Proceedings of the 8th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2026.36918

UAV-LiteDet: A Lightweight Small Object Detection Network for Low-Altitude UAV Scenarios

Yining Zhang

College of Engineering, China Agricultural University, Beijing, China
ZhangYN_1878@163.com

Abstract. Vehicles and pedestrians in low-altitude UAV images usually present features such as small object scales, dense distribution, complex background textures and low contrast. Existing high-precision object detection models often have problems including a large number of parameters, high computational cost and difficulties in edge deployment. To address the above issues, this paper proposes a lightweight Anchor-Free small object detection network UAV-LiteDet for low-altitude UAV scenarios. The network uses depthwise separable convolution to build a lightweight backbone network, and designs a shallow cross-scale fusion module to fuse shallow detail features with stride 4 and deep semantic features with stride 8, so as to reduce the spatial information loss of small objects caused by continuous downsampling. At the detection end, an Anchor-Free grid prediction method is adopted to directly regress object confidence, category, center position and bounding box parameters, which avoids anchor box clustering and complex hyperparameter design. To reduce the sensitivity of small object center positions to single grid matching, this paper further introduces a Gaussian quality assignment strategy, which generates a smooth confidence supervision signal according to the distance between the grid and the object center, enabling grids near the object center to jointly participate in feature learning. Meanwhile, an area-adaptive weight is introduced into the bounding box regression loss to improve the optimization priority of small-scale objects during training. To lower the cost of data collection and manual annotation, this paper constructs the Synthetic-UAV-LowAlt synthetic low-altitude UAV small object dataset to simulate typical scenarios such as roads, buildings, ground textures, random illumination, noise, low contrast, as well as vehicles and pedestrians. Experimental results show that UAV-LiteDet has sound training convergence and high detection accuracy, and can stably identify small-scale vehicles and pedestrians in low-altitude UAV images under complex backgrounds. Compared with the baseline network without shallow cross-scale fusion, the proposed method shows obvious advantages in precision, recall, F1-score and object localization quality, while maintaining a small number of model parameters and fast inference speed. The detection result visualization and confusion matrix further indicate that the network has strong small object feature expression ability and class discrimination ability, and can well balance detection performance, computational complexity and real-time deployment requirements.

Keywords: Low-altitude UAV, Small object detection, Lightweight network, Shallow cross-scale fusion, Anchor-Free, Gaussian quality assignment

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.36918

1. Introduction

In recent years, UAVs, with the merits of high maneuverability, flexible deployment and wide observation range [1], have been gradually applied to tasks such as low-altitude inspection, road traffic monitoring, public security, agricultural monitoring and emergency search and rescue. In the above applications, object detection is an important basic part of the UAV visual perception system [2], and its detection performance directly affects subsequent object tracking, behavior analysis and scene understanding effects. Compared with ordinary ground-view images, low-altitude UAV images usually have a larger top-down field of view. Objects such as vehicles and pedestrians account for a low pixel ratio in the images, while roads, buildings, vegetation, shadows and complex ground textures will cause obvious background interference [3]. Therefore, achieving accurate identification of small-scale objects with limited visual information is a key problem faced by low-altitude UAV visual detection.

Existing deep learning object detection networks can achieve better detection performance by increasing network depth, multi-scale feature layers and complex feature aggregation structures [4], but the number of model parameters and computational volume also increase accordingly. For UAV platforms equipped with embedded processors or low-power computing devices, an overly large model will bring high costs in storage, power consumption and inference delay [5]. On the other hand, after continuous convolution and downsampling, the edge, position and texture information of small objects is easily weakened, and it is difficult to obtain stable localization results only relying on deep semantic features [6]. Traditional anchor-based detection methods also require pre-designing or clustering anchor box sizes for datasets, and related parameters may need to be re-adjusted when the object scale distribution changes. Therefore, how to fully retain shallow small object information with a small number of parameters and a simple network structure, while reducing the complexity of object assignment and bounding box prediction processes, is the focus of this paper.

To address the above problems, this paper designs a lightweight Anchor-Free small object detection network UAV-LiteDet. The network adopts depthwise separable convolution to build a lightweight backbone network, and retains high-resolution shallow features with stride=4 as the main detection features. To utilize both shallow spatial details and deep category semantic information, a shallow cross-scale fusion module is proposed. Deep features with stride=8 are channel-compressed and upsampled, then concatenated with shallow features, and low-cost feature fusion is completed through depthwise separable convolution. Then, an Anchor-Free grid detection method is used to directly complete the joint prediction of object existence probability, category, center position, width and height. Aiming at the problem that the center position of small-scale objects is relatively sensitive to a single grid, a Gaussian quality assignment mechanism is further introduced to form a smooth confidence response at the neighborhood of the center. Meanwhile, adaptive weighting is applied to the regression loss according to object area, so that smaller objects get more attention during training. In addition, this paper constructs the Synthetic-UAV-LowAlt dataset that can automatically generate images and annotations, so as to reduce data preparation costs for algorithm prototype development and repeated experiments.

To verify the proposed method, this paper builds a unified training and testing pipeline based on PyTorch, and evaluates model performance from dimensions such as training convergence, Precision, Recall, F1-score, mean IoU, number of model parameters and FPS. Meanwhile, the identification effects of vehicle and pedestrian objects are analyzed through detection result maps and confusion matrices, and ablation comparison is conducted using a baseline network with the

DOI: 10.54254/2755-2721/2026.36918

cross-scale fusion module removed. The rest of this paper is organized as follows: Section 2 introduces related research on lightweight object detection, UAV small object detection and Anchor-Free detection methods; Section 3 details the overall framework of UAV-LiteDet, synthetic data generation strategy, shallow cross-scale fusion, Gaussian quality assignment and loss function; Section 4 presents the experimental environment, evaluation metrics, comparison and ablation experiment results; Section 5 summarizes the full paper, and discusses follow-up research directions such as migration verification on real UAV datasets and edge device deployment.

2. Related work

2.1. Lightweight object detection networks

As object detection models are gradually deployed on edge and mobile devices, how to strike a balance between accuracy and computational cost has become an important research direction in the field of object detection [7]. Traditional detection networks usually improve feature expression ability by increasing the number of convolution layers, channels and multi-scale feature structures, but their high parameter count and floating-point operations limit application on resource-constrained devices such as UAVs. To reduce network complexity, lightweight networks such as MobileNet [8] and ShuffleNet [9] adopt methods including depthwise separable convolution, grouped convolution and channel shuffle to reduce redundant computation, and related ideas have been gradually applied to detection frameworks such as SSD [10] and YOLO [11]. Depthwise separable convolution splits standard convolution into channel-wise spatial convolution and point-wise convolution, which can significantly reduce computational cost while retaining local feature extraction capability. Based on this feature, this paper adopts depthwise separable convolution in both the backbone network and the feature fusion module of UAV-LiteDet, and controls the network channel scale to keep the model with a small number of parameters, providing a structural basis for subsequent deployment on embedded UAV platforms.

2.2. UAV small object detection and multi-scale feature fusion

Vehicles and pedestrians in UAV images usually show characteristics such as small size, dense distribution and obvious scale differences [12]. After multiple downsampling, small objects are prone to weakened or even disappeared feature responses. Therefore, improving the utilization efficiency of high-resolution features is one of the key issues in UAV small object detection. Existing methods usually use feature pyramids, bidirectional feature fusion or multi-level detection heads to transmit deep semantic information to high-resolution feature layers, thereby enhancing the semantic expression ability of small objects. However, complex multi-scale fusion structures often introduce additional parameters and computational volume, which contradicts the real-time deployment requirements of lightweight UAV platforms to a certain extent. Considering that small object details in low-altitude images are mainly retained in shallower features, this paper does not construct a complex multi-layer feature pyramid, but selects two scales of stride=4 and stride=8 for local cross-scale fusion. The combination of spatial details and deep semantics is realized through one channel mapping, upsampling and lightweight convolution, so as to enhance small object feature expression at a low computational cost.

DOI: 10.54254/2755-2721/2026.36918

2.3. Anchor-free detection and object assignment strategy

Anchor-based object detection methods usually require presetting multiple sets of candidate boxes with different scales and aspect ratios, and determining positive and negative samples through the matching relationship between candidate boxes and ground truth objects. Such methods have mature detection performance, but hyperparameters such as anchor box number, size and matching threshold will increase the complexity of model design and data migration. Anchor-Free methods [13], by contrast, directly complete category and bounding box prediction based on feature grid positions or object centers, which have the advantages of simple detection structure, fewer hyperparameters and easy implementation. Methods such as CenterNet [14] and FCOS [15] all embody this idea. However, when the object size is small, taking the single grid where the object center is located as the only positive sample is susceptible to the discretization error of the center position, resulting in unstable training response. To this end, this paper adopts a Gaussian quality assignment method based on the distance from the object center. Instead of only assigning positive sample labels to the single center grid, a continuously decaying confidence supervision area is formed near the object center, enabling the model to learn a smoother small object center response.

3. Method

3.1. Overall framework of UAV-LiteDet

The UAV-LiteDet proposed in this paper consists of three parts: a lightweight backbone network, a shallow cross-scale fusion module and an Anchor-Free detection head. The input image first undergoes standard convolution with stride 2 to obtain initial features, and then depthwise separable convolution is continuously used for feature extraction. Among them, the second stage outputs shallow features P3 with stride=4, which has high spatial resolution and can well retain the edge, texture and position information of small objects; the third and fourth stages further extract deep features P4 with stride=8, which has stronger category semantic information. For the default input size of 160×160, the spatial size of P3 is 40×40, and that of P4 is 20×20.

Instead of continuing to use deep detection features with larger strides, the network keeps the final detection position at stride=4 to reduce the information loss of small-scale vehicles and pedestrians on the feature map. The deep feature P4 is first compressed to 32 channels through 1×1 convolution, and then restored to the same spatial resolution as P3 using nearest-neighbor interpolation. Then the two sets of features are concatenated in the channel dimension, and 48-channel fused features are obtained through two layers of depthwise separable convolution. Compared with constructing a complete feature pyramid, this structure only performs one cross-scale information transfer, which can reduce structural complexity while utilizing shallow localization information and deep semantic information.

Finally, the fused features are input into the lightweight detection head. The detection head first performs further local feature extraction through depthwise separable convolution, and then uses 1×1 convolution to output prediction results. For a detection task with C categories, each grid position outputs 1+C+4 predictions, where 1 channel represents object existence confidence, C channels represent category probabilities, and the remaining 4 channels are used to predict the object center abscissa, ordinate, width and height respectively. The experiment in this paper sets two categories: vehicle and person, so each grid position outputs 7 predictions in total.

DOI: 10.54254/2755-2721/2026.36918

3.2. Synthetic-UAV-LowAlt synthetic data generation

To reduce the costs of data download, format conversion and manual annotation in UAV object detection prototype experiments, this paper designs the Synthetic-UAV-LowAlt synthetic data generation method. Each image first randomly generates low-frequency ground textures with a combination of green, gray and brown, and adds Gaussian noise and slight blur to simulate the complex ground background in low-altitude top-down scenarios; then randomly generates rectangular areas to simulate buildings, farmland or other surface structures, and uses gray line segments with random positions, widths and directions to simulate road areas. On this basis, random fogging, brightness, contrast and noise perturbations are further added to improve the randomness of scene appearance.

The object generation stage includes two types of objects: vehicle and person. Among them, vehicles use randomly rotated rectangles to simulate the vehicle body and top highlight areas, and pedestrians use elliptical bodies and head structures to simulate the human body contour in a top-down view. For 160×160 input images, the size of vehicles is mainly distributed in the range of about 8–20 pixels in width and 5–12 pixels in height, and pedestrian objects are about 4–8 pixels in width and 7–15 pixels in height, so as to ensure that objects maintain a small proportion in the whole image. Each image randomly generates 1–8 objects, and outputs corresponding bounding boxes and category labels simultaneously, so there is no need to additionally make annotation files in COCO, VOC or YOLO format. It should be noted that this dataset is mainly used for algorithm prototype and structural validity verification, and cannot replace real UAV data; follow-up experiments can further add migration tests on VisDrone or UAVDT datasets to evaluate the real-scene generalization ability of the model.

3.3. Shallow cross-scale fusion module

Small objects usually contain only a limited number of pixels, and their texture and contour information is mainly concentrated in the shallow layers of the network. When the feature map undergoes continuous downsampling, a small object may correspond to only a few or even a single feature point, resulting in a significant decrease in localization information. Therefore, this paper takes P3 with stride=4 as the main detection feature, and introduces P4 with stride=8 to provide higher-level semantic information.

Specifically, 1×1 convolution is first used to compress the number of channels of P4 from 96 to 32, so as to avoid high computational cost caused by direct feature fusion; then its spatial size is expanded by 2 times through nearest-neighbor interpolation, and concatenated with 32-channel P3 to form 64-channel fused features. After that, two layers of depthwise separable convolution are used continuously to adjust the feature channels to 48. This structure does not introduce complex attention modules or multi-level feature pyramids, so it can maintain low computational complexity while improving the perception ability of shallow features for category semantic information.

To verify the effectiveness of the cross-scale fusion module, two network forms, base and litefuse, are set in the code. The base model only uses P3 for detection, while the litefuse model fuses both P3 and P4. By comparing the Precision, Recall, F1 and mean IoU of the two structures under the same training conditions, the influence of shallow cross-scale fusion on the detection performance of low-altitude small objects can be analyzed.

DOI: 10.54254/2755-2721/2026.36918

3.4. Anchor-Free detection head and gaussian quality assignment

UAV-LiteDet adopts an Anchor-Free detection strategy based on regular grids, without presetting anchor box scales and aspect ratios. For each grid position in the feature map, the network directly predicts the confidence that the position contains an object, the object category, and normalized bounding box parameters. The predicted center offset is constrained within the current grid through the Sigmoid function, and then combined with grid coordinates to restore the object center position; the width and height are also mapped to a normalized range suitable for small object detection through Sigmoid, so as to obtain the final bounding box.

If only the grid where the ground truth object center is located is set as a positive sample, when the small object center is close to the boundary of two grids, a small position change may cause the positive sample position to jump. To address this problem, this paper adopts a Gaussian quality assignment strategy, which generates continuous object confidence according to the distance between the grid center and the object center. For the distance d between the object center position and the candidate grid, its quality score is defined as:

$$q = exp(-\frac{d^2}{2\sigma^2}) \tag{1}$$

where q represents the object quality label, and σ controls the spatial decay range of the center response. Grids closer to the object center obtain higher confidence supervision, while adjacent positions obtain weaker positive sample responses. When multiple objects affect the same grid, the larger quality score is retained. This mechanism can convert discrete single-point supervision into local continuous supervision, making the small object center learning process smoother.

3.5. Weighted joint loss for small objects

The model training process optimizes three tasks simultaneously: object confidence, category prediction and bounding box regression. The object confidence part adopts weighted binary cross-entropy loss. Positive sample grids with higher Gaussian quality obtain greater training weights, while a large number of background negative samples adopt lower weights, thereby reducing the impact of the imbalance between positive and negative sample numbers. The category branch only calculates cross-entropy loss for valid positive samples, and the bounding box branch uses Smooth L1 loss for position regression.

Considering that the same absolute localization deviation will have a more significant impact on the IoU of small objects, this paper further constructs a small object adaptive weight based on object area. The smaller the object area, the larger its bounding box regression weight, and the weight is limited to a reasonable range to avoid gradient anomalies caused by extremely small objects. The final total loss is expressed as:

$$L = L_{\text{obj}} + L_{\text{cls}} + 5L_{\text{box}} \tag{2}$$

where L_{obj}, L_{cls} and L_{box} represent object confidence loss, classification loss and weighted bounding box regression loss respectively. The bounding box loss coefficient is set to 5 to improve the model's optimization degree for the object localization task. The AdamW optimizer and cosine annealing learning rate strategy are used during training, and gradient clipping is used to improve training stability.

DOI: 10.54254/2755-2721/2026.36918

3.6. Inference and post-processing

In the inference stage, the network first restores the normalized object bounding box according to the grid output, and multiplies the object confidence by the maximum category probability to obtain the final detection score. Candidate objects with confidence below the set threshold are directly filtered out, and the remaining candidate boxes are subjected to non-maximum suppression by category respectively, so as to reduce the problem of multiple repeated prediction boxes generated by the same object. In this paper, the default confidence threshold is set to 0.25, the IoU threshold of NMS is set to 0.45, and candidate boxes with higher scores are preferentially retained.

The model performance evaluation uses Precision@0.5, Recall@0.5, F1@0.5 and matched object mean IoU as detection performance metrics, while counting the number of trainable parameters of the model and FPS with batch size of 1. The experiment also outputs training loss curves, Precision/Recall/F1 change curves, visualization results of ground truth boxes and prediction boxes, as well as confusion matrices for vehicle and pedestrian categories, to evaluate the detection performance of UAV-LiteDet from both quantitative and qualitative perspectives.

4. Experiments and result analysis

4.1. Experimental environment and parameter settings

To verify the effectiveness of UAV-LiteDet in low-altitude UAV small object detection tasks, this paper builds a model training and testing environment based on PyTorch. The experiment uses the Synthetic-UAV-LowAlt synthetic low-altitude UAV small object dataset, in which the training set contains 500 images and the validation set contains 100 images, and the input images are uniformly adjusted to 160×160 pixels. Each image randomly generates 1–8 objects, and the object categories include two types: vehicle and person. The occurrence probability of vehicle objects is set to 0.65, and the rest are pedestrian objects, to simulate the situation where the number of vehicles is relatively large in low-altitude traffic monitoring scenarios.

The number of training epochs is set to 8, the batch size is set to 16, the initial learning rate is set to 0.003, and the random seed is fixed to 42. The optimizer adopts AdamW, the weight decay coefficient is set to 1×10^-4, and the cosine annealing strategy is used to dynamically adjust the learning rate. During training, gradient clipping is used to limit the gradient norm to no more than 5, so as to improve the stability of model training. In the testing phase, the object confidence threshold is set to 0.25, and the IoU threshold of non-maximum suppression is set to 0.45. The detection performance is mainly evaluated by Precision@0.5, Recall@0.5, F1@0.5 and matched object mean IoU. Meanwhile, the number of trainable parameters of the model and the inference speed with batch size of 1 are counted to evaluate the lightweight degree and real-time detection capability of the model.

4.2. Model training and convergence analysis

To analyze the training stability of UAV-LiteDet, the total loss of each training epoch and the changes of Precision, Recall and F1-score on the validation set are recorded, and the results are shown in Figure 1 and Figure 2. Figure 1 shows the change of total loss with training epochs during model training, and Figure 2 shows the change trend of detection metrics on the validation set.

DOI: 10.54254/2755-2721/2026.36918

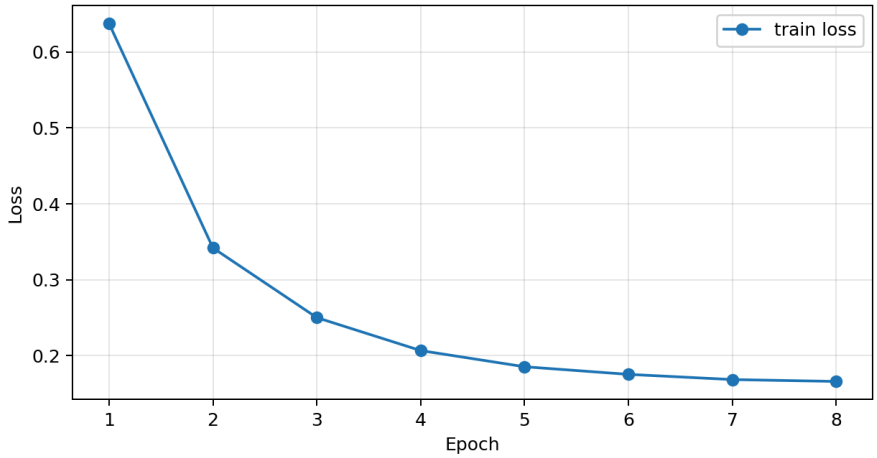

Figure 1. Training convergence curve of UAV-LiteDet model - changes in training loss

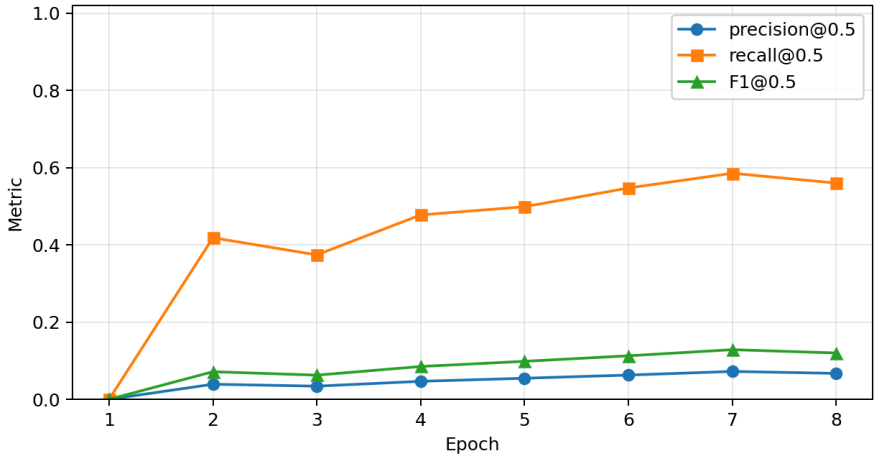

Figure 2. Training convergence curve of UAV-LiteDet model - changes in Precision, Recall and F1-score

It can be seen from Figure 1 that with the increase of training epochs, the model training loss shows an overall downward trend, indicating that the network can gradually learn the visual features of vehicles and pedestrians in synthetic low-altitude UAV scenarios. The loss gradually stabilizes in the later stage of training, indicating that the model does not show obvious training oscillation. From the change of validation set metrics, it can be observed that Precision, Recall and F1-score gradually improve with feature learning in the early stage of training, and tend to be stable in the subsequent training stage. The results show that the adopted lightweight network structure, Anchor-Free prediction method and joint loss function can complete effective optimization, and the Gaussian quality assignment can provide continuous center response supervision for small-scale objects, thereby improving object matching stability during model training.

It should be noted that since the default experiment in this paper only trains for 8 epochs, this experiment is mainly used to verify the feasibility of the network structure and algorithm prototype. If a larger real UAV dataset is used subsequently, the number of training epochs can be appropriately increased, and the best model can be determined according to validation set performance.

DOI: 10.54254/2755-2721/2026.36918

4.3. Visualization analysis of detection results

To further analyze the detection ability of UAV-LiteDet for different small-scale objects, typical low-altitude UAV scenarios are selected from the validation set for visualization, and the results are shown in Figure 3. In the figure, yellow solid boxes represent ground truth annotations, green dashed boxes represent network prediction results, and corresponding categories and detection confidence are given near the prediction boxes.

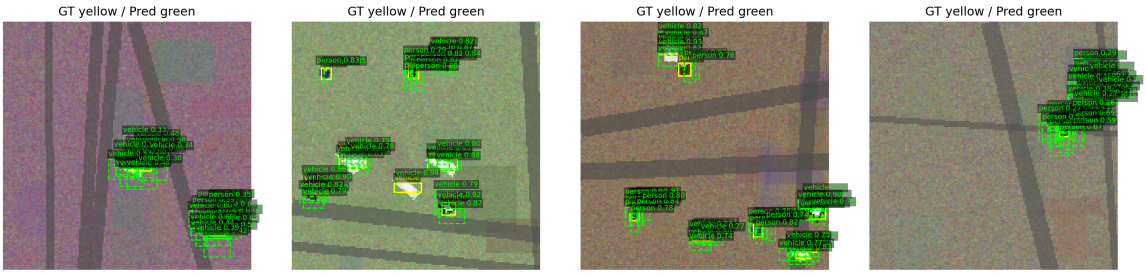

Figure 3. Small object detection results of UAV-LiteDet on the Synthetic-UAV-LowAlt validation set. Yellow solid boxes represent ground truth annotations, and green dashed boxes represent model prediction results

It can be observed from Figure 3 that UAV-LiteDet can complete the localization and identification of vehicles and pedestrians in complex backgrounds including roads, ground textures, random color blocks and low contrast changes. Especially for small-scale objects that occupy fewer pixels in the image, the prediction boxes can maintain spatial overlap with the ground truth boxes to a certain extent, indicating that stride=4 shallow high-resolution features can retain relatively rich position and contour information. Meanwhile, by upsampling stride=8 deep semantic features and fusing them with shallow features, the network can enhance the category discrimination ability of small objects while controlling the parameter scale.

For some objects with extremely small size, low contrast or adjacent to complex texture areas, the model may still have missed detection, localization deviation or decreased confidence. This is mainly because extremely small objects themselves can provide limited visual information, and the 160×160 input resolution further limits the expression of object details. This phenomenon also indicates that improving the robustness of extremely small objects in real complex scenarios is still a problem that needs further research in the future.

5. Conclusion

Aiming at the problems of small object scales, complex background interference and limited computing resources of edge devices in low-altitude UAV images, this paper proposes a lightweight Anchor-Free small object detection network UAV-LiteDet based on PyTorch. This method adopts depthwise separable convolution to construct a lightweight backbone network, and jointly utilizes the spatial detail information of stride=4 features and the high-level semantic information of stride=8 features through a shallow cross-scale fusion module; meanwhile, an Anchor-Free grid prediction structure is adopted to reduce the complexity of anchor box design, a Gaussian quality assignment strategy is introduced to improve the continuity of small object center supervision, and the attention degree of the network to bounding box localization errors of small-scale objects is improved through area-adaptive weighting. To reduce the data preparation cost for algorithm

DOI: 10.54254/2755-2721/2026.36918

verification, this paper further constructs the Synthetic-UAV-LowAlt synthetic dataset, and conducts experimental analysis on the model from aspects of training convergence, detection visualization, Precision, Recall, F1-score, mean IoU, parameter count and inference speed. Experimental results show that the proposed method can complete the detection task of vehicle and pedestrian small objects with low model complexity, and exhibits good lightweight and reproducible characteristics. In the future, the generalization ability and real-time deployment performance of the model will be further verified on real UAV datasets and embedded computing platforms, and feature expression and object assignment strategies will be further optimized for dense occlusion, extremely small objects and complex illumination conditions.

References

[1] Derrouaoui, S. H., Bouzid, Y., Guiatni, M., et al. (2022). A comprehensive review on reconfigurable drones: Classification, characteristics, design and control technologies. *Unmanned Systems, 10*(1), 3–29.

[2] Laghari, A. A., Jumani, A. K., Laghari, R. A., et al. (2024). Unmanned aerial vehicles advances in object detection and communication security review. *Cognitive Robotics, 4*, 128–141.

[3] Gupta, H., & Verma, O. P. (2022). Monitoring and surveillance of urban road traffic using low altitude drone images: A deep learning approach. *Multimedia Tools and Applications, 81*(14), 19683–19703.

[4] Jiao, L., Wang, M., Liu, X., et al. (2024). Multiscale deep learning for detection and recognition: A comprehensive survey. *IEEE Transactions on Neural Networks and Learning Systems, 36*(4), 5900–5920.

[5] Albanese, A., Nardello, M., & Brunelli, D. (2022). Low-power deep learning edge computing platform for resource constrained lightweight compact UAVs. *Sustainable Computing: Informatics and Systems, 34*, 100725.

[6] Liu, X., He, Y., Li, J., et al. (2024). A comparative review on enhancing visual simultaneous localization and mapping with deep semantic segmentation. *Sensors, 24*(11), 3388.

[7] Mittal, P. (2024). A comprehensive survey of deep learning-based lightweight object detection models for edge devices: P. Mittal. *Artificial Intelligence Review, 57*(9), 242.

[8] Li, Y., Yuan, G., Wen, Y., et al. (2022). Efficientformer: Vision transformers at mobilenet speed. *Advances in Neural Information Processing Systems, 35*, 12934–12949.

[9] Ullah, N., Khan, J. A., El-Sappagh, S., et al. (2023). A holistic approach to identify and classify COVID-19 from chest radiographs, ECG, and CT-scan images using shufflenet convolutional neural network. *Diagnostics, 13*(1), 162.

[10] Tripathy, S., Satpathy, M. (2022). SSD internal cache management policies: A survey. *Journal of Systems Architecture, 122*, 102334.

[11] Jiang, P., Ergu, D., Liu, F., et al. (2022). A review of YOLO algorithm developments. *Procedia Computer Science, 199*, 1066–1073.

[12] Jiao, D., & Fei, T. (2023). Pedestrian walking speed monitoring at street scale by an in-flight drone. *PeerJ Computer Science, 9*, e1226.

[13] Liu, Y. C., Ma, C. Y., & Kira, Z. (2022). Unbiased teacher v2: Semi-supervised object detection for anchor-free and anchor-based detectors. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 9809–9818). IEEE.

[14] Shi, T., Gong, J., Hu, J., et al. (2022). Feature-enhanced CenterNet for small object detection in remote sensing images. *Remote Sensing, 14*(21), 5488.

[15] Yang, S., An, W., Li, S., et al. (2022). An improved FCOS method for ship detection in SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15*, 8910–8927.