

An Improved Algorithm for Small Target Detection based on YOLO

Hao Meng, Jieqing Tan

Hefei University of Technology, Hefei Anhui, 230000, China

Abstract: Aiming at the problems of small target size, dense target, missing detection and false detection in UAV aerial images, an improved YOLOv8s small target detection algorithm MS-P2-YOLO is proposed in this paper. First, through several initial convolution layers, the composite convolution module extracts the target feature information, and uses pooling operations of different scales to capture the global context information of the image. Then, the scale adaptive fusion unit module is used to scale and splicing the feature map in a certain form to integrate the feature information from different scales or layers. Then the multi-dimensional feature integration module is used to adjust the features according to the size or number of channels of the input feature map, and to enhance the feature representation through certain forms of scaling and sequence processing. At the same time, P2 detection head is added at the end to further increase the detection ability of small targets, which greatly improves the problems of missing detection, false detection and large number of parameters of small targets. Experiments show that compared with the VisDrone2019 dataset of MS-P2-YOLO and YOLOv8s, P, R and mAP50% have increased by 9%, 12.9% and 12.9% respectively, and the number of parameters has decreased by 4%. At the same time, generalization and comparison experiments were also conducted in YOLOv5s, YOLOv8s, ATSS, TOOD, and Father-RCNN, and the detection results were visualized. The experiments showed that all parameters were improved.

Keywords: YOLO; Small Target Detection; Feature Extraction.

1. Introduction

Object detection is a key task in the field of computer vision, which aims to accurately identify and locate objects of interest from images or videos. With the rapid development of deep learning technology, the performance of object detection algorithms has been significantly improved. Among them, YOLO (You Only Look Once) series algorithms have attracted wide attention in the field of target detection because of its efficient and real-time characteristics. The YOLO algorithm turns the target detection task into a regression problem, and achieves fast and accurate target detection by simultaneously locating and classifying the target through a single neural network.

Small target detection is an important branch of target detection, which mainly focuses on small targets in images, such as pedestrians in video surveillance and small lesions in medical images. Because of the small size of small targets and their features are not obvious, traditional target detection algorithms are often difficult to achieve satisfactory results. However, YOLO algorithm[1-3], with its powerful feature[4] extraction capability and global optimization strategy[5], has shown excellent performance in small target detection.

This paper aims to discuss the small target detection algorithm based on YOLO, and analyze its principle, advantages and application scenarios. We will focus on the application of YOLO algorithm in small target detection, including its specific performance in UAV[6] image target recognition, monitoring scenarios, medical imaging and industrial detection.

In this paper, YOLOv8s [7-9] is selected as the benchmark algorithm, and on this basis, the small target detection[10] task is improved and innovated. The structure of YOLOv8 algorithm model [11] is mainly composed of Input, Backbone and head, among which the head includes Neck network and Detectet.

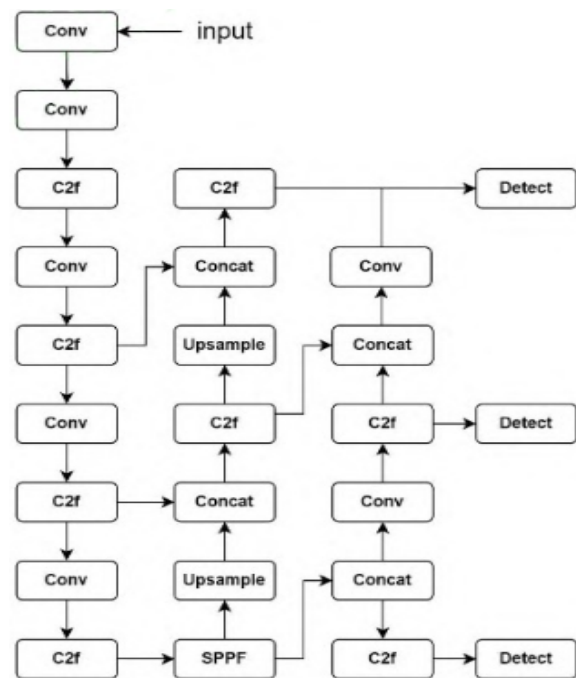


Fig 1. YOLOv8 Structure Diagram

In the input part, the functions include mosaic data enhancement, adaptive anchor calculation and other data enhancement techniques: in the backbone part, CSPDarkNet (Cross Stage Partial, which is the same as YOLOv5[12], is used Networks) structure, however, by changing the convolution of the first layer from 6*6 to 3*3, And according to the ELAN (Efficient Layer Aggregation Networks) of YOLOv7[13], C2F module[14] replaces C3 module. The final SPPF[15](Spatial Pyramid Pooling Fusion) module is composed of a convolution block and a maximum pooling layer[16] (Max The head part is replaced with the mainstream decoupling head, separating the classification task from the

detection task, that is, classification loss and boundary frame loss, so that the model can pay more attention to their respective focuses and improve the efficiency of the model. These improvements enable YOLOv8 to achieve faster detection speed, higher accuracy and ease of use, making it suitable for a variety of tasks including target detection, tracking, instance segmentation, image classification and attitude estimation. The structure of YOLOv8 algorithm model is shown in Figure 1.

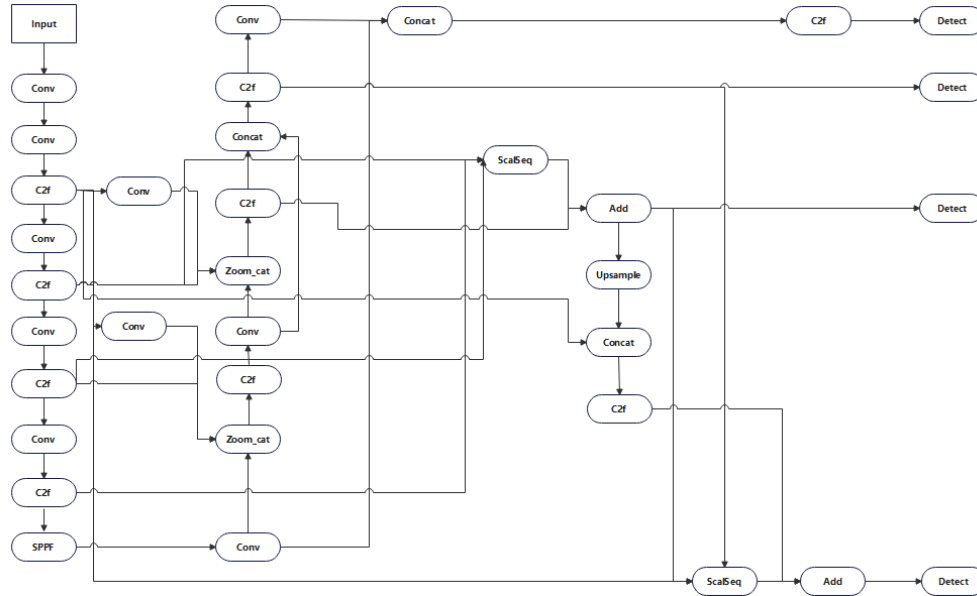


Fig 2. MS-YOLO

2.1. Scale-Adaptive Fusion Unit (SAFU)

Scale adaptive fusion unit

The module is designed to improve the performance of the YOLO v8 algorithm when dealing with objects at different scales, and through its unique "scale adaptive" and "fusion unit" functions, it enables flexible processing of input data and effective fusion of features.

2.1.1. Scale Adaptive

"Scale adaptive" is one of the core features of SAFU module. In actual target detection tasks, the size of the targets often varies greatly, from tiny insects to huge buildings can appear in the same scene. Traditional target detection algorithms often have difficulty in maintaining stable performance when dealing with such scale changes. The SAFU module can automatically adjust its processing strategy according to the size of the object in the input image through its built-in scale adaptive mechanism.

Specifically, the SAFU module contains a number of feature extraction layers or processing units of different scales, which can be selectively activated or suppressed according to different input scales, so as to achieve accurate capture and processing of objects of different scales. This mechanism makes the SAFU module more flexible and robust when dealing with multi-scale targets.

2.1.2. Fusion Unit

The "Fusion unit" is another key component of the SAFU module. In target detection tasks, the effective fusion of features can improve the performance of the algorithm to

It's important. Through its fusion unit, the SAFU module can effectively fuse and integrate features from different layers or scales to generate a richer and more discriminative

2. Improved YOLO Small Target Detection Algorithm

MS-YOLO is a new YOLO framework based on attention scale sequence fusion, which combines spatial and scale features to achieve accurate and fast small target detection. The transformation in this paper mainly includes the following two points, and the structure diagram after the transformation is shown in Figure 2.

feature representation.

This fusion process involves a variety of feature fusion strategies, such as feature concatenation, feature weighting and so on. Through a reasonable fusion strategy, SAFU module can make full use of the complementarity between different layers or different scale features to further improve the algorithm's adaptability to complex scenes.

2.2. Multi-Dimensional Feature Integration Module

Multi-dimensional Feature Integration Module (MDIM)

The MDIM module receives feature maps from different layers as input. These feature maps contain different scales of information, from low-level details to high-level global information. MDIM module fuses and encodes these feature maps of different scales through specific coding strategies to form a comprehensive feature representation.

Specifically, the MDIM module uses operations such as convolution and pooling to extract and integrate feature information. These actions capture the key information in the feature map and fuse it together to form a richer and more detailed representation of the feature. In this way, the MDIM module is able to enhance the model

The ability to detect and segment objects of the same scale.

3. Experiment and Preliminary Preparation

3.1. Experimental Data

The data set used in the experiment is VisDrone2019, which was released by the Machine Learning and Data Mining Laboratory of Tianjin University, and contains 6,471

training images, 5,480 verification images and 1,610 test images, with a total of 8,629 still pictures taken by drones in different scenes. The data set covers 10 different categories, including pedestrians, bicycles, cars, vans, trucks, tricycles, awning tricycles, buses and motorcycles. There are a lot of small targets and occlusion problems in the data set, which makes the detection of the data set very challenging.

3.2. Experimental Environment

The GPU used in the experiment is NVIDIA 3090 24G, using a 15-core CPU, pytorch framework version 2.0.0, python version 3.8.0, cuda version 11.8. The specific experimental parameter Settings are shown in Table 1.

Table 1. Experimental environment

Argument	Settings	Argument	Settings
python	3.8.16	torchvision	0.14.1+cu117
torch	1.13.1+cu117	timm	0.9.8
mencv	2.1.0	mmengine	0.9.0

3.3. Experimental Evaluation Index

When evaluating the performance of YOLOv8s, it is necessary to consider a series of important evaluation indicators.

Using precision (P), recall rate (R), average accuracy (fmAP) and other indicators in the final results of small target detection tasks, and each indicator can be defined by the following formula: TP (True Positives) indicates the number of real defect samples, FP (False Positives) indicates the number of False positives, FN(False Negatives) indicates the number of undetected defect samples, and N indicates the total number of detected categories.

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

$$f_{AP} = \int_0^1 P(R) dR$$

$$f_{mAP} = \frac{1}{k} \sum_{s=1}^k f_{AP_s}$$

The fAP measures the performance of the model by calculating the area under the precision rate-recall curve for each category. fmAP is used to measure the performance of a target detection model across multiple categories because it takes into account the performance of each category rather than just a single category.

4. Model Training and Result Analysis

4.1. Model Test Result

In order to verify the effectiveness of MS-P2-YOLO proposed in this paper, ablation experiments were conducted on YOLOv8s, MS-YOLO and MS-P2-YOLO. The specific parameters are shown in Table 2.

As can be seen from Table 1, MS-YOLO and YOLOv8s have similar detection capabilities. However, after integrating the small target detection layer on the basis of MS-YOLO, the obtained MS-P2-YOLO improves the detection ability of small targets by optimizing the network structure, and the accuracy is increased by 12.9%. At the same time, the number of parameters is reduced by 4%. In summary, it can be seen that MS-P2-YOLO, while maintaining high detection accuracy, reduces the computational complexity and the demand for hardware resources, which makes the detection of small targets more efficient and accurate.

Table 2. Model comparison experiment

Algorithm	P	R	f _{mAP} (%)	Parameter	GFLOPs
YOLO-V8s	49.9%	35.2%	37.2%	44634484	172.3
YOLO-MS	47.5%	46.6%	47.3%	43615860	164.9
YOLO-MS-P2	58.9%	48.1%	50.1%	42666304	199.0

4.2. Ablation Experiment

From the detailed data in Table 2, we can see the remarkable performance improvement of YOLOv8s model after a series of careful optimization and improvement. The original YOLOv8s without additional improvements has demonstrated excellent detection accuracy (37.2% fmAP) and moderate model size (44.6million). However, after the

introduction of MS-YOLO improved network, the model performance ushered in a qualitative leap, the detection accuracy was increased from 37.2% to 47.3%, and the number of parameters was significantly reduced to 42.6million, realizing the initial optimization of the model performance. Furthermore, when P2 small target detection layer technology is integrated on the basis of MS-YOLO, the performance of the model is improved more significantly. The fmAP value

climbed further, to 50.1%, while the number of parameters fell by another 1.9million. This result fully demonstrates that MS-P2-YOLO can efficiently retain critical information while reducing model complexity, thus greatly improving detection accuracy.

4.3. Result Analysis

In order to verify the effectiveness of the proposed algorithm in the transmission line insulator defect detection task, lightweight YOLOv5s [17], YOLOv8s, ATSS [18], TOOD [19], and Fast-RCNN [20] are selected as comparison algorithms. In the experiment, the mean accuracy was used as the evaluation index, where the crossover ratio (IoU)

threshold was set at 0.5. The experimental results are shown in Table 3. As can be seen from Table 3, the proposed algorithm is at least 12.9% higher than the other five comparison algorithms in f_{mAP} on the premise that the model size meets the requirements of edge computing deployment. This significant performance improvement fully proves that the proposed algorithm has higher accuracy in small target detection tasks. At the same time, it is worth noting that the number of parameters of the algorithm in this paper is only 42.6 million, which makes the algorithm more efficient in small target detection and further improves its feasibility in practical applications.

Table 3. Comparative Experiment

Algorithm	$f_{mAP}(\%)$	Parameter/M	GFLOPs
ATSS	20.4%	38.91	110.0
TOOD	20.4%	32.04	199.0
Faster-RCNN	32.9%	41.39	208.0
YOLO-V5s	33.9%	37.2	165.0
YOLO-V8s	37.2%	44.6	172.3
YOLO-MS	47.3%	43.6	164.9
YOLO-MS-P2	50.1%	42.6	199.0

5. Conclusion

Aiming at the requirement of high preparation rate of small target detection task, this paper proposes a lightweight YOLOv8s algorithm, and the conclusions are as follows:

- (1) The YOLOv8s algorithm, based on both real-time and high precision, reduces the technical difficulty of subsequent optimization by its excellent performance and robustness, and provides a solid foundation for the detection of small targets.
- (2) Using MS-YOLO model to add small target detection layer, the number of algorithm parameters is significantly reduced by 4%, while the accuracy is increased by 12.9%. This innovative strategy not only optimizes the model performance, but also improves the detection ability of small targets, ensuring the accuracy and reliability of detection results.
- (3) In order to deal with more complex detection scenarios, this paper uses the publicly available UAV data set VisDrone2019 for training. In the future, we plan to further deploy the proposed algorithm in the equipment for small target detection to verify and further improve its performance in practical applications.

References

- [1] LI Q, ZHAO F, XU Z, et al. Insulator and damage detection and location based on YOLOv5[C]//Proceedings of the International Conference on Power Energy Systems and Applications. Piscataway: IEEE Press, 2022: 17-24.
- [2] HAN G, LI T, LI Q, et al. Improved algorithm for insulator and its defect detection based on YOLOX[J]. Sensors, 2022, 22(16): 6186.
- [3] ZHANG Y, WU Z, WANG X, et al. Improved YOLOv8 insulator fault detection algorithm based on BiFormer [C]// Proceedings of the IEEE 5th International Conference on Power, Intelligent Computing and Systems. Piscataway: IEEE Press, 2023: 962-965.
- [4] LI Y, WEI S, LIU X, et al. An improved insulator and spacer detection algorithm based on dual network and SSD[J]. IEICE Transactions on Information and Systems, 2023, 106(5): 662-672.
- [5] YANG W, BO D, TONG L S. TS-YOLO: an efficient yolo network for multi-scale object detection[C]//Proceedings of the IEEE 6th Information Technology and Mechatronics Engineering Conference. Piscataway: IEEE Press, 2022: 656-660.
- [6] CHEN Y, LIU H, CHEN J, et al. Insu-YOLO: an insulator defect detection algorithm based on multiscale feature fusion[J]. Electronics, 2023, 12(15): 3210.
- [7] GAO Z, YANG G, LI E, et al. Novel feature fusion module-based detector for small insulator defect detection[J]. IEEE Sensors Journal, 2021, 21(15): 16807-16814.
- [8] MUSSINA D, IRMANOVA A, JAMWAL P K, et al. Multi-modal data fusion using deep neural network for condition monitoring of high voltage insulator[J]. IEEE Access, 2020, 8: 184486-184496.
- [9] YU Y, CAO H, WANG Z, et al. Texture-and-shape based active contour model for insulator segmentation[J]. IEEE Access, 2019, 7: 78706-78714.
- [10] HAO Z X, WANG Q. Enhanced algorithm for small target detection in UAV aerial images based on YOLO-v7 [J]. Software Guide, 2024, 23(1): 167-172.
- [11] PAN W, WEI C, QIAN C Y, et al. Improved YOLOv8s model for small object detection from perspective of drones [J]. Computer Engineering and Applications, 2024, 60(9): 142-150.

- [12] CHENG Q H, CHEN D F, WANG X F. Small target detection in UAV images based on NDMYOLOv8[J/OL]. Computer Technology and Development, 19, <http://kns.cnki.net/kcms/detail/61.1450.tp.20240520.0906.012.html>, 2024-07-04.
- [13] Lou H, Duan X, Guo J, et al. DC-YOLOv8: small-size object detection algorithm based on camera sensor[J]. Electronics, 2023, 12(10): 2323.
- [14] Zhang H, Cisse M, Dauphin Y N, et al. mixup: beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.
- [15] Zhang L, Zheng J, Li C, et al. CCDN-DETR: a detection transformer based on constrained contrast denoising for multi-class synthetic aperture radar object detection[J]. Sensors, 2024, 24(6): 1793.
- [16] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(7): 12993-13000.
- [17] Tong Z, Chen Y, Xu Z, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[J]. arXiv preprint arXiv:2301.10051, 2023.
- [18] Rezaatofghi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.
- [19] Meng Caixia, Wang Zhaonan, Shi Lei, et al. Improved railroad foreign object intrusion detection algorithm for YOLOv5s [J]. Journal of Chinese Computer Systems, 2024, 45(4):879-886, doi:10.20009/j.cnki.21-1106/TP.2022-0607.
- [20] CHEN R, ZHENG H F, JIANG H Y, et al. Combination of simul-based transfer learning and adaptive fusion for UAV small object detection[J]. Journal of Chinese Computer Systems, 2023,44(8):17431749, doi:10.20009/j.cnki.21-1106/TP.2021-0905.