance

This approach enables visual comparison of shape differences across aircraft categories, particularly regarding matching degree when aligned at different keypoints. It facilitates the understanding of geometric characteristics across aircraft types and enables identification of key distinguishing features.

The core methodology employs a graph-based topological representation where each keypoint serves as a central node to construct graph structures $\{G_i\}$. Vectors falling outside the representative, stable keypoint set S_{Ma} are excluded from calculations. This multi-perspective approach amplifies discriminative features between target types by considering varied centrality viewpoints, mirroring human cognitive processes in multi-angle comparative analysis. Different reference points yield distinct error profiles when comparing target structures as shown in Figure 10.



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

Figure 10. Visualization of errors with respect to different reference points. (**a**) Amplified errors with respect to distinctive keypoints; (**b**) Limited errors with respect to distinctive keypoints.

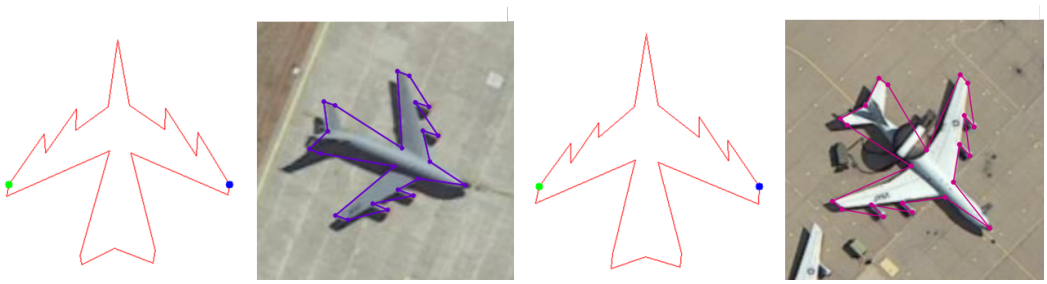
The optimal matching target is determined by

$$i^* = \arg \max_{i \in \{1,2,\ldots,K\}} \sum_{j=1}^{X} \left| G_{\text{pre}}^{j} - G_{GT_i}^{j} \right| \tag{12}$$

where G_{pre}^{j} denotes the graph constructed from predicted keypoints with the j-th keypoint as the central node, $G_{\text{GT}_i}^{j}$ represents the graph constructed from keypoints of the i-th candidate target with the j-th keypoint as the central node, and X is the total number of keypoints. This formulation identifies candidate targets maintaining minimal structural divergence across all centrality perspectives. The three targets d_1, d_2, d_3 with the smallest errors are regarded as candidate categories. The score $Score_i^{\text{ATSD}}$ for each type is defined analogously to $Score_i^{\text{AR}}$.

2.6.5. Color Consistency

Keypoints and their mutual relationships reflect target structural information. However, when two target categories exhibit nearly identical structures, keypoint-based fine-grained aircraft recognition algorithms fail. For example, the KC-135 tanker and E-3 AWACS—both modified from the Boeing 707 platform as shown in Figure 11—share identical airframe structures. Thus, a component color-based identification method is further proposed as follows: if wings are white, targets are more likely classified as E-3; conversely, KC-135 otherwise. Benefiting from accurate dense keypoints, wings can be precisely localized with colors determined simultaneously.



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

Figure 11. KC-135 (**a**) and E-3 (**b**) share similar topological structure.

2.6.6. Actual Dimensions

Since remote sensing is a form of quantitative observation, the actual ground resolution of an image can be quantitatively determined. Based on acquired image resolution and pixel spatial dimensions, a target's actual dimensions can be calculated. Although affected by observation errors and keypoint extraction inaccuracies, these dimensions still enable distinction between structurally similar targets with significant size differences. However, due to the absence of raw geographic information in this study's dataset, this evaluation criterion was not utilized.

3. Results

To validate the proposed method, groups of experiments are conducted based on the proposed keypoints dataset.

3.1. Keypoints-Based Recognition

Based on statistical analysis of target types and quantities in the dataset, the data is partitioned into training, and testing sets similarly as MAR20, which is 1331:2511. Model training was conducted with a 12 vCPU Intel® Xeon® Platinum 8352V CPU @ 2.10 GHz and a vGPU-32 GB. Next, with the keypoints prediction, matching algorithms are conducted to find out the classes.

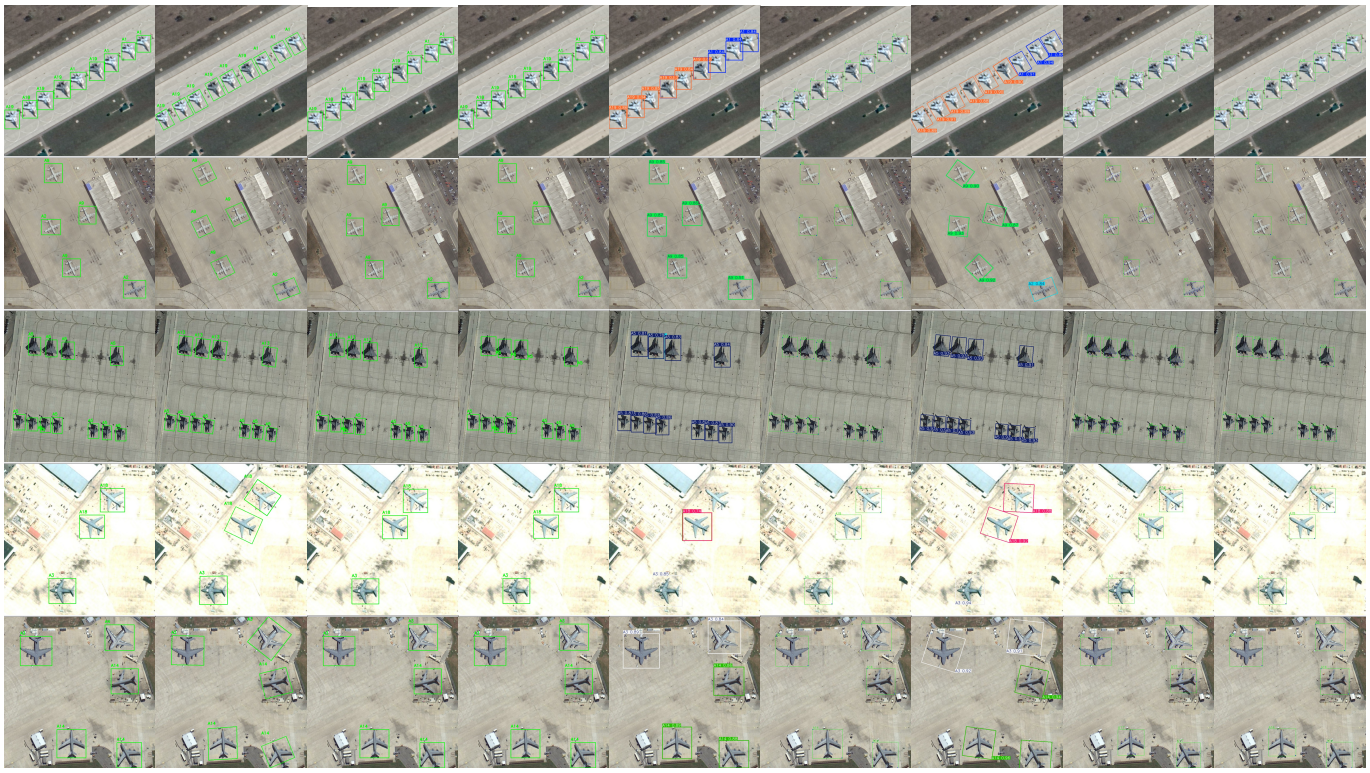
In inference phase, the prposed method overall comprises two key steps. The first stage is YOLO-Pose-based keypoint extraction, whose computational complexity is almost identical to that of general YOLO algorithms. The second stage is multi-dimensional feature calculation along with feature vector matching for aircraft targets. A total of 540 images are randomly selected for testing, containing a total of 6302 targets. The average recognition time per image slice is 33.4 ms, while the recognition time per target is 2.8 ms.

3.1.1. Closed-Set Aircraft Recognition

Comparative experiments are conducted with purely bounding box-based detection and recognition algorithms RoI Transformer [40], Oriented R-CNN [41], TSD [42], FDY-OLOV8 [43], YOLOS [16], YOLO11, and YOLO11-OBB [9]. RoI Transformer [40] applies spatial transformations to RoIs to adapt to object detection in densely arranged and rotating targets within remotely sensed imagery, demonstrating strong versatility. Oriented R-CNN [41] is an efficient two-stage rotated object detector that directly generates high-quality rotated proposals through an Oriented Region Proposal Network (Oriented RPN), significantly reducing computational costs while balancing detection accuracy and speed. TSD [42] is a two-stage detector that optimizes feature representations for classification and regression tasks through its task-aware spatial disentanglement head structure, enhancing detection robustness in complex scenes. FD-YOLOv8 [43] proposed a local detail feature module and a focus modulation mechanism to improve semantic information and local and global features, and achieved high mAP in a public dataset. YOLOS [16] utilizes the Vision Transformer (ViT) module for feature extraction while leveraging the advantages of a single-stage detector. YOLOv11 [9] is the most lightweight algorithm in the YOLO series with optimal comprehensive performance, making it ideal for real-time processing of large-data tasks like remote sensing image interpretation.

As aircraft classes with similar features are clustered into five types, it is much easier for the detector to decide which type an aircraft blongs to. The MAP of five types of aircraft are higher than each class within certain types, indicating better overall performance as shown in Table 6. For the aircraft targets in this dataset, experimental results of our proposed matching recognition algorithm on test data are substantially superior to those achieved by detection networks as shown in Table 7 and Figure 12. This solution avoids the

sample imbalance problem caused by insufficient samples of certain categories. Severely imbalanced aircraft categories cause ineffective models, such as E-8, whose MAP is quite lower than other aircraft in previous work. With the proposed algorithm, E-8 gets quite a high MAP since it is seen as *T*1 similar with other four-engine planes in the first stage followed with a match-based recognition algorithm. Because type T2 and type T5 only include KC-10 and Su-30, respectively, The class recognition MAP is the same as Type recognition MAP. While the number of T2, including KC-10, is much less than that of the other types, its MAP is still relatively lower than that of other types, suffering from the class-imbalance.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**) (**g**) (**h**) (**i**)

Figure 12. Recognition results of different algorithms, (**a**) RoI Transformer [40], (**b**) Oriented R-CNN [41], (**c**) TSD [42], (**d**) FDYOLOV8 [43], (**e**) YOLOS [16], (**f**) YOLO11, (**g**)YOLO11-OBB [9], (**h**) the proposed and (**i**) groundtruth.

Table 6. AP of five aircraft types.

	T1	**T2**	**T3**	**T4**	**T5**	**All-Types**
AP	98.9	94.2	99.3	94.7	95.2	96.5

Algorithms based on Vision Transformer (ViT) are adept at capturing the relationships between patches, which, to some extent, aligns with the human cognitive process of recognizing airplanes—for instance, identifying components like wings, nose, and tail. The algorithm performs well in recognizing the E-3 AWACS due to its distinctive rotodome structure, indicating that the ViT's capability to establish relationships between target patches is particularly effective for recognizing targets with distinctive structures. However, due to the high structural similarities among certain aircraft, relying solely on patch relationships makes accurate differentiation challenging.

Table 7. MAP of the proposed method and traditional detection methods. The bold indicates the first and second high MAP.

Class	Aircraft	[40]	[41]	[42]	[43]	[16]	[9]	[9]-obb	Ours
A2	C-130	81.53	81.73	86.7	98.3	97.6	98.6	**99.0**	**99.2**
A3	C-17	87.61	88.08	93.9	94.7	97.1	**98.5**	98	**98.7**
A4	C-5	78.33	69.57	71.8	**97.1**	96.9	95.9	96.4	**98.1**
A7	E-3	90.24	90.49	96.5	86.4	99	**99.2**	**99.1**	98.7
A8	B-52H	87.58	89.54	91.3	**98.8**	98.8	98.6	97.3	**99.2**
A9	P-3C	87.93	89.78	89.2	**98.1**	97.1	96	97.6	**99.4**
A11	E-8	85.88	87.62	87.3	82.4	88.5	**89.5**	88.6	**96.1**
A14	Tu-95	90.46	90.59	95.5	94.6	98.6	98.2	**98.3**	**99.2**
A17	KC-135	88.2	88.5	89.6	**97.2**	95.1	96.1	96.2	**97.9**
A18	KC-10	74.59	70.5	58.5	91.7	83.6	84.6	**92.4**	**94.2**
A6	Tu-160	90.49	89.92	93.6	87.1	98.1	**98.8**	98.2	**98.9**
A10	B-1B	90.89	90.91	98.1	**99.3**	97.9	99.1	99.2	**99.4**
A15	F-16/2	80.45	75.61	75.7	**90.6**	83.1	79	82.9	**90.1**
A16	Tu-22	89.29	88.39	87.8	87.8	93.6	**94.6**	92.6	**95.2**
A13	F-15	67.24	67.52	67.1	90.6	90.3	91.7	**92.0**	**96.2**
A5	F-22/35	47.85	46.33	34.5	**77.9**	75.4	73.7	72.1	**89.9**
A19	F-18	89.11	88.27	91	**99.1**	96.4	97	97.1	**97.7**
A20	Su-27/35	81.3	78.72	79.6	84.8	85.1	**87.7**	87.4	**89.2**
A12	Su-24	80	80.25	80.1	81	87.9	**89.2**	88.6	**90.9**
A1	Su-30/34	85.4	81.73	87.2	86	92.2	**90.2**	**90.2**	**95.2**

3.1.2. Experiments on Datasets Built on GF-07 Images

A small-scale test dataset is further constructed by extracting aircraft targets from the GF-07 satellite. A total of 1500 crops with 0.5 0.8 m resolution containing eight kinds of aircraft are selected, including A2, A3, A4, A7, A9, A11, A7, and P-8A, to evaluate fine-grained recognition performance. Following the proposed annotation method, key-points are annotated on the targets to create training and testing datasets with a 7:3 ratio. Experimental results, in Figure 13 and Table 8, demonstrate that our proposed approach achieves good performance.

Figure 13. Results of the proposed algorithm on datasets built on GF-07 images.

Table 8. Quantitative evaluation results of the proposed algorithm on datasets built with the GF-07 satellite.

Train	Test	Type	Class	MAP
486	307	A2	C-130	98.9
377	276	A3	C-17	97.1
251	234	A4	C-5	97.5
276	221	A7	E-3	98.1
421	268	A9	P-3C	96.1
286	231	A11	E-8	98.1
532	319	A17	KC-135	97.2
607	338	P-8A	P-8A	99

3.1.3. Open-Set Aircraft Recognition

Furthermore, the validation is extended to other aircraft A400-M (Europe) and An-22, An-12, IL-76 (Russia) for open-set recognition. The experimental results indicate that traditional detection and recognition algorithms are usually incomparable to the keypoint-based method in recognition performance. For type recognition, the established contrastive template mechanism enables accurate target classification without requiring labeled samples, while traditional methods frequently confuse these types with categories within their supported range. In Figure 14a, An-125, IL-76 and A400-M are misidentified as A3 (C-17), since they are all transport aircraft with similar structural designs. A-12 is misidentified as A9 (P-3C), since they share similar similar wing configurations. The An-22 is missed during detection because its design differs significantly from conventional turboprop aircraft like the C-130 and P-3C. While, with the proposed algorithm, most of the out-set aircraft are correctly detected and recognized, as detection networks focus on acquiring universal discrete characteristics instead of holistic representations. The complex recognition problems are transformed into quantitative parameter comparisons, thereby ensuring precision, as shown in Figure 14b.

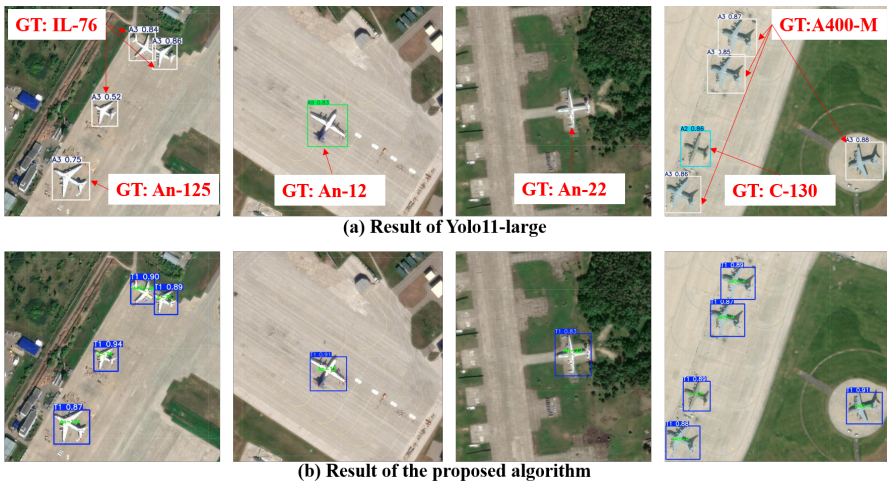


Figure 14. Detection-based results of Yolo11 and the proposed algorithm in open-set test datasets.

3.2. Ablation Study

In this section, ablation studies are further conducted to evaluate the proposed, weighted, structurally constrained keypoint loss, validating their contribution to enhancing keypoint extraction. Additionally, we investigate the effect of implementing a comprehensive matching solution, especially with a simple multi-point cyclic matching strategy, considering the representativeness and stability of keypoints during the recognition stage.

3.2.1. Keypoint Detection Percision with Different Loss Function

To assess keypoint recognition accuracy, OKS is adopted to measure the accuracy of keypoints prediction. Experimental results indicate that structural loss and weighted location loss is positive to develop the performance of keypoint detection. As shown in Figure 15 and Table 9, compared with that of normal keypoints detection algorithms (Figure 15a), keypoint detection results of the proposed algorithms with weighted loss function (Figure 15b) are more in line with the actual distribution, such as the overall direction. Further, the results of algorithms with structural loss (Figure 15c) are more closely aligned with the actual contour.

Table 9. Overall performance with different keypoint loss function.

	Normal Loss	Weighted Loss	Structural Loss
Points percision	96.1%	96.9%	98.2%

With the structural loss and weight loss, overall direction in the keypoint prediction is more consistent with the ground truth.

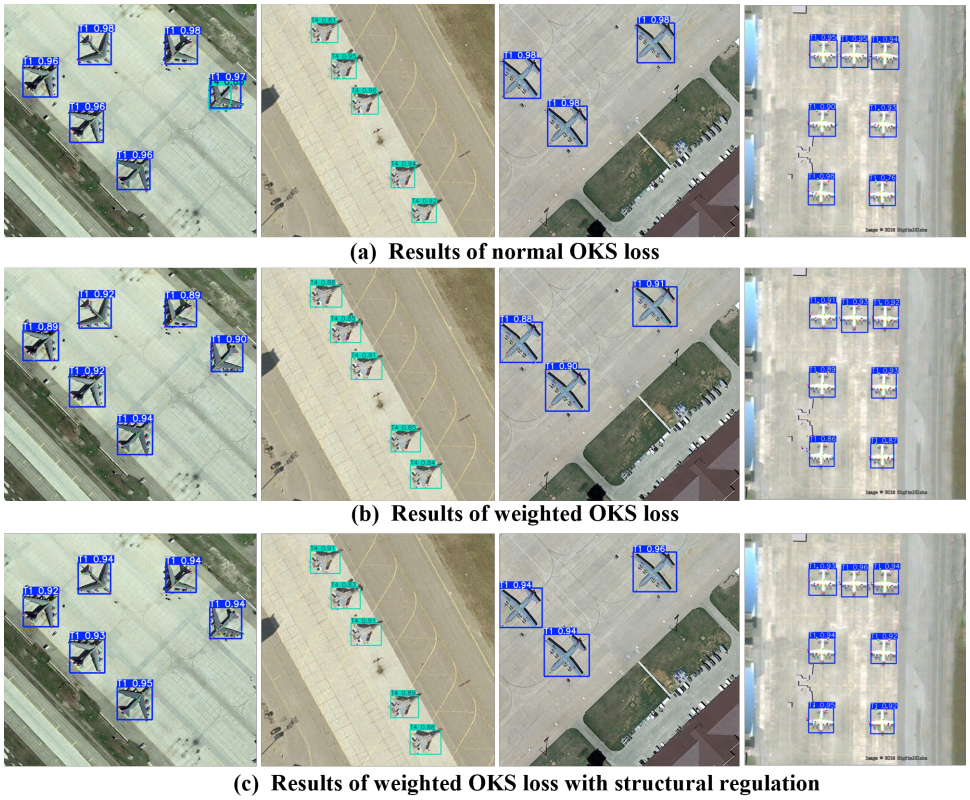


Figure 15. Resultswith different loss function for keypoint detection.

3.2.2. Recognition Percision with Different Matching Methods

Conventional methods based on vector similarity, the proposed matching approach without considering the representative and stable keypoints, and the proposed matching approach based on representative and stable keypoints are compared. Experimental results demonstrate that the proposed method achieves superior classification accuracy under challenging conditions including keypoint detection errors, large viewing angles, and similar target models.

Through variance-based analysis, keypoint stability and representational vectors are selected as items for topological similarity error computation due to their superior stability and discriminative power. This configuration yields an average recognition accuracy improvement of around 5 percentage points compared to alternative matching algorithms, as shown in Figure 16 and Table 10.

Table 10. Overall performance with different matching methods without (w/o) or with (w) stability and discriminative vectors.

	Vector-Based Matching	w/o R and S Vector	w R and S Vector
Percision	92.1%	94.1%	98.9%

Figure 16. Results of vector-based matching (**top row**) and the proposed comprehensive matching algorithm with selected stability and discriminative vectors (**bottom row**).

3.3. Failure Cases

Similar to most of CNN-based and data-driven algorithms, the precision of keypoint detection can be affected by factors such as shadows, atmospheric interference, and partial occlusion of the target.

Analysis on kinds of failure cases caused by partial occlusion of the target are conducted in this section. As shown in Figure 17, benefiting from the proposed structural loss, minor occlusions have minimal impact on the accuracy of keypoint detection. However, the complete absence of an entire component will lead to failures in keypoint extraction. This is partly because the dataset used in this study contains almost no instances where keypoints of the targets are occluded. In such cases, the classification result of the target will be significantly affected. Concurrently, it can be observed that the inference results of the detection and recognition algorithm based on YOLO also vary as the occluded area increases.

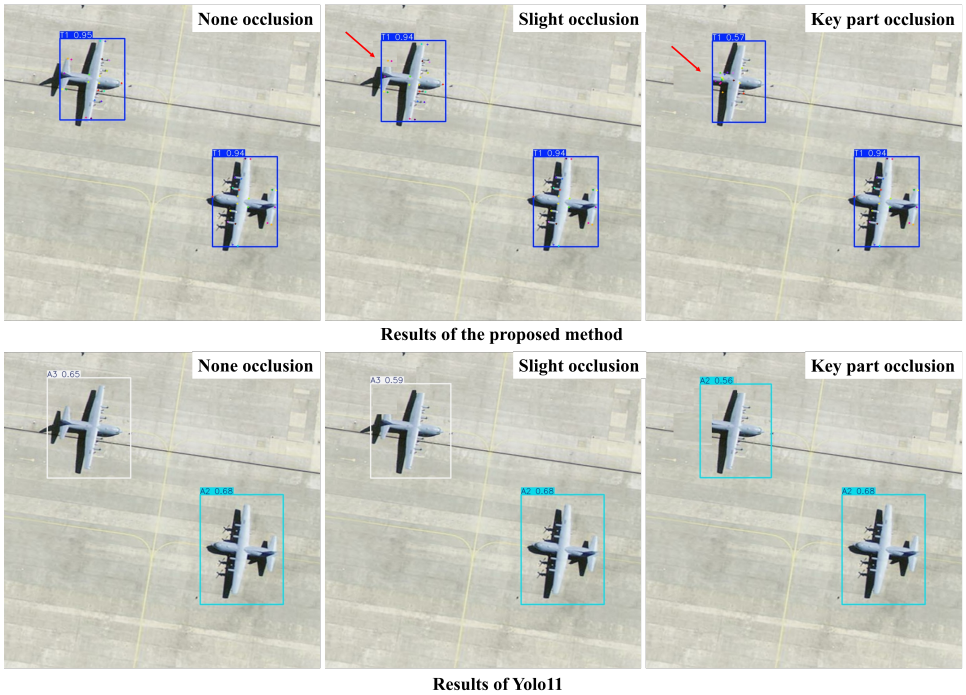


Figure 17. Results of keypoint detection and normal detection algorithm for images with occlusion at different levels.

Overall, the proposed method demonstrates significant robustness against shadows and variations in image quality. However, extensive occlusion may lead to instability in keypoint extraction, which will be the focus of our future work.

4. Discussion

The performance gains of our method originate mainly from three aspects. Target Grouping by Appearance: Objects in the multi-category detection task are consolidated based on visual similarity, treating targets with highly similar appearance as a single class in the detection stage. This strategy significantly improved the precision for categories with few samples, as evidenced by the notable increase in AP for the E-8 aircraft, as shown in Table 7.

Discriminative Feature Extraction via Keypoints: This method of extracting quantitative features based on keypoints for classification effectively addresses the challenge of distinguishing between visually similar aircraft. For instance, the C-17 and C-5 transport aircraft share similar fuselage configurations and paint schemes. Our method accurately extracts keypoints and derives digital characteristics like aspect ratio and fuselage proportions. Since the C-17 and C-5 exhibit distinct differences in fuselage segment proportions, our approach enables accurate differentiation.

Stable and Representative Keypoint Selection: The keypoints utilized were selected through quantitative analysis for high stability and representativeness. The selection fully considered the stability of keypoints across different viewing angles, resulting in a set of keypoints that remain relatively consistent for each target, regardless of perspective. Concurrently, representative analysis was employed to identify the most discriminative keypoints for effectively distinguishing between different categories. This yields a final keypoint set that balances stability and discriminativeness. Generating digital signatures based on this set filters out interference from unstable or ineffective keypoints, precisely captures inter-class differences, and consequently enhances the stability of the keypoint-based matching process.

Building upon the keypoint detection foundation, future work can integrate the powerful reasoning capabilities of large language models to achieve synergistic multimodal processing. The keypoint-based approach enables precise localization of aircraft components and quantitative description of structural proportions, angles, and parameters. These geometric representations can be transformed into natural language descriptors that interact with knowledge graphs of aircraft attributes. Combining this structural understanding with large language model reasoning will advance towards comprehensive aircraft cognitive systems capable of contextual inference.

5. Conclusions

This manuscript proposes a concise method for fine-grained recognition of aircraft targets, based on a constructed aircraft keypoint dataset, and its effectiveness is demonstrated through experiments. This method demonstrates excellent processing capabilities similar to handling open and closed-set data. For aircraft targets not covered in the training set, users can adopt a similar technical approach to establish keypoint templates and datasets tailored to specific scenarios. By further applying the methods proposed in this research for keypoint extraction and aircraft template preparation, it becomes feasible to develop aircraft keypoint detection and recognition algorithms adapted to their requirements.

Author Contributions: Methodology, Q.W.; Software, L.Z. and B.N.; Validation, P.L.; Resources, P.L.; Data curation, Q.W., P.L. and L.Z.; Writing—original draft, Q.W.; Writing—review & editing, B.N.; Supervision, X.G.; Project administration, G.Z. and Y.H.; Funding acquisition, F.W. and G.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the first author, wangqt@aircas.ac.cn. The datasets adopted in this work include MAR20-kp which is based on MAR20 and a dataset built with GF-07 data. The MAR20-KP will be available for every researcher upon request, while, considering the Data Copyright of GF-07 data, the second datasets will be reserved, currently.

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

References

1. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A Large-scale Dataset for Object Detection in Aerial Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018, Salt Lake City, UT, USA, 18–22 June 2018 .
2. Wenqi, Y.U.; Cheng, G.; Wang, M.; Yao, Y.; Han, J. MAR20: A benchmark for military aircraft recognition in remote sensing images. *Natl. Remote Sens. Bull.* **2023**, *27*, 2688–2696.
3. Sun, X.; Wang, P.; Yan, Z.; Xu, F.; Wang, R.; Diao, W.; Chen, J.; Li, J.; Feng, Y.; Xu, T. FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 116–130. [CrossRef]
4. Wei, L.; Dragomir, A.; Dumitru, E.; Christian, S.; Scott, R.; Fu, C.; Berg, A.C. *SSD: Single Shot MultiBox Detector*; Springer: Cham, Switzerland, 2016.
5. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
6. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the Computer Vision and Pattern Recognition 2016, Las Vegas, NV, USA, 27–30 June 2016.
7. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767. [CrossRef]
8. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [CrossRef]
9. Khanam, R.; Hussain, M. YOLOv11: An Overview of the Key Architectural Enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]
10. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2014, Columbus, OH, USA, 24–27 June 2014.
11. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision 2015, Santiago, Chile, 7–13 December 2015.
12. 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]
13. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Houlsby, N. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
14. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Tao, D. A Survey on Visual Transformer. *arXiv* **2020**, arXiv:2012.12556.
15. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Comput. Surv.* **2022**, *54*, 41. [CrossRef]
16. Fang, Y.; Liao, B.; Wang, X.; Fang, J.; Qi, J.; Wu, R.; Niu, J.; Liu, W. You Only Look at One Sequence: Rethinking Transformer in Vision through Object Detection. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 26183–26197.
17. Fu, J.; He, B.; Wang, Z. Aircraft recognition based on feature points and invariant moments. In Proceedings of the 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, China, 29–31 July 2017; pp. 7–12. [CrossRef]
18. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A Unified Embedding for Face Recognition and Clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015, Boston, MA, USA, 7–12 June 2015.
19. Baltrusaitis, T.; Robinson, P.; Morency, L.P. OpenFace: An open source facial behavior analysis toolkit. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision 2016, Lake Placid, NY, USA, 7–9 March 2016.
20. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2014, Columbus, OH, USA, 24–27 June 2014.
21. Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A. VGGFace2: A dataset for recognising faces across pose and age. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, Xi'an, China, 15–19 May 2018.
22. Zhao, A.; Fu, K.; Wang, S.; Zuo, J.; Zhang, Y.; Hu, Y.; Wang, H. Aircraft Recognition Based on Landmark Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1413–1417. [CrossRef]

23. Yuhang, Z.; Hao, S.; Jiawei, Z.; Hongqi, W.; Guangluan, X.; Xian, S. Aircraft Type Recognition in Remote Sensing Images Based on Feature Learning with Conditional Generative Adversarial Networks. *Remote Sens.* **2018**, *10*, 1123.
24. Jia, H.; Guo, Q.; Zhou, R.; Xu, F. Airplane Detection and Recognition Incorporating Target Component Detection. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 12–16 July 2021; pp. 4188–4191. [CrossRef]
25. Liu, S.; Wang, Q.D.; Zhang, L.; Han, X.X.; Wang, B.Q.; Liu, Y.X. Aircraft type recognition method by integrating target segmentation and key points detection. *Natl. Remote Sens. Bull.* **2024**, *28*, 7–12. [CrossRef]
26. Qian, Y.; Pu, X.; Jia, H.; Wang, H.; Xu, F. ARNet: Prior Knowledge Reasoning Network for Aircraft Detection in Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5205214. [CrossRef]
27. Chen, D.; Zhong, Y.; Ma, A.; Zheng, Z.; Zhang, L. Explicable Fine-Grained Aircraft Recognition Via Deep Part Parsing Prior Framework for High-Resolution Remote Sensing Imagery. *IEEE Trans. Cybern.* **2024**, *54*, 3968–3979. [CrossRef]
28. Guo, Z.; Hou, B.; Guo, X.; Wu, Z.; Yang, C.; Ren, B.; Jiao, L. MSRIP-Net: Addressing Interpretability and Accuracy Challenges in Aircraft Fine-Grained Recognition of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5640217. [CrossRef]
29. Xie, X.; Cheng, G.; Rao, C.; Lang, C.; Han, J. Oriented Object Detection via Contextual Dependence Mining and Penalty-Incentive Allocation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5618010. [CrossRef]
30. Zhao, W.X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z.; et al. A Survey of Large Language Models. *arXiv* **2025**, arXiv:2303.18223.
31. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. LLaMA: Open and Efficient Foundation Language Models. *arXiv* **2023**, arXiv:2302.13971. [CrossRef]
32. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q.; Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Adv. Neural Inf. Process. Syst.* **2023**, *35*, 24824–24837.
33. Köstinger, M.; Wohlhart, P.; Roth, P.M.; Bischof, H. Annotated Facial Landmarks in the Wild: A large-scale, real-world database for facial landmark localization. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011.
34. Li, Y.; Yang, S.; Liu, P.; Zhang, S.; Wang, Y.; Wang, Z.; Yang, W.; Xia, S.T. *SimCC: A Simple Coordinate Classification Perspective forHuman Pose Estimation*; Springer: Cham, Switzerland, 2022.
35. Nibali, A.; He, Z.; Morgan, S.; Prendergast, L. Numerical Coordinate Regression with Convolutional Neural Networks. *arXiv* **2018**, arXiv:1801.07372. [CrossRef]
36. Geng, Z.; Sun, K.; Xiao, B.; Zhang, Z.; Wang, J. Bottom-Up Human Pose Estimation Via Disentangled Keypoint Regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021, Nashville, TN, USA, 19–25 June 2021.
37. Zhang, F.; Zhu, X.; Dai, H.; Ye, M.; Zhu, C. Distribution-Aware Coordinate Representation for Human Pose Estimation. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
38. Li, J.; Bian, S.; Zeng, A.; Wang, C.; Pang, B.; Liu, W.; Lu, C. Human Pose Regression with Residual Log-likelihood Estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision 2021, Montreal, BC, Canada, 11–17 October 2021.
39. Maji, D.; Nagori, S.; Mathew, M.; Poddar, D. YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022, New Orleans, LA, USA, 18–24 June 2022.
40. Ding, J.; Xue, N.; Long, Y.; Xia, G.S.; Lu, Q. Learning RoI Transformer for Oriented Object Detection in Aerial Images. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 2844–2853. [CrossRef]
41. Xie, X.; Cheng, G.; Wang, J.; Yao, X.; Han, J. Oriented R-CNN for Object Detection. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 3500–3509. [CrossRef]
42. Song, G.; Liu, Y.; Wang, X. Revisiting the Sibling Head in Object Detector. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11560–11569. [CrossRef]
43. Jiang, X.N.; Niu, X.Q.; Wu, F.L.; Fu, Y.; Bao, H.; Fan, Y.C.; Zhang, Y.; Pei, J.Y. A Fine-Grained Aircraft Target Recognition Algorithm for Remote Sensing Images Based on YOLOV8. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 4060–4073. [CrossRef]