

Article

UAV Image Small Object Detection Based on RSAD Algorithm

Jian Song ¹, Zhihong Yu ^{1,*}, Guimei Qi ², Qiang Su ¹, Jingjing Xie ¹ and Wenhao Liu ¹ 

¹ College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China; songjian9703@emails.imau.edu.cn (J.S.); suqi0203@emails.imau.edu.cn (Q.S.); 18232076601@163.com (J.X.); lwh20222022@126.com (W.L.)

² College of Computer Science and Technology, Inner Mongolia Normal University, Hohhot 010020, China; ciecqgm@imnu.edu.cn

* Correspondence: yzhyq@imau.edu.cn; Tel.: +86-13848172458

Abstract: There are many small objects in UAV images, and the object scale varies greatly. When the SSD algorithm detects them, the backbone network's feature extraction capabilities are poor; it does not fully utilize the semantic information in the deeper feature layer, and it does not give enough consideration to the little items in the loss function, which result in serious missing object detection and low object detection accuracy. To tackle these issues, a new algorithm called RSAD (Resnet Self-Attention Detector) that takes advantage of the self-attention mechanism has been proposed. The proposed RSAD algorithm utilises the residual structure of the ResNet-50 backbone network, which is more capable of feature extraction, in order to extract deeper features from UAV image information. It then utilises the SAFM (Self-Attention Fusion Module) to reshape and concatenate the shallow and deep features of the backbone network, selectively weighted by attention units, ensuring the efficient fusion of features to provide rich semantic features for small object detection. Lastly, it introduces the Focal Loss loss function, which adjusts the corresponding parameters to enhance the contribution of small objects to the detection model. The ablation experiments show that the mAP of RSAD is 10.6% higher than that of the SSD model, with SAFM providing the highest mAP enhancement of 7.4% and ResNet-50 and Focal Loss providing 1.3% and 1.9% enhancements, respectively. The detection speed is only reduced by 3FPS, but it meets the real-time requirement. Comparison experiments show that in terms of mAP, it is far ahead of Faster R-CNN, Cascade R-CNN, RetinaNet, CenterNet, YOLOv5s, and YOLOv8n, which are the mainstream object detection models; In terms of FPS, it slightly inferior to YOLOv5s and YOLOv8n. Thus, RSAD has a good balance between detection speed and accuracy, and it can facilitate the advancement of the UAV to complete object detection tasks in different scenarios.

Keywords: UAV images; RSAD; small object; self-attention mechanism; feature fusion



Citation: Song, J.; Yu, Z.; Qi, G.; Su, Q.; Xie, J.; Liu, W. UAV Image Small Object Detection Based on RSAD Algorithm. *Appl. Sci.* **2023**, *13*, 11524. <https://doi.org/10.3390/app132011524>

Academic Editor: Sungho Kim

Received: 7 September 2023

Revised: 11 October 2023

Accepted: 19 October 2023

Published: 20 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Unmanned aerial vehicles (UAVs) are becoming an essential component of daily life thanks to the ongoing development of UAV technology. UAVs are frequently employed in many fields because they are compact, adaptable, simple to use, and economical [1]. Drone usage is evident in both military and non-military operations. As an illustration, traffic monitoring, power patrol, locating military targets at sea, and emergency disaster aid [2–5]. However, object detection is the primary technology that enables UAVs to complete these tasks. UAVs can capture aerial images in a variety of environments by carrying optical elements on their fuselage [6,7]. Object detection technology is then used to process the images captured by UAVs and analyse the feature information in the images, which can greatly reduce the human cost and improve the monitoring efficiency. Therefore, how to process and analyse UAV images quickly, accurately and intelligently has become a research focus in the field of UAV imagery.

In recent years, the break-through in computer hardware and the emergence of deep learning have led to the rapid development of object detection technology, which, as an important technology for UAV image processing, enhances the UAV's ability to understand different scenarios and can help the UAV to perform tasks in different scenarios [8,9]. More specifically, deep learning algorithms for object detection perform significantly better than traditional methods. The feature learning framework in deep learning solves several problems that existed in traditional methods, such as generating a large number of redundant proposals through time-consuming selective search methods, as well as the manual feature design process, the selective search methods, and the manual feature design process. Recent progress of deep neural networks has significantly boosted the performance of object detection on public benchmarks such as COCO [10]. However, UAV images [11] are different from public benchmarks. Due to the UAV's large field of view, the scale of objects in UAV images varies greatly, and small objects form a significant proportion, whereas current research focuses on developing complex models to achieve high accuracy for small objects on high-resolution UAV images, which requires a lot of processing work. However, complex models are not easy to deploy on UAVs when UAV resources are limited. All these make UAV object detection more difficult than general object detection.

Initial deep learning-based object detection algorithms, such as Faster R-CNN [12], Mask R-CNN [13], and YOLO [14] only perform prediction and regression in the last layer of the backbone network. This detection method is difficult to adapt to changes in the scale of objects and retains less-effective information about small objects, resulting in low accuracy of UAV image detection. The SSD [15] algorithm merges the anchor frame concept of Faster R-CNN with the regression concept of YOLO and proposes the introduction of multiple convolutional classifiers and regressors on feature maps of different scales for object detection. This method ensures robustness to scale changes and offers a new perspective for UAV image detection. Nowadays, most of the object detection algorithms based on UAV images retain the characteristics of multi-scale detection of SSD algorithm [16–18] and improve on it. Zhang et al. [19] proposed a global density fusion convolutional network for small objects in UAV images, which improves the model's detection rate of small objects in UAV images by fusing the feature layers of different scales, but increases the computational burden. In order to enhance the fusion efficiency of different feature layers, Ma Junyan et al. [20] proposed an MFT structure, which gives weights to different feature layers for fusion through the attention mechanism, solving the problem that different feature layers have the same influence on the results when fused, but the specific information in different feature layers is not considered for UAV image detection, and the improvement of the detection accuracy of UAV images is limited. Tian et al. [21] proposed a dual neural network review method, by classifying the secondary features of the suspected object area, quickly screening the missed objects in the primary detection, extracting the feature map of the UAV image by using the VGG backbone, and combining the feature map and the location information of the suspected area for the secondary identification to realize the high-quality detection of the small objects in the UAV image; however, this structure increases the complexity of the model and makes the detection speed to decrease. The object detection algorithm of UAV image at the present stage improves the detection accuracy of UAV image by inefficient feature fusion and increasing the attention mechanism, but the detection speed is slow, and the practicality is poor.

In this study, a RSAD algorithm for object detection in UAV images is proposed from SSD, and a feature fusion module SAFM with self-attention mechanism is designed, through which information from different layers is efficiently fused, and residual learning and Focal Loss loss function are utilized to improve the detection accuracy of UAV images with less loss in detection speed.

2. Related Work

2.1. Object Detection in UAV Images

Object detection was initially applied in the detection of general images. It is usually classified into two main categories, two-stage object detection algorithms and one-stage object detection algorithms, and they differ in their strengths and limitations.

The two-stage algorithm is known for its high detection accuracy, such as the Faster R-CNN mentioned above. It first generates the proposed region from the image and then classifies and localises the objects in it. However, for UAV images, its detection accuracy is drastically reduced. To address this phenomenon, many scholars have conducted research. Liu et al. [22] use VGGNet and ResNet as the feature extraction networks in Faster R-CNN to detect maize tassels, both with images from UAV. They modified the anchor size in the region proposal network to match the real size of the tassel pixels, which improved the detection accuracy. However, detection efficiency has a long way to go. The ROI transformer, developed by Ding et al. [23], produces rotational ROIs using RPN from horizontal ROIs. This considerably improves the detection accuracy of oriented objects. However, the network became heavy and sophisticated due to fully linked layers and a ROI alignment mechanism that rotated ROIs. For UAV images, a label-oriented object format called gliding vertexes was introduced by Xu et al. [24] and acquired four vertex gliding offsets on the regression branch of the Faster R-CNN head. It can accurately detect multi-directional objects, but it also introduces a certain amount of computation and increases the computational burden.

One-stage object detectors are efficient but usually have lower accuracy, such as SSD mentioned above. The core of the one-stage object detection algorithm is regression, which eliminates the region proposal aspect of the two-stage method; there will be no preclassification or regression process, and the specific categories are directly classified and regressed against the boxes [25]. Arunehru L. et al. [26] achieved real-time detection of UAV images by performing target frame clustering analysis, pre-training the model, employing multi-scale training, and optimizing the filtering rules of candidate frames for the YOLO algorithm from the perspective of detection efficiency. However, the detection accuracy is slightly lower. Inspired by YOLO architecture, Elhagry A. et al. [27] propose a novel single-stage detection architecture. Oghaz D. et al. [28] used SSD as a baseline network and improved SSD using feature enhancement modules such as super-resolution, inverse plethysmography and feature fusion. These improvements effectively enhanced the detection accuracy of SSD for UAV images, but at the expense of a significant amount of detection speed. Yundong, L. I. et al. [29] designed a multi-block SSD mechanism consisting of three steps to address the limitations of small object detection for UAV surveillance of railway scenes. The original input image is first divided into four overlapping blocks. The next step is to send each block to an SSD independently for detection. The overlapping frames of the sublayers are then eliminated using non-maxima splicing and integration of the detection results from the four blocks. This can lower the leakage rate and considerably increase the detection accuracy of small objects in UAV images. However, it decreases the detection efficiency and makes the model more complex.

2.2. SSD Object Detection Framework

SSD algorithm, which combines the anchor mechanism from Faster R-CNN and the regression concept from YOLO, is a suggested object detection technique. According to [15], in the common public datasets VOC and COCO, SSD not only outperforms Faster R-CNN in detection accuracy, but also outperforms YOLO in detection speed. It is an end-to-end single-point multi-frame one-stage object detection algorithm, which classifies and regresses the predictions of objects on different scales of feature layers through the anchor frame mechanism, and finally excludes low-scoring prediction frames through the Non-Maximum Suppression (NMS) algorithm to obtain the detection results [30]. The specific detection process is as follows: the backbone network VGG-16 extracts image features to create six feature maps at different scales for simultaneous detection of objects at

different scales; then, a set of anchor boxes with different aspect ratios and sizes are tiled by pixels on the feature maps at different scales, and scores are generated for the presence of each object category in each default box as well as adjusting the shape. Finally, using NMS to output the location and category of the object box, the algorithm structure is shown in Figure 1.

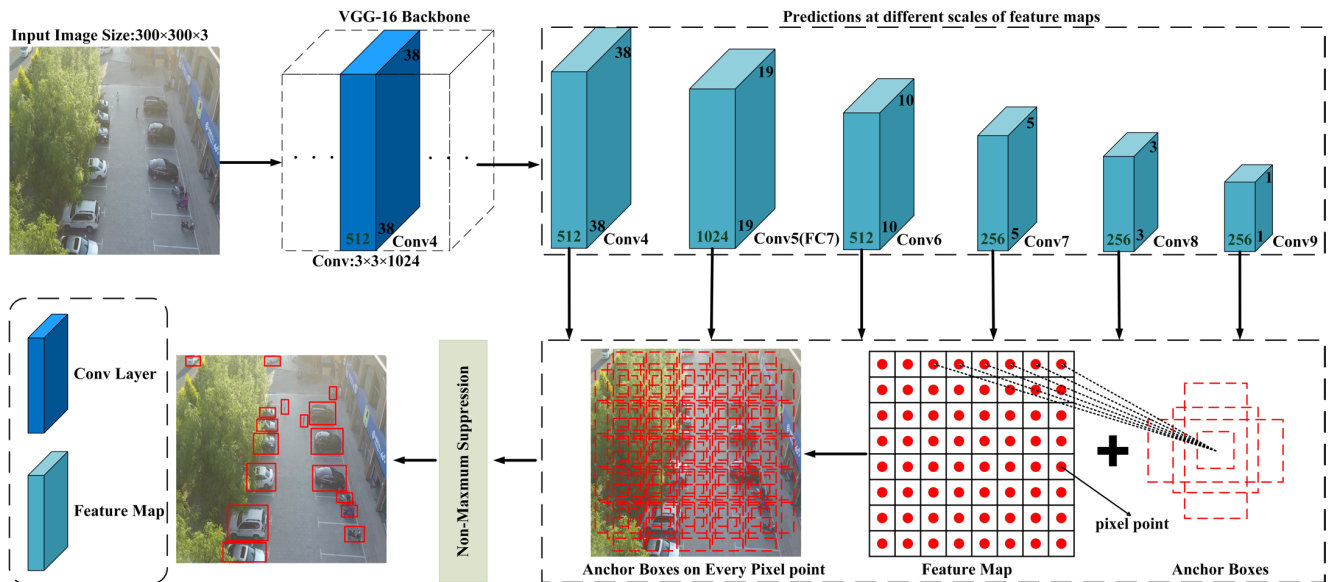


Figure 1. The network structure of SSD.

The SSD object detection algorithm has poor accuracy in detecting small objects in UAV images [31], mainly due to the following reasons: The layers of the backbone network VGG-16 are too shallow for feature extraction of small objects [32], and more advanced semantic features cannot be obtained; the algorithm's six feature maps for classification and regression are independent of one another, and the shallow feature maps used for small object recognition cannot fully utilize the deeper feature maps' semantic features [33]. Small objects, which are hard samples, contribute less to the SSD loss function during model training, causing the model to pay them less attention and resulting in lower detection accuracy. In summary, the network's structure of the SSD algorithm needs to be further modified in order to increase its detection performance for small objects in UAV images.

2.3. Datasets and Evaluation Metrics

The VisDrone UAV image dataset [34] is used for the experiments in this study; the training set has 6471 samples, the validation set contains 777 samples, and each sample is 1920×1080 pixels in size. The dataset consists of 10 classes: pedestrian, people, bicycle, car, van, truck, tricycle, awning tricycle, bus, and motor. More than 90% of the dataset's objects are small, which is a predominance (Figure 2).

The assessment of object detection properties often relies on two metrics: Average Precision (AP) and mean Average Precision (mAP) [35], computed based on the intersection over union (IOU) between ground truth and predicted bounding boxes. AP is obtained from the area surrounded by the precision-recall curves (PR curve) and the axes. Additionally, mAP is the averaged AP value for all classes. The precision can gauge how well the object detection algorithm can classify objects. Additionally, the recall can gauge how well the model can identify an object. The higher value of precision and recall, the object detection algorithm performs better in classifying and identifying. The following demonstrates how the *Precision* and *Recall* are calculated:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

In Equations (1) and (2), TP is true positive, which denotes the number of positive samples whose prediction result is true; FP is false positive, which denotes the number of negative samples whose prediction result is true; and FN , false negative, denotes that the number of positive samples whose prediction result is false. Apart from that, the number of model parameters and the FPS were picked as indicators of the model's complexity and detection speed, respectively.



Figure 2. Display of the VisDrone UAV dataset and object's width and height distribution.

2.4. Motivation and Contribution

Combined with the previously mentioned, the existing object detection algorithms for UAV images do not balance detection accuracy and detection speed. They often sacrifice a large amount of detection speed in pursuit of high detection accuracy. This does not take advantage of the practical application of object detection algorithms on UAVs. Therefore, we propose RSAD, which makes a good balance between detection speed and detection accuracy. Our main contributions are as follows:

- (1) By comparing backbone networks geared towards real-world applications and considering UAV resource constraints, we chose ResNet-50 as the backbone network for RSAD and continued to use multiscale detection.
- (2) We designed a feature fusion module based on the self-attention mechanism. Different from the traditional attention mechanism, we pay more attention to the weight relationship within the image. By reshaping and concatenating the information extracted into the image by the backbone network, the correlation between the information inside the image is established. It helps the network to better understand the information in the UAV image.
- (3) We introduced the Focal Loss loss function to enhance RSAD's focus on small targets. The parameters of the Focal Loss loss function are adjusted through experiments with a small number of iterations to make RSAD pay the best attention to small objects.

3. Proposed Method

3.1. Establishment of the Backbone Network

The object detection algorithm's foundation, the backbone network, directly influences how well it performs in terms of object detection [36]. As we all know, medium and large object recognition is more advantageous with deeper backbone network layers since more information can be gleaned from the image. Because there are a lot of small objects in UAV images and the background is complicated, the deeper layer backbone network's feature

extraction effect on small objects is less effective than it is on medium and large objects. As the layers are deepened, the information pertaining to small objects is also lost or even completely lost. At the same time, considering the practical application of the algorithm, the number of parameters of the backbone network should not be too large, otherwise it is difficult to be integrated into the UAV system.

In this study, in order to compare the quality of different backbone networks oriented to real applications, we collected and analysed the accuracy values (based on LSVRC dataset) reported in the literature are shown in Figure 3 [37–42]. From the figure, it can be seen that ResNet-152, ResNet-101, and ResNet-50 achieved the top three accuracies. The reason why the ResNet backbone network has such a good performance is because of the use of the residual structure. The proposed residual structure cleverly solves the problem of accuracy degradation by skip-connection, which will allow early feature $s(x)$ to propagate further in the architecture ($H(x)$), as opposed to something like VGG-16 as in SSD. Figure 4 shows the difference between the normal structure and the residual structure. However, considering the limited resources of the UAV. The number of parameters of the backbone network and the size of the computation affect whether it can be applied on UAVs or not. In this study, the number of parameters and the amount of computation of these backbone networks are calculated by Pytorch in Python and are shown in the following table.

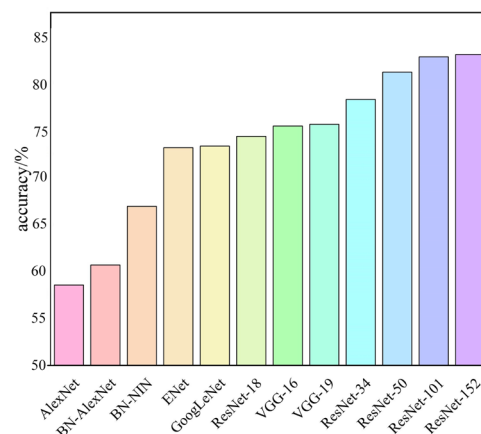


Figure 3. Different backbone network accuracy.

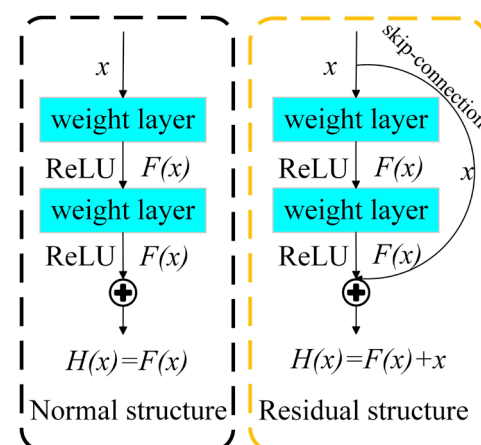


Figure 4. Comparison of normal and residual structures.

From Figure 3 and Table 1, it can be seen that AlexNet and BN-AlexNet are poor both in terms of accuracy and the number of parameters and flops. The BN-MIN, Enet, GoogLeNet, ResNet-1 and ResNet-34 has suitable Flops, few number of parameters, but less accuracy. The VGG-16 and VGG-19 though have a good accuracy, the parameters and flops

are too large. The ResNet-50, ResNet-101 and ResNet-152 have a better accuracy, however, The ResNet-101 and ResNet-152 are complicated compared to ResNet-50. Comprehensive consideration, the ResNet-50 is more appropriate for UAV images. As a result, ResNet-50 is used in this study as the RSAD backbone network, retaining the multi-scale feature map identification capabilities of the SSD object detection algorithm.

Table 1. Number of parameters and Flops for different backbone networks.

Backbone	A	B	C	D	E	F	G	H	I	J	K	L
Parameters/M	61.1	65	6.8	4.5	6.6	11.7	138.4	143.7	21.8	25.6	44.6	60.2
Flops/G	1.2	0.8	1.15	0.6	2.6	3.4	27.2	34.6	6.9	7.71	14.9	21.5

note: A: AlexNet; B: BN-AlexNet; C: BN-NIN; D: ENet; E: GoogLeNet; F: ResNet-18; G: VGG-16; H: VGG-19; I: ResNet-34; J: ResNet-50; K: ResNet-101; L: ResNet-152.

3.2. Self-Attention Mechanism Based Feature Fusion

The self-attention mechanism is a variant of the attention mechanism that is less dependent on external information and more adept at capturing the internal correlations of data or features. The self-attention mechanism gives weights to features and correlations between features and features through convolution and softmax to establish global feature dependencies. In this study, a self-attention mechanism feature fusion module SAFM is designed based on this, as shown in Figure 5. The method aims to organically fuse the high-level semantic information of small objects in the deep layer of the UAV image through the self-attention mechanism with the shallow layer after establishing the correlation, which is no longer a simple stacking. At the same time, large weights are given to important information and small weights are given to background and noise information.

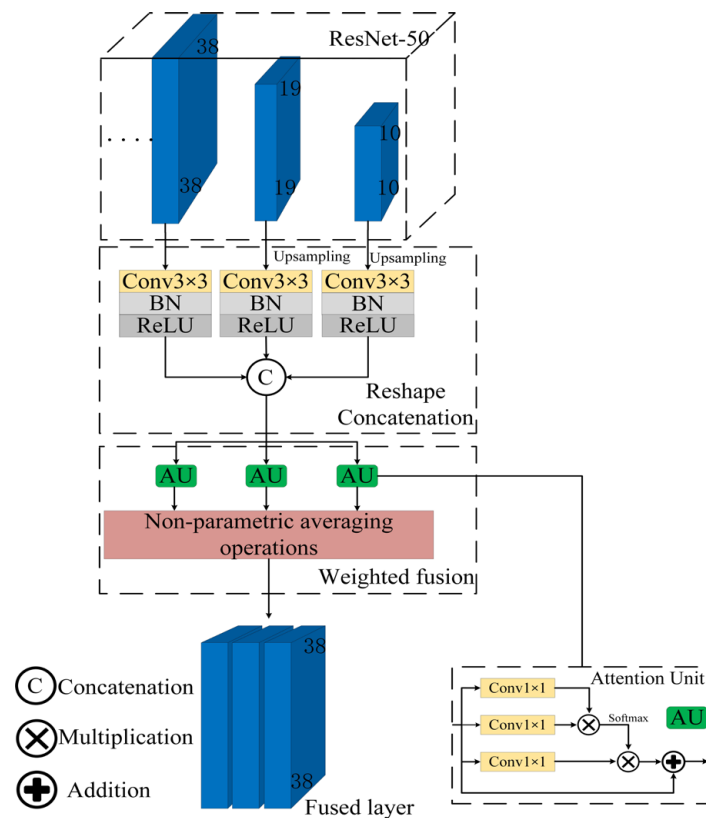


Figure 5. SAFM structure.

SAFM first reshaped and concatenated the features extracted from the ResNet-50 backbone network, and then obtains the global information. The global information is obtained by the following equation:

$$X = [f_i^{3 \times 3}(s_i)], \quad i = 1, 2, \dots, n \quad (3)$$

where: s_i is the feature information extracted by the backbone network, $X \in \mathbb{R}^{C \times H \times W}$, C denotes the resultant tensor's channel dimension, $H \times W$ denotes the resolution of the feature map, $f_i^{3 \times 3}$ denotes that the convolutional layer with 3×3 kernel size to unify the channel dimension. $[]$ denotes the concatenation calculation, and n represents the number of fused feature layers in a pyramid.

The acronym AU refers for self-attention mechanism, which enables the computation of the pertinent data in a single feature and the further establishment of the global feature correlation at the pixel level. Note that the AU is shown in the lower right corner of Figure 4. Receiving a single tensor obtained by concatenation, AU uses two lightweight convolutional layers with the kernel size of 1×1 are adopted to generate features Q and K . $Q, K \in \mathbb{R}^{C \times H \times W}$, $C' = C/8$ is used to reduce computing overhead. Then, based on the features Q and K , the global weighted feature mapping A can be obtained by matrix multiplication followed by softmax operation

$$D = Q^T K \quad (4)$$

$$A = \frac{e^{D_{ij}}}{\sum_{j=1}^{H \times W} e^{D_{ij}}}, \quad i, j = 1, 2, 3, \dots, H \times W \quad (5)$$

where $D \in \mathbb{R}^{HW \times HW}$ and A_{ij} represents the elements of the i th row and j th column.

After obtaining A , $V \in \mathbb{R}^{C \times H \times W}$ is obtained by another convolution kernel. Then, the weighted sum of all positions in the same feature is obtained by multiplying the feature with the weighted mapping A . Finally, the original concatenated features and the multiplication result are added to obtain the output of the AU.

$$X' = X + (AV^T)^T \quad (6)$$

In order to alleviate the problem of increasing the order of magnitude of feature information, after obtaining the outputs of the AU units, a nonparametric averaging operation is used in the fusion process to obtain the final fusion result:

$$H_i = \frac{f_i^{1 \times 1}(X'_i) + f_{i+1}^{1 \times 1}(X'_{i+1}) + f_{i+2}^{1 \times 1}(X'_{i+2})}{3}, \quad i \in n-1 \quad (7)$$

where: layer i is shallower than layers $i+1$ and $i+2$, $f^{1 \times 1}$ is used to reduce the number of channels.

SAFM feature fusion is different from traditional feature fusion in that it is not simply adding the features of different layers, but reshaping and concatenating the features of different layers and then weighting and filtering them through AU units. On the 38×38 -sized feature layer, it is enriched with a lot of geometric information and some background noise due to its large receptive field. Directly using it for detecting UAV images is unhelpful for detecting small objects in the image. Therefore, the information in the 19×19 -sized, 10×10 -sized feature layer is introduced into the shallow layer using SAFM. The information extracted in the 19×19 , 10×10 feature layer is more specific (semantic information) due to its small receptive field. After reshaping, concatenating, and fusing with the information in the 38×38 layer, the correlation between geometric and semantic information is established through the AU unit, which is more conducive to the detection of small objects in the UAV image, and also suppresses the interference caused by the background noise. The fusion layer H_i obtained through the SAFM module contains both the geometric

information of the shallow feature layer and the semantic information of the deep feature layer, which can effectively guide the model to detect small objects in UAV images.

3.3. Loss Function

Since small objects in UAV images occupy fewer pixels and contribute less to the loss function, it leads to insufficient attention to small objects in the training phase of the algorithm, which in turn leads to poor overall detection accuracy of the trained model for small objects in UAV images. In order to solve such problems, the Focal Loss has been proposed and achieved certain effects [43]. The Focal Loss can increase the importance of small objects in the loss function by increasing the value of the modulation coefficient γ , increasing the algorithm's focus on small objects in the training phase, thus improving the model's detection accuracy of small objects, the focus loss function formula is as follows:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (8)$$

In the formula, $FL(p_t)$ is the Focal Loss function, α_t is a decimal between 0 and 1, which is used as a weight to adjust the proportion of positive and negative samples, p_t represents the classification probabilities of the different classes, and γ ranging from 0 to 5, which is used to adjust the rate at which the weights of the samples are easily divided. In the case of imbalance between the positive and negative samples, the weight factor α_t , which affects the proportion of the positive and negative samples contributing to the loss function, can be set flexibly according to the data set. On the other hand, when the difficult and easy samples are unbalanced, the γ can be adjusted so as to reduce the contribution of the easy-to-categorize samples to the loss function, which makes the model training more focused on the difficult-to-categorize samples. When the authors proposed and analyzed the Focal Loss function in the study, it was based on binary classification, and Equation (8) was also derived based on binary classification. In this study, when object detection is carried out, it is a multiclassification case, so Equation (8) can be changed to Equation (9):

$$FL(p_c) = -\alpha_c(1 - p_c)^\gamma \log(p_c) \quad (9)$$

where: c denotes the category and the other parameters are the same as in Equation (8).

Equation (9) is only applied on classification, for the localisation loss which is also known as IoU loss. The localisation loss in SSD will be followed and the formula is as follows:

$$L_{loc}(x, l, g) = \sum_{i \in Pos, m \in \{cx, cy, w, h\}}^N \sum x_{ij}^k smooth_{L1} \left(l_i^m - \hat{g}_j^m \right) \quad (10)$$

where Pos stands for position, N represents the number of prediction boxes; m for centroid coordinates and width and height of the surrounding box; x_{ij}^k for whether the i th prediction box matches the j th real box with respect to category k : 0, 1; l_i^m stands for centroid coordinates and width and height information of the i th prediction box, $\hat{g}_j^m$ stands for centroid coordinates and width and height information of the i th real box. The difference between the prediction boxes and the ground truth boxes is smoothed by smoothL1 after the difference between the prediction boxes and the ground truth boxes has been determined. This allows the loss to be more stable and not deviate from it along with the anomalous values, reducing the sensitivity of the loss to those values.

Since γ dominates in Focal Loss. So, we set α_c to the default value of 0.25. For the determination of the value of γ , we did a pre-experiment with a small number of iterations on the VisDrone dataset. We set γ to 0, 1, 2, 3, 4, 5, and other parameters are the same as SSD, and performed 1000 iterations of the experiment to obtain the detection accuracy of various classes, respectively (Figure 6).

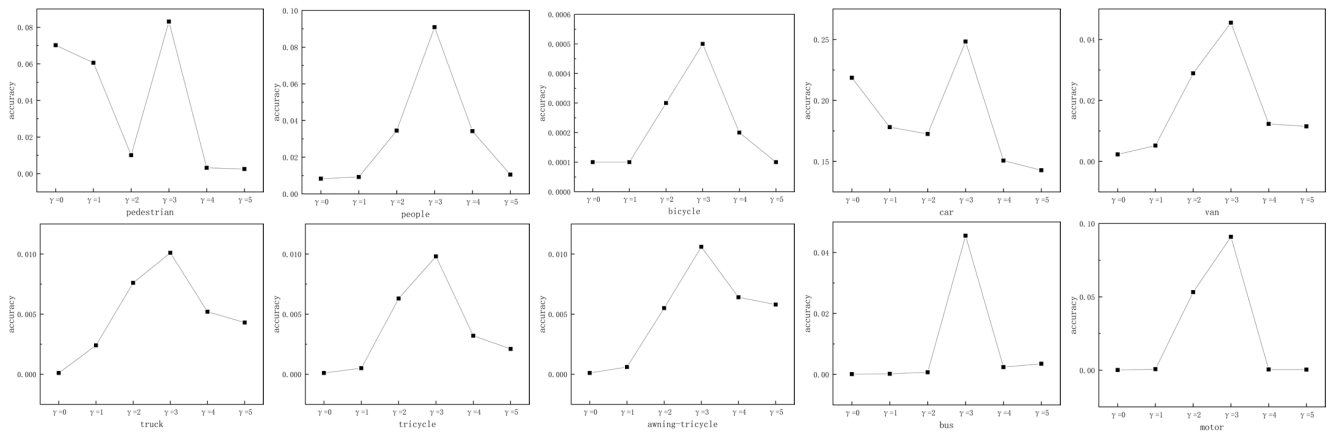


Figure 6. Effect of different γ values on detection accuracy.

It can be observed from Figure 6 that after 1000 iterations, the model has a slight ability to detect different kinds of detection targets. When the value of γ is 3, the model has outstanding detection ability for each kind. Therefore, the γ value is set to 3 for training in this study. The framework diagram of the RSAD network obtained after the above improvements is shown in Figure 7.

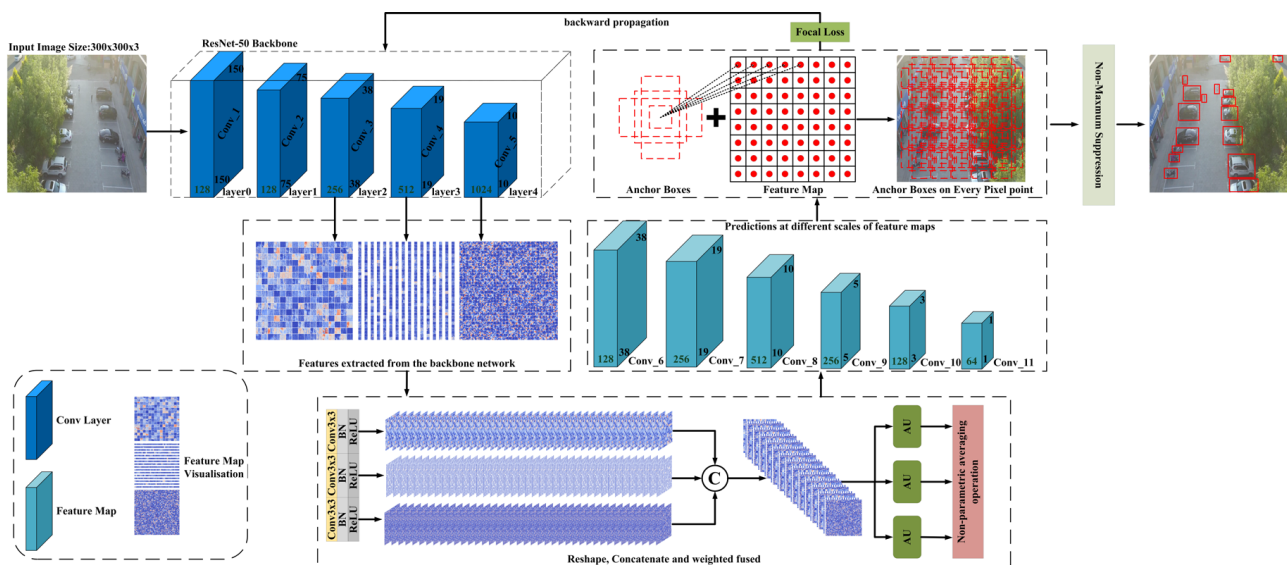


Figure 7. RSAD network structure.

4. Results

4.1. Experimental Setup

The experimental environment of this study is: in terms of hardware, we adopt the 12th Gen Intel Core i5-12400F hexa-core and the graphic card of the Nvidia GeForce RTX 3060 Ti 8 GB. In terms of software, we chose the Windows 10 Professional 64-bit operating system, Pytorch deep learning framework, and Python 3.7 programming language. We also used the Nvidia software CUDA 11.6 and CUDNN 11.6 to accelerate the deep learning.

In the training phase, we utilised some of the weights of the pretrained SSD model, significantly reducing the training time. To ensure consistency and fairness of the experiments, we trained the model to 120,000 iterations and fixed the batch size to 32. For efficiency, we normalised the width and height of the inputs to 300. We used the SGD optimiser and set all other parameters to the default configuration of SSD. The loss value results are taken every 10 iterations, and the accuracy results are taken every 1000 iterations.

4.2. Ablation Experiments

In order to verify the effectiveness of the improvement of each unit, a series of ablation experiments were conducted on the test set in this study. The mAP, the number of parameters, and the FPS (3060 Ti) were selected as the evaluation metrics, and the results of the experiments are shown in Table 2. The results of the ablation experiments show that after adding the SAFM feature fusion, the mAP improvement rate is the largest, and the improvement rate reaches 7.4%, which shows from the side that the SAFM module obtains global weighted feature mapping by reshaping and concatenating the features of the backbone network, which efficiently and selectively organically fuses the high-level semantic information in the deep layer and the geometric information in the shallow layer together, and the feature layer used for detection contains rich semantic and geometric information at the same time, which helps the model to better detect small objects in UAV images.

Table 2. Ablation experiment results.

VGG-16	ResNet-50	SAFM	Focal Loss	mAP/%	Parameters/B	FPS (3060 Ti)
✓				19.9	22.9	55
	✓			21.2	29.8	59
	✓	✓		28.6	37.4	52
	✓	✓	✓	30.5	37.4	52

The increase in accuracy also brings about an increase in the number of parameters and a slight decrease in FPS, but it also meets the demand for real-time performance. The application of the new backbone network and the introduction of the Focal Loss also have some improvement, with an improvement rate of 1.3% and 1.9%, respectively. The improvement in detection accuracy reflects that the feature extraction ability of ResNet-50 is better than that of VGG-16, and the loss function of Focal Loss increases the model's focus on small objects in UAV images. Although the model's parameters are improved by 14.5B, the FPS is not substantially reduced compared to the original SSD algorithm.

In Figure 8 the mAP curves during training are shown. The AP curve of the model rises rapidly between 0 and 20,000 iterations and converges by the time it reaches 100,000 iterations. From the figure, it can be seen that the method in this study clearly outperforms SSD. When replacing the backbone network with ResNet-50, it reduces the convergence speed of the model, but this effect is eliminated after adding SAFM and Focal Loss. Moreover, SAFM improves the model mAP significantly, proving that the SAFM module fusion strategy proposed in this study is feasible. After the introduction of Focal Loss, no obvious enhancement effect can be seen between 0 and 60,000 iterations, and the effect of Focal Loss is demonstrated after 60,000 iterations.

In order to validate the feature extraction capability of this study's model, the visualized feature heatmaps of this study's model and the SSD model were compared. The results are shown in Figure 9, from which it can be seen that before adding SAFM, the model extracted the feature information more roughly and there was a lot of background noise information. After adding SAFM, it can be clearly seen that the model extracts the features more accurately and eliminates a lot of background and noise information.

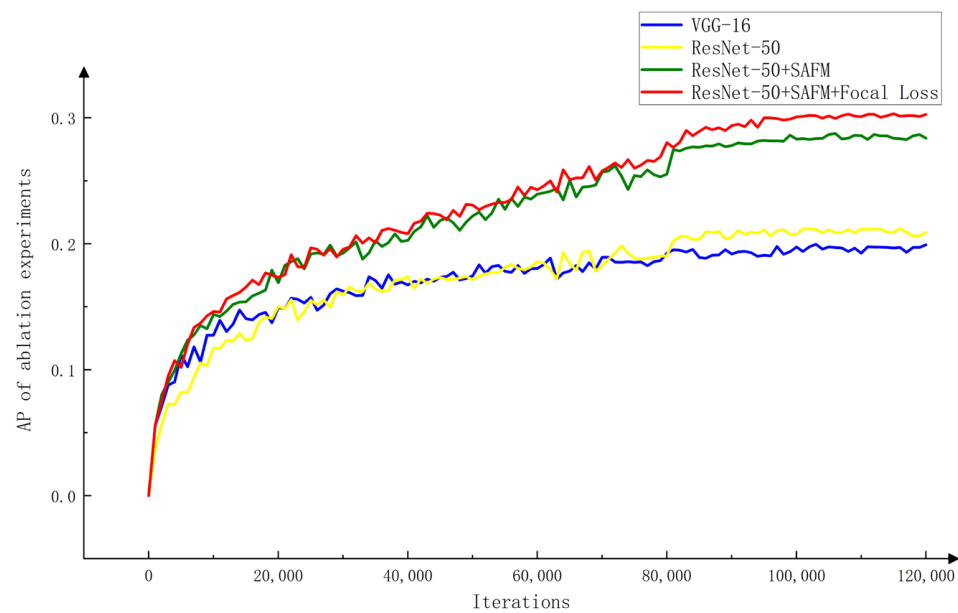


Figure 8. The mAP of ablation experiments.

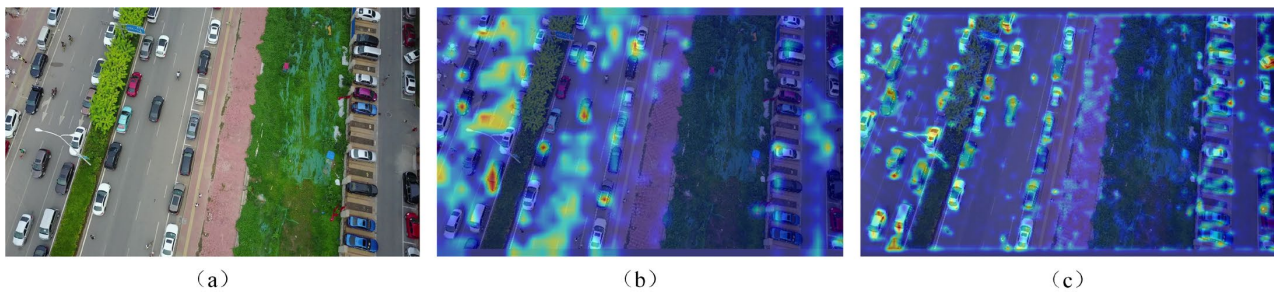


Figure 9. Comparison of visualization feature heatmaps. (a) Original image. (b) Feature map of the SSD model. (c) Feature map of the proposed model.

4.3. Comparison Experiment

The above experiments proved that the model in this study is better than the SSD model. In order to further explore the advantages and disadvantages of the RSAD object detection model proposed in this study with the existing excellent object detection models, this study trained a variety of excellent object detection models under the same experimental environments and datasets, including: the Faster R-CNN [44], the Cascade R-CNN [45], Retain-Net [46], Center-Net [47], YOLOv5s [48], and YOLOv8n [49], and the AP values for each object and the mAP values for all categories were obtained on the test dataset. Table 3 lists the results of each model on the test set. From the table, it can be seen that the model proposed in this study achieved the highest mAP value of 30.5% for the image input size of 300×300 , which gives the best overall detection performance for UAV images. The optimal AP value is achieved for the following values: bicycle, van, truck, tricycle, awning tricycle, bus single object. Slightly poorer detection accuracy was achieved for Pedestrian and Car, probably due to the high number of instances of Pedestrian and Car in the dataset, and the modulation factor of the Focal Loss function reduced the model's focus on these two categories, resulting in lower APs for these two object categories. The AP values on People and Motor individual objects are second only to the two-stage object detection model, but the detection speed is far superior to them, making it easier to be deployed on UAVs for real-world applications.

Table 3. Results of different algorithms on VisDrone test dataset.

Model	Backbone	AP/%										mAP /%	FPS
		P1	P2	B1	C	V	T1	T2	A	B2	M		
Faster R-CNN	ResNet-50	21.4	15.6	6.7	51.7	29.5	19.0	13.1	7.7	31.4	20.7	21.7	9.3
Cascade R-CNN	ResNet-50	22.2	14.8	7.6	54.6	31.5	21.6	14.8	8.6	34.9	21.4	23.2	12.1
RetinaNet	ResNet-50	13.0	7.9	1.4	45.5	19.9	11.5	6.3	4.2	17.8	11.8	13.9	16.7
CenterNet	Hourglass-104	14.8	13.2	5.6	50.2	24.0	21.3	20.1	17.4	37.9	23.7	22.8	14.0
YOLOv5s	CSPDarknet53	19.7	13.7	3.84	62.0	27.2	22.4	15.7	6.9	40.3	19.8	23.2	57.2
YOLOv8n	CSPNet	13.6	11.6	1.8	55.9	21.0	18.3	10.6	5.63	30.1	15.4	18.4	61.5
SSD300	VGG-16	13.6	11.3	9.1	42.6	25.8	26.1	13.2	6.8	37.9	12.6	19.9	55.0
RSAD(ours)	ResNet-50	15.1	14.0	10.5	51.8	43.0	46.1	25.5	19.1	59.5	19.9	30.5	52.0

note: image-size: 300×300 ; bolded text is the optimal result; P1: pedestrian; P2: people; B1: bicycle; C: car; V: van; T1: truck; T2: tricycle; A: awning tricycle; B2: bus; M: motor.

The above experiments proved that the RSAD model proposed in this study has the best integrated detection performance for UAV images compared to other excellent models when the image input is 300×300 . In order to more intuitively reflect the excellent performance of the RSAD model, the YOLOv5 model, which is second only to RSAD in terms of mAP value and has a similar detection speed to RSAD in Table 2, is selected for the comparison of detection results. Figure 10 shows the detection results of YOLOv5 and RSAD models in different environments as well as under different light conditions. From the Figure 10a, b, we can intuitively see that the YOLOv5 model has missed detection of objects in the image, while the RSAD model can accurately detect them. In Figure 10c, d, where the light is insufficient and the environment is more complex, the YOLOv5 model has missed detection due to the complex environment and insufficient light, but the RSAD model still maintains excellent detection performance and shows strong anti-interference ability. The detection results show that the RSAD model proposed in this study exhibits excellent detection performance for UAV images with a large number of small objects and large-scale transformations during daytime and nighttime, and RSAD can effectively suppress the interference of image background noise information, selectively mine important feature information, and effectively guide the UAV to complete the task of target detection in different environments.

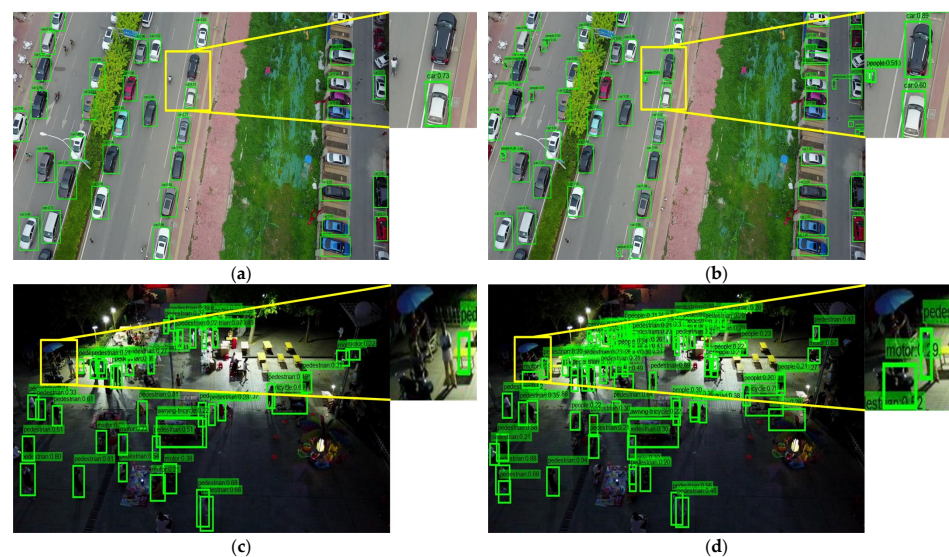


Figure 10. Display of results ((a,b) are a comparison of the detection effect of YOLOv5 and RSAD models in daytime; (c,d) is a comparison of the detection effect of YOLOv5 and RSAD models for nighttime).

5. Discussion

The constructed ResNet-50 backbone network and retaining the multi-scale detection enhance the model's ability of extracting features from small objects in UAV images, and keep good robustness to scale changes. The SAFM feature fusion module reshaped, concatenated, and weighted non-parametric averaging for features to improve the model's recognition of small object features in UAV images and efficiently and fully utilize the semantic and geometric information in the deep and shallow layers. The introduction of Focal Loss function improves the detection accuracy of the algorithm on small objects in UAV images without affecting the detection speed of the model and increases the model's focus on small objects in UAV images.

Through experiments we also found that when the input image size is 300×300 , the detection accuracies of many good object detection models show a substantial decrease [50–52], which is due to the fact that images with small sizes contain less feature information, and much of the feature information is gradually lost in the process of convolution. Many object detection models with high object detection accuracy often have large input image sizes, and the computational burden of large input image sizes increases, which is not conducive to the application of object detection models on UAVs. Real-time performance is one of the necessary conditions for object detection models to be applied on UAVs. Taking a 300×300 image as the input to the model can significantly reduce the computational overhead and increase the detection speed, which is conducive to the deployment of object detection models on UAVs and achieving real-time detection.

In the subsequent research on UAV image detection, for the problem of low detection accuracy of very small objects in UAV images, data augmentation and image super-resolution are considered to enhance the information contained in the image, which makes the backbone network better to extract its features. Meanwhile, the lightweighting of the model is further investigated, and the model is mounted on a UAV to realize a real-time object detection model for intelligent processing of UAV images in practical applications.

6. Conclusions

In this study, the RSAD algorithm is proposed from the basic framework of SSD. Firstly, the feature information in the image is deeply mined by ResNet-50; secondly, selective fusion of deep and shallow information is carried out by SAFM module, which makes full use of the effective information in the image and improves the robustness of the model; lastly, the Focal Loss function is used to increase the model's focus on the difficult-to-detect objects, which improves the overall detection performance of the model. Through experiments, it is proved that the method proposed in this study has excellent detection performance for UAV images. Compared with the SSD object detection model, the mAP is improved by 10.6%, and the detection results for individual objects are all better than those of the SSD model. Additionally, in the input image size of 300×300 , the model obtained in this study shows optimal detection accuracy compared with other excellent target detection models, which is second only to YOLOv8 in detection speed.

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

Funding: National Natural Science Foundation of China. Research on the mechanism of high efficiency and low-consumption strip cutting and kneading/crushing 52265035; Collaborative Intelligence-based Multi-mobile Robot Collaborative Handling System 2021GG0218.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to acknowledge the funding of National Natural Science Foundation of China and Collaborative Intelligence-based Multi-mobile Robot Collaborative Handling System project and thank our team members for their help and encouragement.

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

References

1. Lan, Y.; Huang, Z.; Deng, X.; Zhu, Z.; Huang, H.; Zheng, Z.; Lian, B.; Zeng, G.; Tong, Z. Comparison of machine learning methods for citrus greening detection on UAV multispectral images. *Comput. Electron. Agric.* **2020**, *171*, 105234. [\[CrossRef\]](#)
2. Liekai, C.; Martin, D.; Danxun, L. Airborne Image Velocimetry System and Its Application on River Surface Flow Field Measurement. *J. Basic Sci. Eng.* **2020**, *28*, 1271–1280.
3. Jiang, B.; Qu, R.; Li, Y.; Li, C. Object detection in UAV imagery based on deep learning: Review. *Acta Aeronaut. Astronaut. Sin.* **2021**, *42*, 137–151.
4. Tong, K.; Wu, Y. Deep learning-based detection from the perspective of tiny objects: A survey. *Image Vis. Comput.* **2022**, *123*, 104471.
5. Li, X.; Song, S.; Yin, X. Real-time Vehicle Detection Technology for UAV Imagery Based on Target Spatial Distribution Features. *China J. Highw.* **2022**, *35*, 193–204.
6. Liu, W.; Quijano, K.; Crawford, M.M. YOLOv5-Tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 is based on transfer learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8085–8094. [\[CrossRef\]](#)
7. Wu, X.; Li, W.; Hong, D.; Tao, R.; Du, Q. Deep Learning for Unmanned Aerial Vehicle-Based Object Detection and Tracking: A survey. *IEEE Geosci. Remote Sens. Mag.* **2021**, *10*, 91–124. [\[CrossRef\]](#)
8. Apolo-Apolo, O.E.; Martinez-Guanter, J.; Egea, G.; Raja, P.; Pérez-Ruiz, M. Deep learning techniques for estimation of the yield 556 and size of citrus fruits using a UAV. *Eur. J. Agron.* **2020**, *115*, 126030. [\[CrossRef\]](#)
9. Yang, J.; Yang, H.; Wang, F.; Chen, X. A modified YOLOv5 for object detection in UAV-captured scenarios. In Proceedings of the 2022 IEEE International Conference on Networking, Sensing and Control (ICNSC), Shanghai, China, 15–18 December 2022; pp. 1–6.
10. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common objects in context. In Proceedings of the Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part V13. Springer International Publishing: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
11. Saeed, Z.; Yousaf, M.H.; Ahmed, R.; Velastin, S.A.; Viriri, S. On-Board Small-Scale Object Detection for Unmanned Aerial Vehicles (UAVs). *Drones* **2023**, *7*, 310. [\[CrossRef\]](#)
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. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [\[CrossRef\]](#) [\[PubMed\]](#)
14. 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.
15. 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 Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; pp. 21–37.
16. 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.
17. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
18. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
19. Zhang, R.; Shao, Z.; Huang, X.; Wang, J.; Li, D. Object Detection in UAV Images via Global Density Fused Convolutional Network. *Remote Sens.* **2020**, *12*, 3140. [\[CrossRef\]](#)
20. Junyan, M.; Yanan, C. MFE-YOLOX: Dense small target detection algorithm under UAV aerial photography. *J. Chongqing Univ. Posts Telecommun. (Nat. Sci. Ed.)* **2023**, 1–8.
21. Gangyi, T.; Jianran, L.; Wenyuan, Y. A dual neural network for object detection in UAV images. *Neurocomputing.* **2021**, *443*, 292–301.
22. Liu, Y.; Cen, C.; Che, Y.; Ke, R.; Ma, Y.; Ma, Y. Detection of Maize Tassels from UAV RGB Imagery with Faster R-CNN. *Remote Sens.* **2020**, *12*, 338. [\[CrossRef\]](#)
23. Ding, J.; Xue, N.; Long, Y.; Xia, G.; Lu, Q. Learning roi transformer for oriented object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2849–2858.
24. Xu, Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.; Bai, X. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1452–1459. [\[CrossRef\]](#)
25. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. Yolov6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.

26. Jawaharlalnehru, A.; Sambandham, T.; Sekar, V.; Ravikumar, D.; Loganathan, V.; Kannadasan, R.; Khan, A.A.; Wechtaisong, C.; Haq, M.A.; Alhussen, A.; et al. Target Object Detection from Unmanned Aerial Vehicle (UAV) Images Based on Improved YOLO Algorithm. *Electronics* **2022**, *11*, 2343. [\[CrossRef\]](#)
27. Elhagry, A.; Dai, H.; El Saddik, A.; Gueaieb, W.; De Masi, G. CEAFOD: Cross-Ensemble Attention-based Feature Fusion Architecture Towards a Robust and Real-time UAV-based Object Detection in Complex Scenarios. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023; pp. 4865–4872.
28. Maktab Dar Oghaz, M.; Razaak, M.; Remagnino, P. Enhanced Single Shot Small Object Detector for Aerial Imagery Using Super-Resolution, Feature Fusion and Deconvolution. *Sensors* **2022**, *22*, 4339. [\[CrossRef\]](#)
29. Yundong, L.I.; Han, D.; Hongguang, L.I.; Xueyan, Z.; Baochang, Z.; Zhifeng, X. Multi-block SSD based on small object detection for UAV railway scene surveillance. *Chin. J. Aeronaut.* **2020**, *33*, 1747–1755.
30. Bowei, L.; Ming, H.; Qing, L.; Wenlong, X. Improved SSD Domestic Garbage Detection Algorithm. *Mach. Des. Manufacture.* **2023**, *9*, 157–162.
31. Liu, X.; Li, Y.; Shuang, F.; Gao, F.; Zhou, X.; Chen, X. ISSD: Improved SSD for Insulator and Spacer Online Detection Based on UAV System. *Sensors* **2020**, *20*, 6961. [\[CrossRef\]](#)
32. Zhai, S.; Shang, D.; Wang, S.; Dong, S. DF-SSD: An Improved SSD Object Detection Algorithm Based on DenseNet and Feature Fusion. *IEEE Access* **2020**, *8*, 24344–24357. [\[CrossRef\]](#)
33. Leng, J.; Liu, Y. An enhanced SSD with feature fusion and visual reasoning for object detection. *Neural. Comput. Appl.* **2019**, *31*, 6549–6558. [\[CrossRef\]](#)
34. VisDrone. Available online: <https://github.com/VisDrone/VisDrone-Dataset> (accessed on 16 May 2021).
35. Jian, J.; Liu, L.; Zhang, Y.; Xu, K.; Yang, J. Optical Remote Sensing Ship Recognition and Classification Based on Improved YOLOv5. *Remote Sens.* **2023**, *15*, 4319. [\[CrossRef\]](#)
36. Canziani, A.; Paszke, A.; Culurciello, E. An Analysis of Deep Neural Network Models for Practical Applications. *arXiv* **2016**, arXiv:1605.07678.
37. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
38. Adam, P.; Abhishek, C.; Sangpil, K.; Eugenio, C. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv* **2016**, arXiv:1606.02147.
39. Min, L.; Qiang, C.; Shuicheng, Y. Network in network. *arXiv* **2013**, arXiv:1312.4400.
40. Christian, S.; Wei, L.; Yangqing, J.; Pierre, S.; Scott, R.; Dragomir, A.; Dumitru, E.; Vincent, V.; Andrew, R. Going deeper with convolutions. *arXiv* **2014**, arXiv:1409.4842.
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *arXiv* **2015**, arXiv:1512.03385.
42. Karen, S.; Andrew, Z. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
43. Li, X.; Lv, C.; Wang, W.; Li, G.; Yang, L.; Yang, J. Generalized focal loss: Towards efficient representation learning for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 3139–3153. [\[CrossRef\]](#)
44. Chen, Y.; Wang, H.; Li, W.; Sakaridis, C.; Dai, D.; Van Gool, L. Scale-aware domain adaptive faster r-cnn. *Int. J. Comput. Vis.* **2021**, *129*, 2223–2243. [\[CrossRef\]](#)
45. 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.
46. 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 (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
47. 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.
48. Zhao, J.; Zhang, X.; Yan, J.; Qiu, X.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W. A Wheat Spike Detection Method in UAV Images Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 3095. [\[CrossRef\]](#)
49. Li, Y.; Fan, Q.; Huang, H.; Han, Z.; Gu, Q. A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition. *Drones* **2023**, *7*, 304. [\[CrossRef\]](#)
50. Alibaba, B.M.; Ozer, S. SyNet: An ensemble network for object detection in UAV images. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 10227–10234.
51. Cao, Y.; He, Z.; Wang, L.; Wang, W.; Yuan, Y.; Zhang, D.; Zhang, J.; Zhu, P.; Van Gool, L.; Han, J.; et al. VisDrone-DET2021: The vision meets drone object detection challenge results. In Proceedings of the IEEE/CVF International Conference on Computer vision, Montreal, QC, Canada, 10–17 October 2021; pp. 2847–2854.
52. Du, D.; Zhu, P.; Wen, L.; Bian, X.; Lin, H.; Hu, Q.; Peng, T.; Zheng, J.; Wang, X.; Zhang, Y.; et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 213–226.

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