Improvement and Implementation of UAV Target Detection Algorithm Based on YOLOv10

Chang Cai[1,a,*], Yanwen Wang[1], Pengfei Ma[1], Bo Fang[1], Minghao Cao[1]

[1]School of Electronic Information, Xijing University, Xi'an, China
[a]3445904421@qq.com
**Corresponding author*

Abstract: *With the wide application of drones in agriculture, surveying and mapping, power inspection, security and other fields, its safety management problems are becoming increasingly prominent. Traditional UAV detection methods such as radar, infrared, acoustic and radio frequency detection have limitations such as low accuracy, high cost or poor anti-interference in complex environments. Visual inspection technology has become a research hotspot due to its high resolution and good target recognition ability. Focusing on the UAV target detection task, based on the YOLOv10 network structure, this paper proposes an improved detection algorithm to improve the detection accuracy and small target perception ability. Experimental results show that the improved algorithm can significantly improve the recognition effect of multi-scale UAV targets while maintaining the detection speed, and has good engineering application value.*

Keywords: *UAV Detection, Deep Learning, Target Detection*

1. Introduction

With the wide application of UAV technology in military reconnaissance, agricultural plant protection, power inspection, logistics and transportation and other fields, how to achieve efficient identification and tracking of UAVs has become one of the current research hotspots[1]. Unmanned aerial vehicles (UAVs) are small in size, low in flight altitude, fast in motion, and often in complex and changeable backgrounds, which makes their detection tasks face great challenges in visual perception[2].

Traditional UAV detection methods include radar, infrared, acoustic, and radio frequency technologies, but these methods have problems such as high cost, complex deployment, and susceptibility to interference in the actual environment. In contrast, the image object detection technology based on computer vision can directly identify and locate UAVs from videos or images, which has stronger environmental adaptability and detection accuracy, especially in visual surveillance systems[3].

The UAV target detection task is essentially a target detection problem, and its core is to automatically identify the position and category of the UAV in the image through algorithms. In recent years, with the rapid development of deep learning technology, convolutional neural network (CNN)-based object detection algorithms have gradually replaced the traditional manual feature method and become the mainstream solution[4]. In particular, the YOLO (You Only Look Once) series algorithms have been widely studied and applied in the field of UAV detection due to their end-to-end, one-stage, and real-time characteristics.

The current research mainly focuses on improving the detection ability of the model for small-target UAVs, the robustness in complex backgrounds, the multi-scale feature extraction ability, and the balance between detection accuracy and speed[5]. Therefore, how to construct a lightweight, efficient, and practical UAV target detection model has become an important direction of this research.

2. Related research

2.1 Basic theory of the YOLOv10 model

YOLOv10 is an end-to-end single-stage object detection algorithm that simplifies object detection into a unified regression problem. The overall structure of the algorithm consists of four main components: Input, Backbone, Neck, and Head.The network structure is shown in Figure 1.

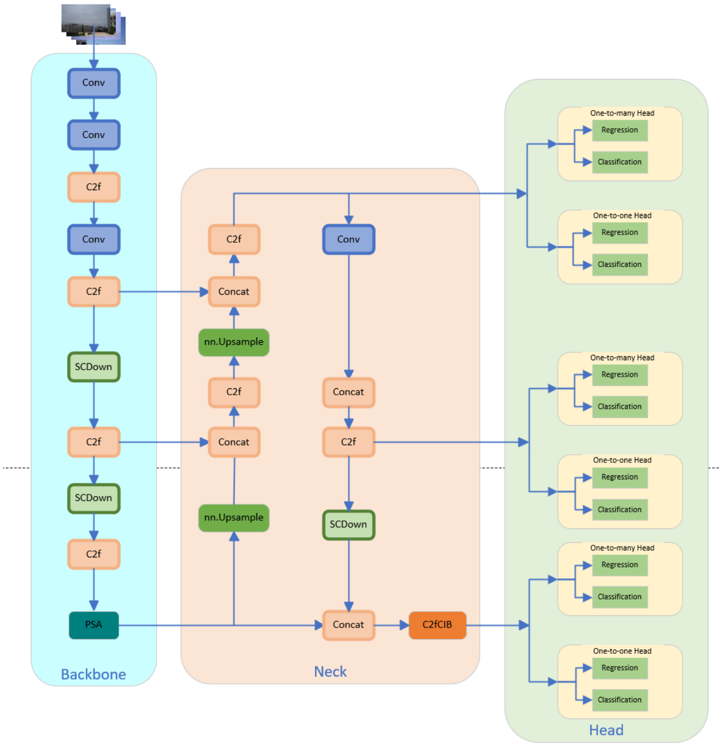

Figure 1: Yolov10 network structure.

The Input module preprocesses the input images, supporting multi-scale resolutions (320×320, 640×640, etc.), and performs operations such as resizing, normalization, and pixel value standardization to meet the network's input requirements and accelerate convergence[6].

The Backbone is responsible for feature extraction. Based on improvements to YOLOv8, it adopts more efficient modular convolutional structures such as C2f (Cross Stage Partial Fusion) and C3 modules, which reduce the number of parameters and computational complexity while maintaining strong feature representation capabilities. The image is progressively downsampled (e.g., from 320×320 down to 20×20), and a SPPF (Spatial Pyramid Pooling Fast) module is used to fuse multi-scale contextual information, enhancing the model's perception of targets at different scales.

The Neck module performs lateral transmission and vertical integration of multi-scale feature maps through operations such as upsampling, concatenation, and convolution. It effectively fuses deep semantic features with shallow spatial details to build a multi-scale semantic foundation for detection. It outputs three feature maps corresponding to large, medium, and small object detection, known as P5, P4, and P3 layers.

Finally, the Head performs prediction operations on each scale of the fused feature maps. YOLOv10 adopts a decoupled detection head design, which separates bounding box regression, objectness confidence, and classification tasks. Each scale's feature map outputs multiple bounding box predictions, each containing object confidence scores and class probabilities.

2.2 Share the convolutional feature pyramid module (FPSConv)

In the original YOLOv10 network structure, the SPPF (Spatial Pyramid Pooling - Fast) module is used at the end of the backbone network, and its main function is to expand the receptive field and aggregate multi-size context information through the maximum pooling operation of different sizes, so as to improve the semantic understanding ability of deep features to large targets[7]. However, there are still some shortcomings in this module, one is that it relies on a fixed pooling scale and lacks adaptability to local features, and the other is that the coverage effect of the receptive field of small targets is limited, which is easy to lead to the loss of feature details[8].

In order to solve the above problems, this paper uses a Feature Pyramid Shared Conv module to replace the original SPPF module. The module constructs a lightweight multi-scale feature fusion mechanism through a set of shared convolution kernels and dilated convolutions with multiple expansion rates, and has the multi-scale modeling idea of pyramids, and its module structure is shown in Figure 2.

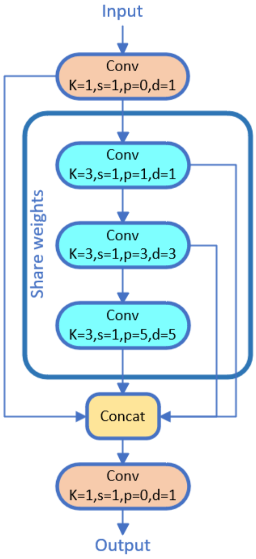

Figure 2: FPSConv structure.

2.3 Multi-Path Multi-Scale Feature Pyramid Network (MP-MSFPN)

In order to further improve the performance of the YOLOv10 network in multi-scale object detection tasks, especially to enhance its ability to detect small targets, we have made comprehensive and in-depth improvements to the neck structure of the original YOLOv10 network. The existing YOLOv10 neck structure has some problems in the process of multi-scale feature fusion, such as single fusion mode, insufficient feature expression ability, and low efficiency of scale information interaction, which seriously limits the detection performance of the network in complex scenes[9]. In order to solve these problems, a novel and efficient neck improvement structure was proposed by combining the fusion module and the multi-scale convolutional enhancement module (CSP_MSCB) to enhance the feature fusion and expression ability of the network. The module structure is shown in Figure 3.

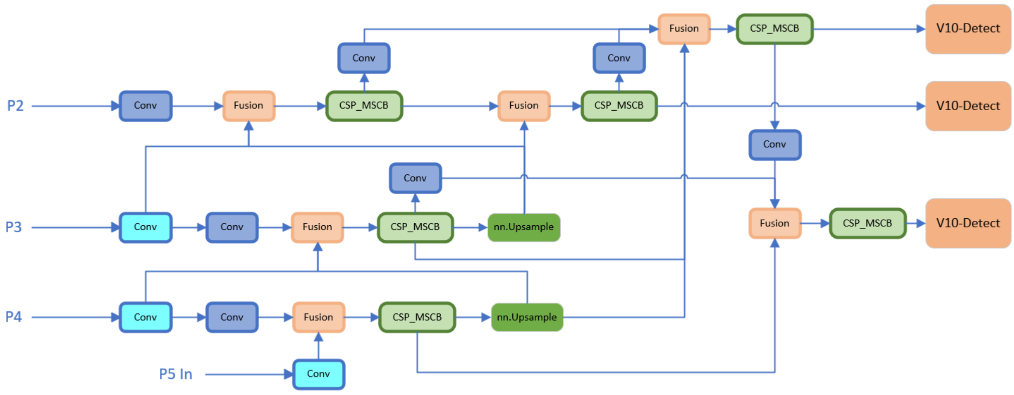

Figure 3: MP-MSFPN structure.

The structure designs a cross-layer connection path from P5 to P3, so that the feature maps of different levels can interact repeatedly in multiple directions through the upsampling, convolution and fusion modules. The features of each layer are enhanced by multi-scale convolution through the CSP_MSCB module, which further excavates the local and global information expression capabilities. This combination of multi-layer interconnection, multi-scale dynamic fusion, and multi-stage enhancement processing significantly improves the richness and adaptability of features in spatial, semantic, and scale dimensions, and the overall structure presents obvious multi-path feature fusion and enhanced interaction characteristics.

3. Experiments and analysis of results

3.1 Introduction to datasets

The datasets used in all experiments in this section are DUT Anti-UAV datasets[10].The DUT Anti-UAV dataset is divided into two subsets: object detection and target tracking, and this paper uses a drone detection subset, which includes a training set: 5,200 images, 5,243 targets, a validation set: 2,600 images, 2,621 targets, a test set: 2,200 images, 2,245 targets, and the image resolution varies from 160×240 to 3744×5616, providing multi-scale training and test data.

3.2 Experimental environment

The experiments were conducted using the PyTorch deep learning framework on the Anaconda platform. The server was equipped with an AMD EPYC 7532 32-Core Processor, configured with 8 cores and 30 GB of RAM. The GPU used was an NVIDIA GeForce RTX 4060 Ti with 16 GB of VRAM, and the CUDA version was 12.1. The programming language was Python 3.8, running on the Linux operating system. The development environment was based on PyCharm as the integrated development tool.

All input images are uniformly resized to 640×640 pixels. In order to improve the efficiency of model training, the stochastic gradient descent (SGD) optimizer was used, the initial learning rate was set to 0.01, the momentum coefficient was set to 0.937, the weight decay parameter was set to 0.0005, and the batch size was set to 16.

3.3 Ablation experiments

In order to verify the role of each improved module in improving the performance of YOLOv10 network, four sets of ablation experiments were designed, and FPSConv and MP-MSFPN neck structures were gradually introduced, and their effects on the average accuracy, detection accuracy, recall rate, computational size, model complexity and inference speed were evaluated, respectively, and the results are shown in Table 1.

Table 1: Ablation experimental results of the improved algorithm module on the DUT Anti-UAV dataset.

Group	FPSConv	MP-MSFPN	mAP@0.5(%)	P/%	R/%	Parameters	GFLOPS
1	×	×	85.5	91.7	77.1	2265363	6.5
2	√	×	86.8	92.7	78.7	2412819	6.5
3	×	√	86.9	92.6	78.5	1850084	5.9
4	√	√	87.1	94	76.6	1997540	6.5

The ablation experiments were conducted to assess the individual and combined impact of the FPSConv and MP-MSFPN neck modules on the YOLOv10 network's performance. From the results:

Baseline (Group 1): Without either FPSConv or MP-MSFPN, the baseline model achieves a mAP@0.5 of 85.5%, with 6.5 GFLOPs and 2.26M parameters.

FPSConv only (Group 2): Introducing only the FPSConv module improves the mAP to 86.8% and recall to 78.7%, showing that FPSConv effectively enhances feature representation and detection recall without increasing computational complexity.

MP-MSFPN only (Group 3): Using only MP-MSFPN achieves a slightly higher mAP of 86.9%, with the lowest parameter count (1.85M) and lowest GFLOPs (5.9) among all groups, demonstrating its superior efficiency in feature fusion and lightweight design.

Combined (Group 4): When both modules are used together, the best precision (94%) and highest mAP (87.1%) are achieved, indicating a strong complementary effect. Although the recall slightly drops to 76.6%, the overall detection performance improves, and the parameter count remains reasonably low (1.99M).

The combined use of FPSConv and MP-MSFPN achieves the best trade-off between accuracy, efficiency, and model complexity, confirming the effectiveness of the proposed improvements in enhancing multi-scale feature representation and detection precision.

3.4 Comparative experiments

In order to reasonably evaluate the detection performance of the model and illustrate the feasibility of the optimization method, the table is compared with YOLOv8n and other algorithms based on YOLOv10n network improvement, and the table is arranged according to the value of mAP@0.5 (%) from small to large. The comparison results are shown in Table 2.

Table 2: Comparative experiments of different algorithms on the DUT Anti-UAV dataset.

Group	model	parameters	GFLOPs	mAP@0.5(%)	P/%	R/%
1	YOLOv8n	3005843	8.1	84.1	92.2	75.9
2	YOLOv10n-AFGC	2270587	6.5	84.9	90.9	76.9
3	YOLOv10n-CBAM	2270775	6.5	85.2	92.6	76.3
4	YOLOv10n	2265363	6.5	85.5	91.7	77.1
5	YOLOv10n-CPCA	2283819	7.1	85.8	92.7	76.2
6	YOLOv10n-ASF	2304979	6.9	86.5	92.1	77.3
7	Ours	3054467	6.5	87.1	94	76.6

The comparative experiments demonstrate that the proposed improved YOLOv10 model achieves the best overall performance. It obtains the highest detection accuracy with a mAP@0.5 of 87.1%, outperforming YOLOv8n (84.1%) and other YOLOv10n variants. While maintaining a moderate parameter size (3.05M) and low computation cost (6.5 GFLOPs), it also achieves the highest precision (94%) and a competitive recall rate (76.6%). These results validate the effectiveness and efficiency of the proposed improvements.

3.5 Visual comparison of test results

Figures 4 through 7 show the results of the improved YOLOv10 (left) and YOLOv10 (right) in different environments, bright light, complex backgrounds, open scenes, and nighttime environments, respectively. The results show that the improved YOLOv10 is much better than the original network in these environments, with a 3% increase in confidence under the influence of bright light, an enhanced detection capability in complex backgrounds, and an 11% increase in open space and night conditions.

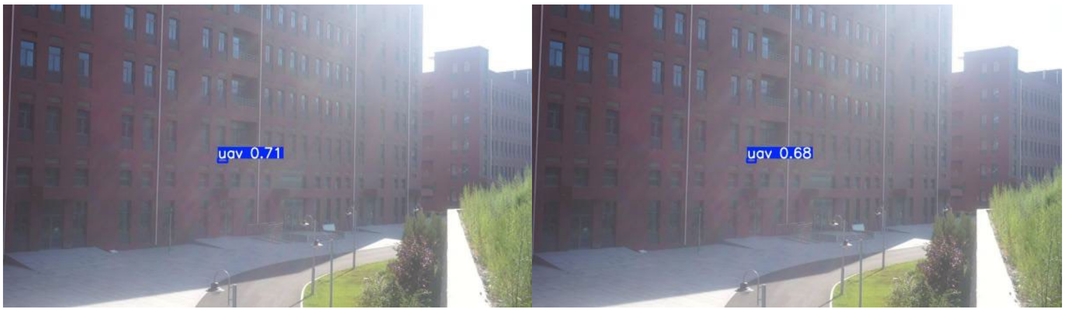

Figure 4: Comparison of the detection effect of the improved YOLOv10 and YOLOv10 under the influence of strong light.

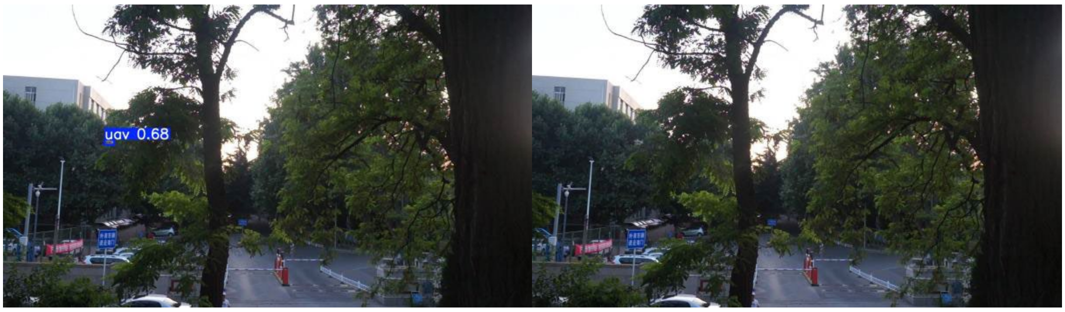

Figure 5 Comparison of the detection effect of improved YOLOv10 and YOLOv10 in complex background.

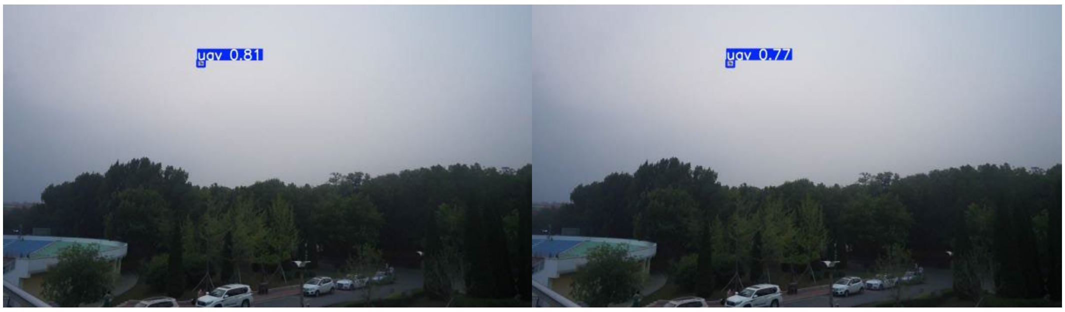

Figure 6 Comparison of the detection effect of improved YOLOv10 and YOLOv10 in open scenes.

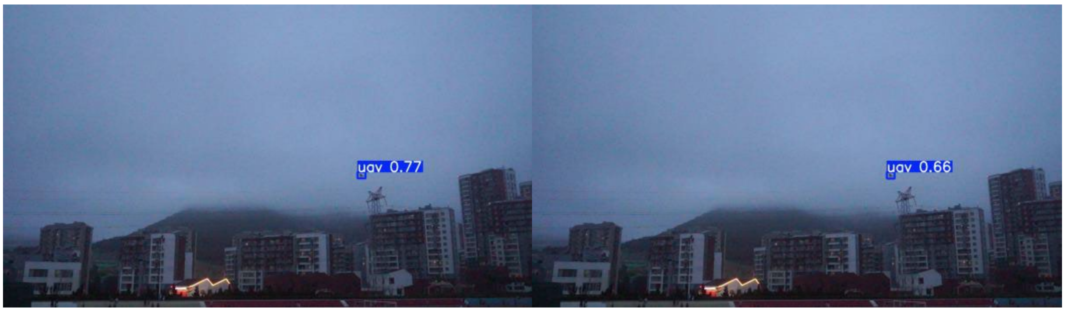

Figure 7 Comparison of the detection results of the improved YOLOv10 and YOLOv10 in the evening environment.

4. Conclusions

According to the problems existing in the small-target UAV detection algorithm, a new and improved YOLO object detection algorithm is proposed in this chapter, and in view of the limitations of the original model in small-target detection, the shared convolutional feature pyramid module is introduced, and the multi-scale dilated convolution and shared weight mechanism are used to enhance the feature perception ability and improve the expression accuracy of multi-scale targets. A multi-path and multi-scale Enhanced Feature Fusion Neck (MP-MSFPN) structure was designed to enhance the semantic information interaction and fusion capabilities between features at different scales. In order to verify the role of each improved module in improving the performance of YOLOv10 network, four sets of ablation experiments were designed, and FPSConv and MP-MSFPN neck structures were gradually introduced, and their effects on the average accuracy, detection accuracy, recall, computational size, model complexity and inference speed were evaluated respectively.

Acknowledgments

The paper was supported by Key R&D Program Projects in Shaanxi Province, China (Grant No. 2023-YBGY-115) and Special Fund for High-level Talents of Xijing University, China (Grant No. XJ18B01).

References

[1] Kim J K , Shin Y J , Kim J H ,et al .Faraway Small Drone Detection based on Deep Learning[J]. International journal of computer science and network security, 2020, 20(1).

[2] Saqib M , Sharma N , Khan S D ,et al. A Study on Detecting Drones Using Deep Convolutional Neural Networks[C]//AVSS 2017.IEEE Computer Society, 2017.DOI:10.1109/AVSS.2017.8078541.

[3] Xu X , Sun Y , Ding L ,et al. A Novel Infrared Small Target Detection Algorithm Based on Deep Learning[C]//ICAIP 2020: 2020 4th International Conference on Advances in Image Pr-ocessing.2020. DOI:10.1145/3441250.3441258.

[4] Al-Lqubaydhi N , Alenezi A , Alanazi T ,et al. Deep learning for unmanned aerial vehicles detection: A review[J].Computer Science Review, 2024, 51.DOI:10.1016/j.cosrev.2023.100614.

[5] Lenhard T R , Weinmann A , Jger S ,et al. YOLO-FEDER FusionNet: A Novel Deep Learning Architecture for Drone Detection[J].IEEE, 2024.DOI:10.1109/ICIP51287.2024.10647355.

[6] Ghazlane Y , Ahmed E H A , Hicham M . Real-time lightweight drone detection model: Fin-e-grained Identification of four types of drones based on an improved Yolov7 model[J]. Neurocomputing, 2024, 596(000):15.DOI:10.1016/j.neucom.2024.127941.

[7] Jiang R , Zhou Y , Peng Y . A Review on Intrusion Drone Target Detection Based on Deep Learning[J]. IEEE, 2021.DOI:10.1109/IMCEC51613.2021.9482092.

[8] Mei J , Zhu W . BGF-YOLOv10: Small Object Detection Algorithm from Unmanned Aerial Vehicle Perspective Based on Improved YOLOv10[J].Sensors (14248220), 2024, 24(21).DOI:10. 3390/s24216911.

[9] Wang H , Zhang Y , Zhu C . DAFPN-YOLO: An Improved UAV-Based Object Detection Algorithm Based on YOLOv8s[J].Computers, Materials & Continua, 2025, 83(2).DOI:10.32604/cmc.2025. 061363.

[10] Bo C , Wei Y , Wang X ,et al. Vision-Based Anti-UAV Detection Based on YOLOv7-GS in Complex Backgrounds[J]. Drones (2504-446X), 2024, 8(7).DOI:10.3390/drones8070331.