remote sensing

MDPI

Article

Aircraft-LBDet: Multi-Task Aircraft Detection with Landmark and Bounding Box Detection

Yihang Ma [1], Deyun Zhou [1], Yuting He [1], Liangjin Zhao [2,3], Peirui Cheng [2,3], Hao Li [2,3] and Kaiqiang Chen [2,3,*]

[1] School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China; mayhaircas@mail.nwpu.edu.cn (Y.M.); dyzhou@nwpu.edu.cn (D.Z.); heyt@nwpu.edu.cn (Y.H.)
[2] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China; zhaolj004896@aircas.ac.cn (L.Z.); chengpr@aircas.ac.cn (P.C.); lihao@aircas.ac.cn (H.L.)
[3] Key Laboratory of Network Information System Technology (NIST), Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China
* Correspondence: chenkq@aircas.ac.cn; Tel.: +86-188-1090-0679

Abstract: With the rapid development of artificial intelligence and computer vision, deep learning has become widely used for aircraft detection. However, aircraft detection is still a challenging task due to the small target size and dense arrangement of aircraft and the complex backgrounds in remote sensing images. Existing remote sensing aircraft detection methods were mainly designed based on algorithms employed in general object detection methods. However, these methods either tend to ignore the key structure and size information of aircraft targets or have poor detection effects on densely distributed aircraft targets. In this paper, we propose a novel multi-task aircraft detection algorithm. Firstly, a multi-task joint training method is proposed, which provides richer semantic structure features for bounding box localization through landmark detection. Secondly, a multi-task inference algorithm is introduced that utilizes landmarks to provide additional supervision for bounding box NMS (non-maximum suppression) filtering, effectively reducing false positives. Finally, a novel loss function is proposed as a constrained optimization between bounding boxes and landmarks, which further improves aircraft detection accuracy. Experiments on the UCAS-AOD dataset demonstrated the state-of-the-art precision and efficiency of our proposed method compared to existing approaches. Furthermore, our ablation study revealed that the incorporation of our designed modules could significantly enhance network performance.

Keywords: aircraft detection; multi-task learning; landmark detection; bounding box detection

Citation: Ma, Y.; Zhou, D.; He, Y.; Zhao, L.; Cheng, P.; Li, H.; Chen, K. Aircraft-LBDet: Multi-Task Aircraft Detection with Landmark and Bounding Box Detection. *Remote Sens.* **2023**, *15*, 2485. https://doi.org/10.3390/rs15102485

Academic Editor: Naoto Yokoya

Received: 3 April 2023
Revised: 27 April 2023
Accepted: 4 May 2023
Published: 9 May 2023

1. Introduction

Remote sensing has emerged as a powerful tool for gathering data and information about the ground's surface from a long distance. Recent developments in deep learning have revolutionized the field of remote sensing, providing researchers with new and sophisticated methods for image processing and interpretation. Among these methods, classification [1–5], object detection [6–12], change detection [13–16] and semantic segmentation [17–19] are some of the most popular uses for deep learning-based remote sensing, which have significantly accelerated the realization of related applications in the fields of environmental monitoring, traffic security and national defense.

As typical remote sensing targets, aircraft are important transportation carriers and military targets. The accurate detection of aircraft targets plays a vital role in diverse applications, such as air transportation, emergency rescue and military surveillance. Compared to object detection methods based on other specific targets, such as ships and constructions [20,21], it is difficult to detect aircraft in remote sensing images, mainly because of their small size, dense distribution and complex backgrounds. Therefore, aircraft detection is an important task within the field of remote sensing.

Over the past few years, researchers have been striving to develop efficient and accurate methods for aircraft detection, which can be broadly categorized into two groups based on the techniques used: bounding box regression and landmark detection. Bounding box regression methods employ general detection models, such as the R-CNN series [22–24] and the YOLO series [25–27], to extract features by selecting a large number of region proposals for subsequent regression and classification. In contrast, landmark detection-based methods first locate different types of landmarks on objects and then form detection boxes based on these landmarks, which can make better use of the structural characteristics of aircraft.

However, aircraft detection has some characteristics that are different from those of general object detection. Firstly, as shown in Figure 1, aircraft targets in remote sensing images are usually densely arranged in airports and are prone to interference between different targets. Secondly, the head, wings and tail of an aircraft contain strong structural information that is crucial for accurate detection. Finally, the fine-grained classification of aircraft categories [28] relies heavily on the structural information of aircraft. The existing approaches [29–33] do not fully leverage these features, despite some improvements in aircraft detection methods. For example, bounding box regression-based methods suffer from irrelevant background interference within the rectangular anchor boxes, while implicit feature extraction fails to make full use of the strong structural information of aircraft targets. Moreover, in the case of a large number of closely spaced targets, landmark-based methods are prone to errors when grouping keypoints. Therefore, it is necessary to develop new and more effective methods for aircraft detection that can overcome these challenges and take full advantage of the unique characteristics of aircraft targets.

Figure 1. The challenges of aircraft detection, including dense arrangement, different aircraft scales, different structure information and complex background interference. Existing anchor-based methods do not fully utilize aircraft scale and structural information, while keypoint methods have difficulty clustering and grouping densely arranged objects.

To address the characteristics of aircraft detection and the problems of existing algorithms, we propose an aircraft detection algorithm called Aircraft-LBDet (aircraft landmark and bounding box detection), which is a multi-task algorithm model combining bounding box detection and landmark detection. The bounding box-based approach can effectively compensate for the grouping errors of landmark detection in the case of dense arrangement, thus improving the accuracy of detection. By adding landmark supervision information to the bounding box-based approach, the structural features of aircraft can be fully utilized to

provide more detailed position information for bounding box localization, which is helpful for subsequent fine-grained classification tasks. We can achieve more accurate aircraft detection by combining bounding box detection with landmark supervision. In this paper, we present our method and evaluate its performance on relevant remote sensing datasets. The experimental results demonstrated that the proposed method outperformed existing state-of-the-art algorithms and was robust to various environmental conditions.

The main contributions of our method are as follows:

- We propose a multi-task joint training method for remote sensing aircraft detection, within which landmark detection provides stronger semantic structural features for bounding box localization in dense areas, which helps to improve the accuracy of aircraft detection and recognition;
- We propose a multi-task joint inference algorithm, within which landmarks provide more accurate supervision for the NMS filtering of bounding boxes, thus substantially reducing post-processing complexity and effectively reducing false positives;
- We optimize the landmark loss function for more effective multi-task learning, thereby further improving the accuracy of aircraft detection.

In the rest of this paper, we first review the related work in the field of remote sensing and identify the gaps that our proposed method aims to address. Next, we describe our proposed method and explain how it differs from existing approaches. Then, we present the results of our experiments, including comparisons to existing methods and ablation studies. Finally, we summarize our findings and discuss the implications of our work for future research within the remote sensing field.

2. Related Work

With the rapid development of deep neural networks, breakthroughs have been made in text, image and speech processing [34–37]. As a crucial task in the field of computer vision, object detection has attracted a lot of research attention. In this section, we provide a brief overview of existing deep learning-based object detection algorithms for general methods and remote sensing methods.

2.1. General Object Detection Methods

Object detection has traditionally involved utilizing feature matching templates and sliding windows for detection. In recent years, researchers have proposed numerous anchor-based detection methods based on deep CNNs. These methods are typically categorized as either two-stage or one-stage methods. In two-stage detectors, anchors are generated and then regions of interest are subsequently utilized for classification and regression. One classic two-stage detector, the Fast-RCNN [24], uses RoI pooling to enhance the semantic descriptions of features and improve efficiency. Building on this, the Faster-RCNN [22] method improves this approach further by introducing a region proposal network to extract regions and enable end-to-end training. However, the region proposal stage requires many parameters and computational costs. The Cascade R-CNN [38] introduces a multi-stage perception mechanism that reduces the proportion of false positive samples. While two-stage methods have high detection accuracy, they require a significant number of parameters and computational costs. In contrast, one-stage methods offer faster inference speeds, making them ideal for real-time object detection in resource-constrained environments. One-stage methods have been reported to reach inference efficiency. YOLOv1 [25] performs regression and classification directly with a grid for images instead of generating anchors. Subsequently, other methods, such as YOLOv2 [26], YOLOv3 [27] and SSD [39], use anchors with multi-stage feature maps to improve detection accuracy.

Although one-stage methods are more efficient, their detection accuracy is a big problem compared to that of two-stage methods. Therefore, more precise anchor-based one-stage methods have been explored and the keypoint-based method has become an essential direction. By utilizing a single convolutional neural network, CornerNet [40] is able to detect bounding boxes that are defined by keypoints situated at the top left and bottom

right corners, transforming the detection problem into keypoint prediction and clustering. Duan et al. [41] found that CornerNet can only extract the edge features of objects, so they proposed CenterNet and introduced keypoint triplets to determine objects, allowing the network to obtain the internal features of objects and distinguish whether each bounding box is correct. To address the problem that features extracted from the regular cells of bounding boxes are easily affected by invalid features in background and foreground areas, RepPoints [42] uses deformable convolutional kernels, which can adaptively extract object features and contain the semantic information of multi-stage levels, thereby solving the problem of the limited feature extraction capability of rectangular boxes. The creators of CentripetalNet [43] found that relying solely on appearance-based embeddings to group keypoints had significant limitations, so they proposed a new corner-matching method based on CornerNet, which improves robustness by learning additional centripetal offsets.

2.2. Object Detection in Remote Sensing Images

There are many challenges in the field of object detection in remote sensing images compared to object detection in standard images, such as extensive modifications, intricate surroundings and compact objects at multiple scales. Li et al. [44] addressed the the issue of rotation by designing multi-angle anchors and proposing a double-channel feature fusion network to enhance the joint representation of ambiguous features. Fu et al. [45] built a rotation-aware detector based on the Faster R-CNN to cover objects in arbitrary directions. These methods offer feature fusion frameworks to produce and fuse hierarchical features at multiple scales. To address the challenge of detecting multi-scale objects in aerial imagery, Qian et al. [46] proposed a method called multi-level feature fusion (MLFF), which combines the multi-scale features output by FPNs. Yao et al. [47] designed a unified EssNet backbone to preserve the resolution of deep features by using dilated convolution. The approach involves generating high-quality feature maps that enable the detection of objects at varying scales. Liu et al. [48] designed a feature pyramid model, which can be used to combine multi-scale features through selective refinement modules between different spaces and channels. Then, rich semantic information can be added to multi-scale object detection using a context enhancement module. Taking into account the object distribution patterns observed in selected datasets, Ye et al. [49] applied a stitcher to make one image comprising objects of diverse scales, effectively balancing the proportions of targets.

The advancement of deep convolutional neural networks has led to notable advancements in the detection of aircraft in remote sensing images. Yu et al. [29] proposed an aircraft detection network in a remote sensing GLF-Net utilizing the encoder and decoder fusion of multi-scale features with both global and local information, which addresses the characteristics of small targets and the complex backgrounds in aircraft samples. To address the problem of interference by clouds, Zhang et al. [30] proposed a novelty feature aggregation network called AFA-Net. The method includes a self-attention module that dynamically emphasizes the local features of the exposed sections of aircraft, across both the feature map's channel and spatial dimensions. X-LineNet [31] is an aircraft detector that utilizes local object features based on lines. This method transforms the aircraft detection task into the prediction and clustering of pairs of intersecting lines within objects, making clustering easier by enhancing the dimensionality of point-to-line. Considering the geometric semantics of the sword-shaped elements of aircraft structures, S2CGNet [32] transforms the task of detecting aircraft into the estimation of sparse instance-level masks. It introduces SAMs (sword attenuation masks) to capture the geometric appearance of aircraft, which enriches local appearance features and improves the accuracy of the bounding boxes. Zhao et al. [33] introduced a module for fusing multi-scale features called BFPCAR, which incorporates semantic features that are prioritized during the fusion of information to reduce information loss between different layers and overcome the issue of imbalanced attention between non-adjacent layers. Liu et al. [50] proposed an aircraft detection CNN with a corner cluster algorithm. The method begins by detecting the corners of binary images using mean-shift, which generates potential regions of interest. Subsequently, a CNN

layer is employed to extract features and classify the regions that could potentially include aircraft. To further confirm the locations of any aircraft, an extra screening process is then conducted. The DPANet [51] combines deconvolution operations with a position attention method for two-stage aircraft detection, which aims to capture external structural features and enhance the network's capability to differentiate between aircraft and the background.

Although there are various kinds of algorithms concerning the aircraft detection task, the realization of multi-stage feature extraction and the understanding of aircraft-related information remain important problems that need to be researched.

3. Proposed Method

3.1. Overview

The suggested method employs YOLO [52] for enhanced efficiency and reduced computational resource usage. As illustrated in Figure 2, the method consists of three components: a feature extraction backbone, a multi-scale feature pyramid module and an object and landmark detection head. The feature extraction backbone utilizes a residual structure, with a deeper network layer designed to facilitate in-depth feature extraction from input remote sensing images. The multi-scale feature pyramid module is crucial to the overall method as it combines the features obtained from the backbone network, thereby enriching the diversity of the learned features and boosting the network's detection performance. It was mainly designed by SPP (spatial pyramid pooling) [53], with small convolution kernels and parallel modes for aircraft detection tasks. The object and landmark detection head is the module that is used to locate aircraft precisely and regress any landmarks, which was mainly designed by decoupling to resolve conflicts between tasks.

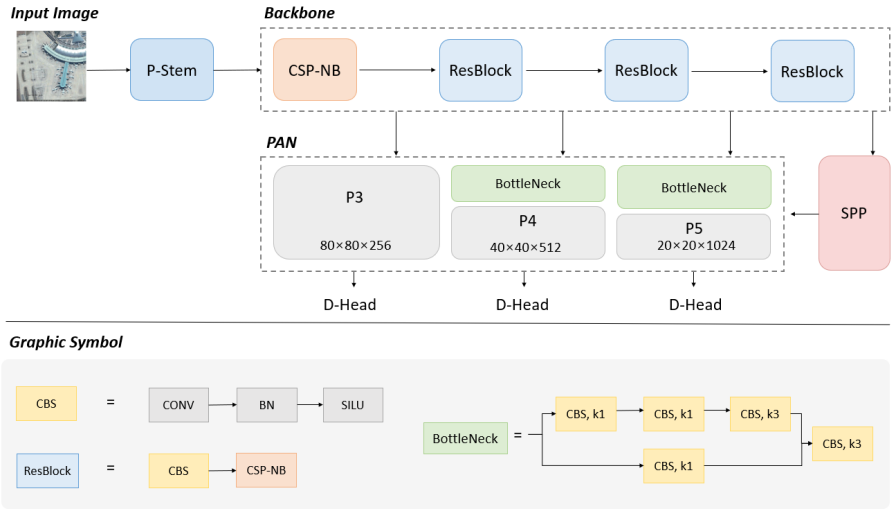

Figure 2. The architecture of Aircraft-LBDet, consisting of a P-stem module, a backbone, an SPP, a PAN module and a D-Head module. Some modules are described in more detail below. In the bottleneck module, k1 and k3 mean the 1×1 and 3×3 kernel size of convolution.

3.2. Feature Extraction Backbone

The feature extraction backbone contains a stem, CSPNet and ResBlock. Stem is usually the initial module in whole networks, which provides better information extraction in deeper dimensions. In YOLOv5, the initial stem layer employs a slicing operation to divide high spatial resolution feature maps into several low spatial resolution maps. This way, the information attenuation caused by downsampling can be reduced. As depicted in Figure 3, the P-Stem module supersedes the original Focus layer, effectively reducing computational demands while essentially maintaining feature extraction capability.

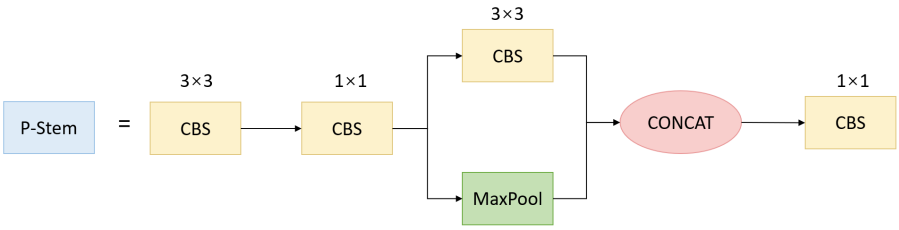

Figure 3. P-Stem was designed to be a lightweight module with 1×1 convolution.

CSPNet is reported to mitigate the issue of redundant gradient information, which leads to inefficient optimization and expensive inference computations. In order to streamline the architecture, decrease the number of parameters and more effectively leverage the fuse operation during inference, we redesigned the CSP module and named it CSP-NB. In Figure 4, the CSP-NB module is shown. This module was inspired by the DenseNet [54] and PeleeNet [55], which have dynamic numbers of channels. The CSP-NB module divides the original input into two branches and then uses N neckblock residual convolution operations after CBS operations to deepen the network channel dimensions and improve feature extraction. After merging features, CBS operation is performed. The CSP-NB module is applied to feature extraction and the backpropagation effect between layers is increased through the nested residual structure to avoid the disappearance of the gradient caused by the deepening of the network layer, which helps to obtain deeper features without degrading the module.

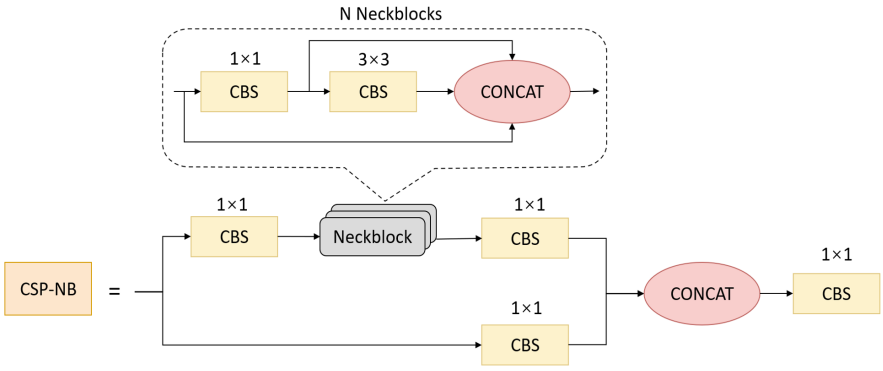

Figure 4. The CSP-NB module with N neckblocks.

3.3. Multi-Scale Feature Pyramid Module

The multi-scale feature pyramid module contains spatial pyramid pooling (SPP) [53] in Figure 5 and a prototype alignment network (PAN) [56]. The SPP module extracts multi-scale features, which are advantageous when dealing with different target sizes within images to be detected. The core lies in the consistent adjustment of the spatial dimensions of feature maps when passing through the multi-scale MaxPool layer, which is convenient for subsequent feature stacking and merging. The three kernel sizes were revised to $7 * 7$, $5 * 5$ and $3 * 3$ in our method for small aircraft targets. In addition, we redesigned SPP as an equivalent serial mode using three $3 * 3$-kernel MaxPool operations to accelerate the running speed of the module. Input images undergo parallel processing through multiple MaxPool operations of varying sizes, followed by further fusion, which can address the multi-scale issue of targets to some degree.

The PAN architecture involves two main components: a prototype network and an alignment network. The prototype network generates a set of class prototypes for each type, which represent the common features of each class. The prototype generation module takes the feature characteristics generated by the backbone and produces a series of class prototypes for each class. Each prototype is a weighted average of the support images of the

feature vector belonging to the same class. The alignment module aligns the prototypes to query images by computing the similarity degree between the features at each pixel location. The PAN can further fuse multi-scale features and realize the alignment of multi-scale feature maps so as to provide features with richer information for subsequent detection.

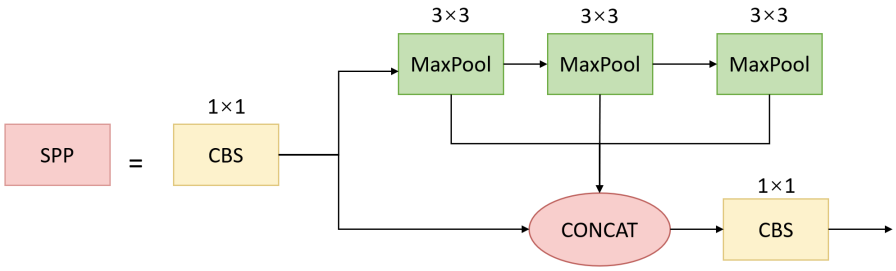

Figure 5. The SPP module with small kernels and parallel mode.

3.4. Object and Landmark Detection Head

General object detection heads do not include landmark detection; however, in the context of aircraft detection tasks, detecting landmarks is a crucial component for accurately identifying aircraft. Therefore, we recommend a brand-new approach for improving overall mission accuracy by adding landmark detection as a parallel task.

To achieve this, we decouple the head into a multi-task detection module where the network can simultaneously output the target bounding box coordinates, classification confidence for prediction and landmark coordinates. However, it is worth noting that for object detection tasks in the remote sensing field, there is the well-known problem of conflict between classification, regression and landmark detection tasks [57]. This conflict arises due to misalignment between different tasks in the spatial dimension, which greatly limits overall performance and results in a highly constrained trade-off. To overcome this limitation, we adopt a $1 * 1$ convolution layer for each level of the PAN feature, which reduces the channel number of the feature layer to 256. Additionally, we introduce three parallel branches, each with two $3 * 3$ convolution layers, to handle the regression, classification and landmark detection tasks. These changes improve the output labels for the head, as shown in Figure 6.

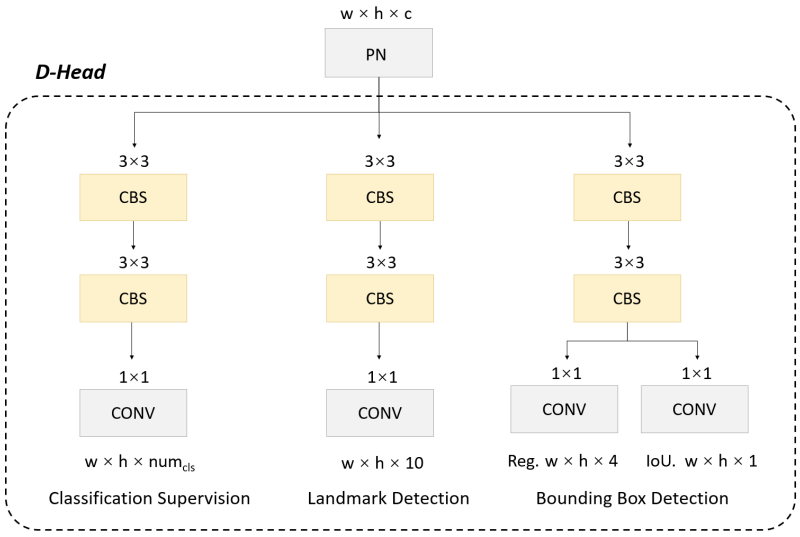

Figure 6. The D-Head module, with three parallel branches each for classification regression and landmark detection tasks. PN is the output of the PAN structure.

3.5. Central-Constraint NMS

The NMS operation is a common component of the post-processing stage of object detection. After an input image is predicted by a network, the network will predict a large number of bounding boxes. Then, the appropriate bounding boxes need to be reserved so that the object confidence of each box is at its highest and the same object will not be detected repeatedly. For aircraft detection tasks, when the range of a bounding box is out of accuracy, especially for large-scale tasks, the images include redundant background and parts of surrounding targets, which makes it difficult for the basic NMS operation to completely eliminate redundant boxes. However, in terms of aircraft detection, the landmark points of aircraft have relatively fixed spatial geometric relationships with each other. Hence, we can use aircraft landmark points to further constrain the selection of bounding boxes for more accurate output results. In this module, we further strengthen the constraints during the NMS operation using the landmark detection results output by the decoupled detection head, which aims to filter out highly redundant candidate boxes. In detail, there should be only five landmark points in each candidate box and the center coordinates of the five landmark points should coincide with the center coordinates of the box as much as possible. Therefore, we propose an additional landmark point confidence $conf_{lms}$ to update the original object confidence $conf_{obj}$. Supposing the geometric center coordinate for a candidate bounding box is b_c, where the geometric center coordination of the five landmark points is p_c, the update formula is:

$$conf_{lms} = max\{0, 1 - \frac{dist(p_c - b_c)}{min(w,h)/2}\}, \tag{1}$$

$$conf_{obj} = conf_{lms} \cdot ori\_conf_{obj}, \tag{2}$$

where $dist(\cdot)$ is the L2 distance between two points (i.e., the Euclidean distance) and w and h are the width and height of the candidate box, respectively. Then, the updated object confidence can be used to remove redundant boxes during NMS processes. For clarity, the process of Central-Constraint NMS is shown in Algorithm 1.

Algorithm 1 Central-constraint non-maximum suppression (central-constraint NMS) algorithm.

Inputs: $A = \{a_1, \cdots, a_N\}, P = \{p_1, \cdots, p_N\}, t$; A represents the list of initial detection boxes; P contains the corresponding detection scores; t denotes the NMS threshold.

Output: R, P

$R \leftarrow \{\}$
while $A \neq$ empty **do:**
  for a_i, p_i in A, P **do:**
    $p_i \leftarrow$ **ScoreUpdate**(p_i, A_i)
  end
  $n \leftarrow arg\max P$
  $N \leftarrow a_n$
  $R \leftarrow R \bigcup N; A \leftarrow A - N$
  for a_i in A **do:**
    if $IoU(N, a_i) \geq t$ **then:**
      $p_i \leftarrow p_i \cdot (1 - IoU(N, a_i))$
    end
  end
end
return R, P

3.6. Landmark Box Loss Function

Due to the small pixels of aircraft target samples, the L2 loss function has a poor fitting effect. When the input increases, the derivation of the L2 loss concerning the input also increases, which causes the gradient to become unstable. Therefore, we adopt a piecewise loss function method to smooth the gradient loss during model training.

$$Func(y) = \begin{cases} f \cdot \ln\left(1 + \frac{|y|}{d}\right), \text{if y < f} \\ |y| - B, otherwise \end{cases} \tag{3}$$

The non-negative f has a nonlinear range of $[-f, f]$, d limits the curvature of the nonlinear region's scope and B is a constant that links the piecewise-defined linear and nonlinear parts in a smooth way.

$$loss_K = \sum_i Wing(s - s'), \tag{4}$$

where s is the predicted landmark and s' is the ground truth.

Furthermore, for the sake of simultaneously constraining the relationships between landmarks and anchor boxes to make them pair, the constraint loss was designed as follows:

$$loss_{relation} = \sum_i L_2(p_c - b_c), \tag{5}$$

where p_c denotes the geometric center of the landmark group and b_c refers to the geometric center of the rectangular regression box.

The $loss_{pull}$ and $loss_{push}$ refer to the method of pairing constraints between points in CornerNet [40] and extend to five points, as follows:

$$loss_{pull} = \frac{1}{N} \sum_{k=1}^{N} [(e_{tlk} - e_k)^2 + (e_{blk} - e_k)^2 + (e_{trk} - e_k)^2 + (e_{brk} - e_k)^2], \tag{6}$$

$$loss_{push} = \frac{1}{N(N-1)} \sum_{k=1}^{N} \sum_{j=1, j \neq k}^{N} \max(0, 1 - e_k - e_j), \tag{7}$$

where e_k is the mean of e_{tlk}, e_{blk}, e_{trk} and e_{brk}. Through the $loss_{pull}$ and $loss_{push}$, the nose, tail and wings of each aircraft target can be effectively clustered. On the basis of the loss between landmarks and bounding boxes, the point and the box of the same aircraft target can be effectively paired and the two-way regression of multi-task aircraft detection can be effectively realized. The new overall landmark box loss function was designed as follows:

$$L_{all} = \alpha \cdot loss_K + \beta \cdot loss_{relation} + \gamma(loss_{pull} + loss_{push}), \tag{8}$$

where α, β and γ are the hyperparameters, which can be adjusted based on the dataset distribution during training.

4. Results

In this section, we describe the experiments related to our proposed approach. We start by describing the dataset used for our experiments, followed by details of the implementation. Then, we discuss the evaluation metrics that were employed to measure the performance of our method. Subsequently, we compare our method to other state-of-the-art techniques on the same dataset. Furthermore, we conduct ablation experiments to analyze the contributions of the individual components of our approach. Finally, we provide visualizations of our results and discuss our findings regarding the fine-grained performance.

4.1. Dataset

During the experimental stage, we chose the UCAS-AOD dataset [58] to validate the performance of our model. The UCAS-AOD is a dataset comprising aerial images for object detection in remote sensing applications, which was developed by the University of the Chinese Academy of Sciences (UCAS) and contains high-resolution images captured by an unmanned aerial vehicle (UAV) over a university campus. The images have a resolution of 0.1 m per pixel and cover an area of approximately 2.4 square kilometers.

The UCAS-AOD dataset is annotated with ground truth object bounding boxes for three object classes: cars, buildings and trees. The annotations were performed manually by experts in remote sensing and computer vision. The dataset has been extensively utilized to demonstrate the effectiveness of object detection algorithms for aerial images, particularly those relying on deep CNNs. There are 648 images for training, including 4720 instances, and 432 images for testing, including 3490 instances. During the experiments, the original image input size was cropped to 832 $\times$ 832.

4.2. Implementation Details

We conducted experiments utilizing the PyTorch 1.8.1 framework in conjunction with the Nvidia Geforce GTX 1080 GPU. We employed the YOLOv5-4.0 as our initial version and implemented modifications that were specific to our research. Our experiments utilized a batch size of 16 and training was executed for a duration of 200 epochs using an SGD optimizer. We set the initial learning rate to 10^{-2} and the final learning rate to 10^{-5} in order to effectively optimize the model performance over time. Furthermore, a momentum value of 0.8 was utilized for the first five warm-up epochs, subsequently transitioning to 0.9 for later steps to enhance the stability and computational efficiency of the training process. In an effort to increase the resilience of our proposed model and improve its generalizability, we also enlarged the dataset through a range of techniques, including flipping, rotation and random cropping.

4.3. Comparison Experiments

We set up comparative experiments in several areas to demonstrate the validity of our proposed model.

Comparison of our model to other generic detection models.

The proposed Aircraft-LBDet method was evaluated in comparison to several advanced object detection methods on the UCAS-AOD dataset. The results are presented in Table 1. The proposed method achieved an impressive AP of 0.904, outperforming the other methods by significant margins. Specifically, it outperformed the Faster R-CNN, SSD, CornerNet, Yolo v3, RetinaNet+FPN and Yolo v5s by 4.5%, 0.8%, 13.9%, 4%, 0.3% and 4.5%, respectively. In addition to its superior performance, the proposed method was also highly efficient. The Yolo v5s, which is known for its efficient and real-time detection, achieved a frame rate of 80.6 FPS; however, the Aircraft-LBDet method further improved the efficiency of the YOLO method, achieving a frame rate of 94.3 FPS, which was 17.0% higher than that of the Yolo v5s. Overall, these results demonstrated that the proposed Aircraft-LBDet method is a highly effective and efficient method for object detection, particularly for aircraft detection.

Table 1. The comparisonbetween the different methods on the UCAS-AOD dataset.

Method	AP	FPS	Model Size
Faster R-CNN [22]	0.859	11	243.5 MB
SSD [39]	0.896	17	144.2 MB
CornerNet [40]	0.765	6.9	804.9 MB
Yolo v3 [27]	0.864	25	248.1 MB
RetinaNet+FPN [59]	0.901	7.2	228.4 MB
Yolo v5s	0.859	80.6	14.17 MB
Ours	**0.904**	**94.3**	**13.8** MB

Comprehensive comparison of multiple indicators between our approach and the baseline.

In Table 2, a comparison between the Aircraft-LBDet method and the baseline YOLOv5s method is presented. The performance of these two methods was evaluated from several aspects. In terms of the average aircraft detection precision, Aircraft-LBDet performed slightly better than the baseline when the threshold was adjusted from 0.5 to 0.95. Additionally, the parameter of Aircraft-LBDet was significantly less than that of YOLOv5s. The false alarm rate (FA) is an important evaluation metric and it is evident from the table that Aircraft-LBDet had a 39.7% lower false alarm rate than the baseline method. The F1 score, which is a measure of the balance between precision and recall, was also significantly higher for Aircraft-LBDet than the baseline. Therefore, the results indicated that Aircraft-LBDet outperforms the baseline in terms of both accuracy and efficiency.

Table 2. The comparison to the baseline.

Method	$AP_{0.5-0.95}$	Flops (G)	FA	F1
Yolo v5s	0.667	26.3	0.121	0.928
Ours	**0.675**	**15.3**	**0.073**	**0.956**

Comparison of other aircraft detection methods.

In the field of aircraft detection, there have been numerous studies exploring effective models to enhance the accuracy of detection [60–64]. We conducted comparative experiments between our model and other models for aircraft detection. The results are summarized in Table 3. It is evident that our approach outperformed all of the other methods, including FR-O [22], ROI-trans [60], FPN-CSL [61], R_3Det-DCL [62], P-RSDet [63] and DARDet [64], which had AP scores ranging from 0.834 to 0.903. Our method achieved an AP score of 0.904, which was 0.001 higher than the best performing method in the literature. Moreover, our approach utilized the smallest backbone (CSP-ResBlock) compared to the other methods, indicating the superiority of our architecture in terms of computational efficiency. The reason for this superior performance was the re-designed CSP-ResBlock architecture in our method. The CSP-ResBlock architecture effectively balances the accuracy and efficiency of the model. It combines the advantages of a residual block and a cross-stage partial (CSP) structure, which allows the model to capture more useful features from input images. Our method outperformed the other SOTA methods and had the smallest backbone (CSP-ResBlock) compared to ResNet-50 and ResNet-101, which indicated the effectiveness of our re-designed CSP-ResBlock architecture and innovative collaborative learning detection head.

Table 3. The comparison of other methods for aircraft detection. All methods were evaluated on the UCAS-AOD dataset.

Method	Backbone	AP
FR-O [22]	ResNet-101	0.834
ROI-trans [60]	ResNet-101	0.889
FPN-CSL [61]	ResNet-101	0.892
R_3Det-DCL [62]	ResNet-101	0.893
P-RSDet [63]	ResNet-101	0.900
DARDet [64]	ResNet-50	0.903
Ours	CSP-ResBlock	**0.904**

4.4. Ablation Experiments

The detection of aircraft in remote sensing images is a challenging task that requires the accurate localization of aircraft landmarks and bounding boxes. With respect to this matter, we propose a novel method for detecting aircraft in remote sensing images by integrating

various modules and loss functions. To evaluate the effectiveness of our proposed method, we conducted a series of ablation experiments on the UCAS-AOD dataset.

The ablation experiments comprised a quantitative analysis of the different modules and loss functions to determine their contributions to the overall performance of our proposed method. The results demonstrated that the landmark box loss function enhanced the detection performance of the algorithm by constraining the relationships between landmarks and bounding boxes. Consequently, the algorithm's detection performance was boosted by 4.9%.

Building on this initial improvement, we specifically added the CSP-NB module to improve the network's feature extraction ability by deepening the network channel, which improved the detection performance of the algorithm by 4.2%. We then incorporated a P-Stem module to improve the average precision (AP) of the algorithm by 1.6%. Finally, we added a central-constraint NMS operation to further increase the detection precision of the algorithm by 0.2%. The ablation study results are listed in Table 4, where each row corresponds to a different combination of modules and loss functions.

Overall, the ablation experiments indicated the effectiveness of our proposed method for detecting aircraft. By integrating different modules and loss functions, we were able to achieve a significant improvement in the detection performance of the algorithm. Our proposed method could be used for various applications, including aerial surveillance, border patrol and disaster response. Further research should explore the possibility of our method being used in other domains and applications.

Table 4. The ablation studies of different modules and loss functions in our proposed model.

ID	a	b	c	d	e
Landmark Box Loss		✓	✓	✓	✓
CSP-NB			✓	✓	✓
P-Stem				✓	✓
Central-Constraint NMS					✓
AP	0.795	0.844	0.886	0.902	0.904
Comparison	-	0.049 ↑	0.042 ↑	0.016 ↑	0.002 ↑

4.5. Visualization

This paper proposes an innovative method for the precise detection of dense objects in large-scale remote sensing imagery. In particular, the proposed method leverages highlight detection and keypoint detection to realize dense object detection and fine-grained classification. In this section, we present some example predictions (Figures 7 and 8) to demonstrate the effectiveness and superiority of our method.

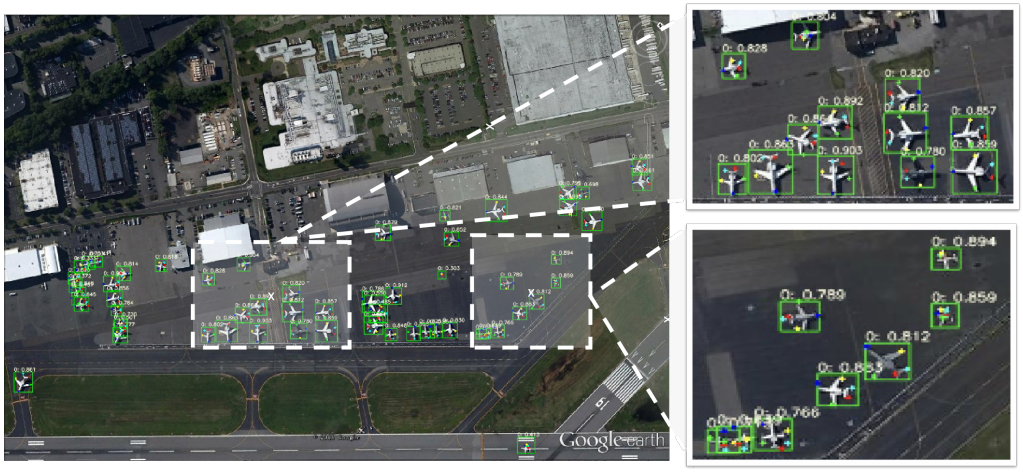

Figure 7. *Cont.*

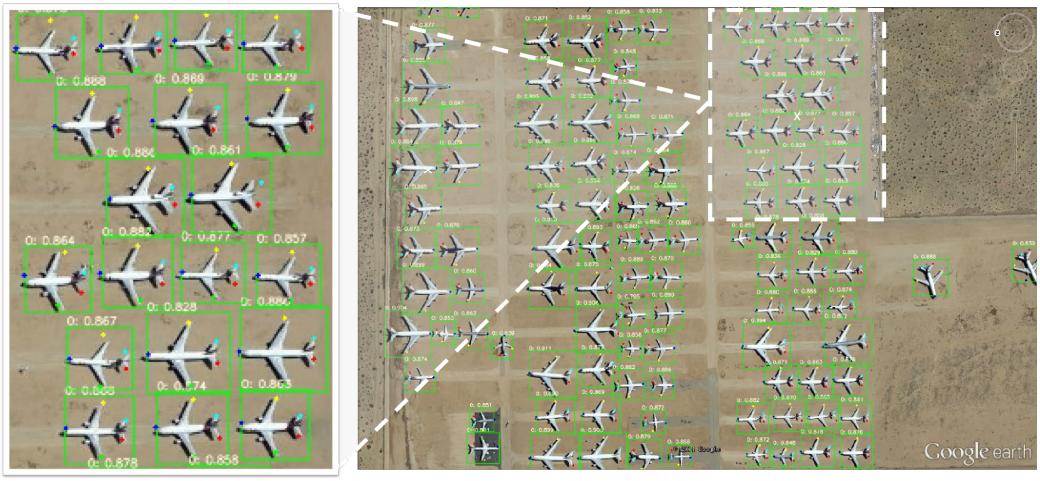

Figure 7. A visualization of the detection results for images with densely arranged targets, showing that our proposed approach had good robustness with different backgrounds and obtained accurate landmarks.

Figure 8. A more detailed visualization of the detection results at a higher resolution, showing that our method maintained its good performance with different backgrounds and obtained accurate landmarks.

Figures 7 and 8 show two densely arranged scenes captured by large-size remote sensing imagery at different resolutions. From these figures, we can see that our proposed

method could accurately detect aircraft and precisely identify their positions to achieve dense object detection. The results showed that our method is capable of handling complex and cluttered environments and could be a promising approach for real-world applications.

Overall, the proposed method achieved a significant improvement over existing methods for the detection of dense objects in large-scale remote sensing imagery. By leveraging highlight detection and keypoint detection, our approach offers a more accurate and precise solution for a wide range of real-world scenarios.

5. Discussion

In Figure 9, we demonstrate that our proposed approach can judge specific types of aircraft, including the MD-90, A330, Boeing787, Boeing777, ARJ21 and Boeing747, by calculating the size of detected keypoints. Different points represent different parts of aircraft. The purple point represents the head, while the green and yellow points represent the left and right wings, respectively. The red and blue points represent the left and right of the tail, respectively. This fine-grained classification is achieved by utilizing highlight-detected keypoints, which enables the proposed method to accurately distinguish between different types of aircraft.

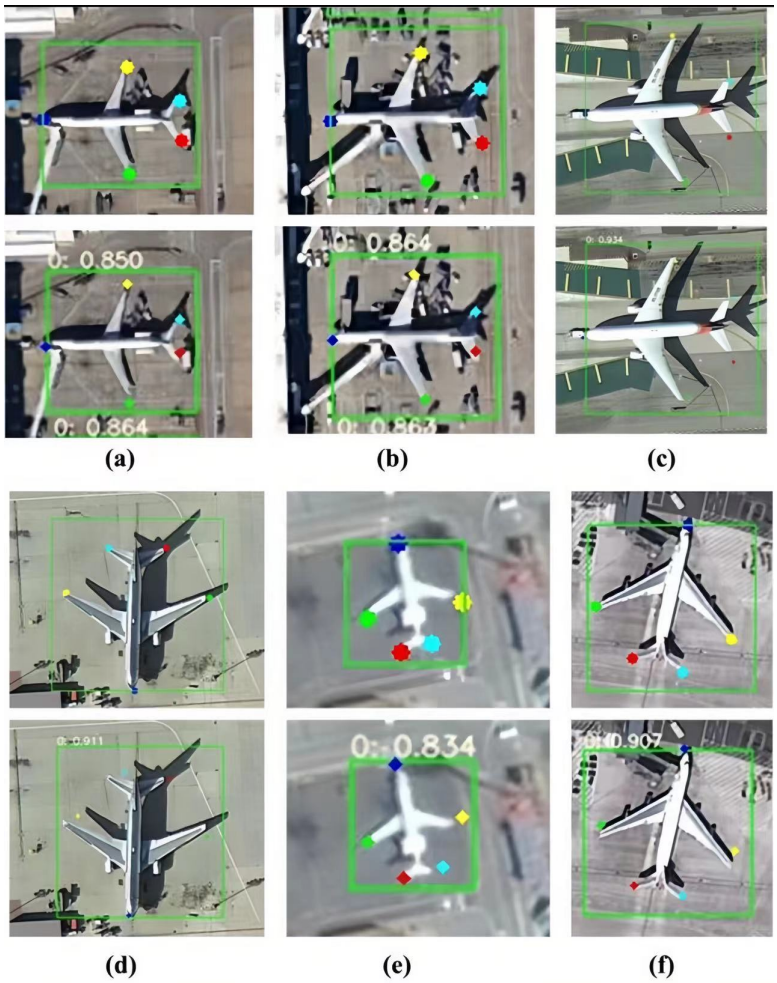

Figure 9. Avisualization of the detection results between the ground truth (top) and the prediction (bottom): (**a**) MD-90; (**b**) A330; (**c**) Boeing787; (**d**) Boeing777; (**e**) ARJ21; (**f**) Boeing747.

Furthermore, we calculated the wingspan and fuselage length of each aircraft category and compared them to relevant data on Wiki. The results of this specific contrast are shown in Table 5. The contrast between the actual and theoretical data for the aircraft categories was conducted to aid the fine-grained classification. The results suggested that the actual

data tended to exhibit slightly increased wingspan and fuselage lengths, ranging from 5–15%, which was because of the angle and inclination of the samples during measurement. Our analysis of the relevant data within each category consistently indicated a close resemblance, which showed that length is a reliable parameter for fine-grained classification. Therefore, the focus on wingspan and fuselage length provides a more comprehensive understanding of aircraft categories and better fine-grained classification performance.

Table 5. The comparison of the wingspans and fuselage lengths (m) of different aircraft.

Aircraft	Theoretical		Actual	
	Wingspan	Fuselage Length	Wingspan	Fuselage Length
MD-90	32.9	39.5	35.4 (7.6% ↑)	42.8 (8.4% ↑)
A330	60.3	58.8	64.1 (6.3% ↑)	62.9 (7.0% ↑)
Boeing787	60.1	57.7	68.0 (13.1% ↑)	60.5 (4.9% ↑)
Boeing777	64.8	63.7	71.4 (10.2% ↑)	65.4 (2.7% ↑)
ARJ21	22.5	33.5	25.8 (14.7% ↑)	36.7 (9.6% ↑)
Boeing747	68.5	70.6	69.3 (1.2% ↑)	73.2 (3.7% ↑)

6. Conclusions

In this paper, we proposed an end-to-end multi-task aircraft detection method using landmark box detection in remote sensing images. The proposed method enables the more accurate detection of closely spaced aircraft targets by combining bounding boxes and landmark detectors, thereby enhancing the reliability and efficacy of remote sensing aircraft detection. Meanwhile, the designed multi-task joint training and inference algorithm proves the satisfactory practical applicability of our model. Our work validates the importance of considering aircraft structures and combining different supervisory information for remote sensing aircraft detection. In future work, we aim to incorporate special component semantic information to achieve more efficient fine-grained classification and guarantee the sufficient perception of visual aircraft features. We believe that future research will be inspired by our work in this area and that this work could contribute to the development of advanced remote sensing technologies for aircraft detection.

Author Contributions: Y.M. conceived and designed the experiments; D.Z. and Y.H. performed the experiments and analyzed the data; Y.M. wrote the paper; L.Z., P.C. and H.L. contributed materials; K.C. supervised the study and reviewed this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable

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

References

1. Li, Y.; Zhang, H.; Xue, X.; Jiang, Y.; Shen, Q. Deep learning for remote sensing image classification: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1264. [CrossRef]
2. Chen, L.; Li, S.; Bai, Q.; Yang, J.; Jiang, S.; Miao, Y. Review of image classification algorithms based on convolutional neural networks. *Remote Sens.* **2021**, *13*, 4712. [CrossRef]
3. Rajendran, G.B.; Kumarasamy, U.M.; Zarro, C.; Divakarachari, P.B.; Ullo, S.L. Land-use and land-cover classification using a human group-based particle swarm optimization algorithm with an LSTM Classifier on hybrid pre-processing remote-sensing images. *Remote Sens.* **2020**, *12*, 4135. [CrossRef]
4. Wu, H.; Prasad, S. Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Trans. Image Process.* **2017**, *27*, 1259–1270. [CrossRef]
5. Zhou, F.; Hang, R.; Liu, Q.; Yuan, X. Hyperspectral image classification using spectral-spatial LSTMs. *Neurocomputing* **2019**, *328*, 39–47. [CrossRef]
6. Cheng, G.; Han, J. A survey on object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 11–28. [CrossRef]

7. Zhao, P.; Gao, H.; Zhang, Y.; Li, H.; Yang, R. An aircraft detection method based on improved mask R-CNN in remotely sensed imagery. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; IEEE: Piscataway, NJ, USA , 2019; pp. 1370–1373.
8. Han, Q.; Yin, Q.; Zheng, X.; Chen, Z. Remote sensing image building detection method based on Mask R-CNN. *Complex Intell. Syst.* **2021**, *8*, 1847–1855. [CrossRef]
9. Chen, J.; Sun, J.; Li, Y.; Hou, C. Object detection in remote sensing images based on deep transfer learning. *Multimed. Tools Appl.* **2022**, *81*, 12093–12109. [CrossRef]
10. Yu, D.; Ji, S. A new spatial-oriented object detection framework for remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [CrossRef]
11. Shivappriya, S.; Priyadarsini, M.J.P.; Stateczny, A.; Puttamadappa, C.; Parameshachari, B. Cascade object detection and remote sensing object detection method based on trainable activation function. *Remote Sens.* **2021**, *13*, 200. [CrossRef]
12. Deng, Z.; Sun, H.; Zhou, S.; Zhao, J.; Lei, L.; Zou, H. Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 3–22. [CrossRef]
13. Gu, W.; Lv, Z.; Hao, M. Change detection method for remote sensing images based on an improved Markov random field. *Multimed. Tools Appl.* **2017**, *76*, 17719–17734. [CrossRef]
14. Shafique, A.; Cao, G.; Khan, Z.; Asad, M.; Aslam, M. Deep learning-based change detection in remote sensing images: A review. *Remote Sens.* **2022**, *14*, 871. [CrossRef]
15. Hussain, M.; Chen, D.; Cheng, A.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 91–106. [CrossRef]
16. Chen, H.; Qi, Z.; Shi, Z. Remote sensing image change detection with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [CrossRef]
17. Yuan, X.; Shi, J.; Gu, L. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Syst. Appl.* **2021**, *169*, 114417. [CrossRef]
18. Kemker, R.; Salvaggio, C.; Kanan, C. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 60–77. [CrossRef]
19. Li, R.; Zheng, S.; Zhang, C.; Duan, C.; Su, J.; Wang, L.; Atkinson, P.M. Multiattention network for semantic segmentation of fine-resolution remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–13. [CrossRef]
20. Dong, Y.; Chen, F.; Han, S.; Liu, H. Ship object detection of remote sensing image based on visual attention. *Remote Sens.* **2021**, *13*, 3192. [CrossRef]
21. Jian, L.; Pu, Z.; Zhu, L.; Yao, T.; Liang, X. SS R-CNN: Self-Supervised learning improving mask R-CNN for ship detection in remote sensing images. *Remote Sens.* **2022**, *14*, 4383. [CrossRef]
22. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Volume 28.
23. 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, Washington, DC, USA, 23–28 June 2014; pp. 580–587.
24. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
25. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
26. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
27. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
28. Maji, S.; Rahtu, E.; Kannala, J.; Blaschko, M.; Vedaldi, A. Fine-grained visual classification of aircraft. *arXiv* **2013**, arXiv:1306.5151.
29. Yu, L.; Hu, H.; Zhong, Z.; Wu, H.; Deng, Q. GLF-Net: A target detection method based on global and local multiscale feature fusion of remote sensing aircraft images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
30. Zhang, N.; Xu, H.; Liu, Y.; Tian, T.; Tian, J. AFA-NET: Adaptive feature aggregation network for aircraft fine-grained detection in cloudy remote sensing images. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 1704–1707.
31. Wei, H.; Zhang, Y.; Wang, B.; Yang, Y.; Li, H.; Wang, H. X-LineNet: Detecting aircraft in remote sensing images by a pair of intersecting line segments. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 1645–1659. [CrossRef]
32. Liu, C.; Yu, H.; Wei, H.; Sun, X.; Fu, K. S2CGNet: A robust aircraft detector based on the sword-shaped component geometry. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [CrossRef]
33. Zhao, Y.; Li, J.; Li, W.; Shan, P.; Wang, X.; Li, L.; Fu, Q. MS-IAF: Multi-Scale information augmentation framework for aircraft detection. *Remote Sens.* **2022**, *14*, 3696. [CrossRef]
34. Kwon, H. Adversarial image perturbations with distortions weighted by color on deep neural networks. *Multimed. Tools Appl.* **2023**, *82*, 13779–13795. [CrossRef]
35. Kwon, H.; Kim, S. Dual-Mode Method for Generating Adversarial Examples to Attack Deep Neural Networks. *IEEE Access* **2023**, 1. [CrossRef]

36. Kwon, H.; Lee, S. Toward Backdoor Attacks for Image Captioning Model in Deep Neural Networks. *Secur. Commun. Netw.* **2022**, *2022*, 1525052. [CrossRef]
37. Kwon, H.; Lee, J. AdvGuard: fortifying deep neural networks against optimized adversarial example attack. *IEEE Access* **2020**, 1. [CrossRef]
38. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
39. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
40. Law, H.; Deng, J. Cornernet: Detecting objects as paired keypoints. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 734–750.
41. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6569–6578.
42. Yang, Z.; Liu, S.; Hu, H.; Wang, L.; Lin, S. Reppoints: Point set representation for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9657–9666.
43. Dong, Z.; Li, G.; Liao, Y.; Wang, F.; Ren, P.; Qian, C. Centripetalnet: Pursuing high-quality keypoint pairs for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10519–10528.
44. Li, K.; Cheng, G.; Bu, S.; You, X. Rotation-insensitive and context-augmented object detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 2337–2348. [CrossRef]
45. Fu, K.; Chang, Z.; Zhang, Y.; Xu, G.; Zhang, K.; Sun, X. Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *161*, 294–308. [CrossRef]
46. Qian, X.; Lin, S.; Cheng, G.; Yao, X.; Ren, H.; Wang, W. Object detection in remote sensing images based on improved bounding box regression and multi-level features fusion. *Remote Sens.* **2020**, *12*, 143. [CrossRef]
47. Yao, Q.; Hu, X.; Lei, H. Multiscale convolutional neural networks for geospatial object detection in VHR satellite images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 23–27. [CrossRef]
48. Liu, Y.; Li, Q.; Yuan, Y.; Du, Q.; Wang, Q. ABNet: Adaptive balanced network for multi-scale object detection in remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [CrossRef]
49. Ye, Y.; Ren, X.; Zhu, B.; Tang, T.; Tan, X.; Gui, Y.; Yao, Q. An adaptive attention fusion mechanism convolutional network for object detection in remote sensing images. *Remote Sens.* **2022**, *14*, 516. [CrossRef]
50. Liu, Q.; Xiang, X.; Wang, Y.; Luo, Z.; Fang, F. Aircraft detection in remote sensing image based on corner clustering and deep learning. *Eng. Appl. Artif. Intell.* **2020**, *87*, 103333. [CrossRef]
51. Shi, L.; Tang, Z.; Wang, T.; Xu, X.; Liu, J.; Zhang, J. Aircraft detection in remote sensing images based on deconvolution and position attention. *Int. J. Remote Sens.* **2021**, *42*, 4241–4260. [CrossRef]
52. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo algorithm developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [CrossRef]
53. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef]
54. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
55. Wang, R.J.; Li, X.; Ling, C.X. Pelee: A real-time object detection system on mobile devices. In Proceedings of the Advances in Neural Information Processing Systems 2018, Montreal, QC, Canada, 3–8 December 2018; Volume 31.
56. Wang, K.; Liew, J.H.; Zou, Y.; Zhou, D.; Feng, J. Panet: Few-shot image semantic segmentation with prototype alignment. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9197–9206.
57. Song, G.; Liu, Y.; Wang, X. Revisiting the sibling head in object detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11563–11572.
58. Zhu, H.; Chen, X.; Dai, W.; Fu, K.; Ye, Q.; Jiao, J. Orientation robust object detection in aerial images using deep convolutional neural network. In Proceedings of the 2015 IEEE International Conference on Image Processing, Quebec City, QC, Canada, 27–30 September 2015; pp. 3735–3739.
59. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
60. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2849–2858.
61. Yang, X.; Yan, J. Arbitrary-oriented object detection with circular smooth label. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 677–694.
62. Yang, X.; Hou, L.; Zhou, Y.; Wang, W.; Yan, J. Dense label encoding for boundary discontinuity free rotation detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 July 2021; pp. 15819–15829.

63. Zhou, L.; Wei, H.; Li, H.; Zhao, W.; Zhang, Y.; Zhang, Y. Arbitrary-oriented object detection in remote sensing images based on polar coordinates. *IEEE Access* **2020**, *8*, 223373–223384. [CrossRef]
64. Zhang, F.; Wang, X.; Zhou, S.; Wang, Y. DARDet: A dense anchor-free rotated object detector in aerial images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]