

# Research on Traffic Sign Detection Based on Improved YOLOv8

Zhongjie Huang<sup>1</sup>, Lintao Li<sup>1</sup>, Gerd Christian Krizek<sup>2</sup>, Linhao Sun<sup>1</sup>

<sup>1</sup>School of Computer and Information Engineering, Xinxiang University, Xinxiang, China

<sup>2</sup>Department of Applied Mathematics and Physics, UAS Technikum Wien, Vienna, Austria

Email: 619768159@qq.com, 5691214@qq.com, gerd.krizek@technikum-wien.at, 1913314504@qq.com

**How to cite this paper:** Huang, Z.J., Li, L.T., Krizek, G.C. and Sun, L.H. (2023) Research on Traffic Sign Detection Based on Improved YOLOv8. *Journal of Computer and Communications*, 11, 226-232.  
<https://doi.org/10.4236/jcc.2023.117014>

**Received:** June 23, 2023

**Accepted:** July 28, 2023

**Published:** July 31, 2023

---

## Abstract

Aiming at solving the problem of missed detection and low accuracy in detecting traffic signs in the wild, an improved method of YOLOv8 is proposed. Firstly, combined with the characteristics of small target objects in the actual scene, this paper further adds blur and noise operation. Then, the asymptotic feature pyramid network (AFPN) is introduced to highlight the influence of key layer features after feature fusion, and simultaneously solve the direct interaction of non-adjacent layers. Experimental results on the TT100K dataset show that compared with the YOLOv8, the detection accuracy and recall are higher.

## Keywords

Traffic Sign Detection, Small Object Detection, YOLOv8, Feature Fusion

---

## 1. Introduction

Traffic signs play an important role in traffic flow, which contribute to the development of assisted driving and autopilot technology and accelerates the construction of intelligent transportation system. Currently, the main solution for traffic signs detection is to collect images by vehicle-mounted cameras and use computer vision and pattern recognition methods for detection and recognition [1]. With the rapid development of deep learning technology, target detection algorithms, such as Faster RCNN, YOLO, FCOS, etc. have been widely used in traffic signs detection.

In research on traffic sign detection, Li *et al.* [2] proposed an improved Faster R-CNN detection algorithm to deal with the problem of complicated background and small traffic sign detections in wide-field traffic scenarios. Luo *et al.* [3] introduced Bi-FPN for feature fusion and used GAM (global attention mechanism)

[4] to enhance the feature extraction ability of the network to verify the effectiveness of the improved YOLOv5 algorithm for traffic sign detection. Chen *et al.* [5] proposed a multi-scale feature fusion detector based on the improved FCOS (full-convolution single-stage object detection) [6] to solve the problems of small traffic sign detection. The above-mentioned have achieved certain results in the detection of traffic signs, but there are still problems such as low detection accuracy and missed detection in actual detection. Therefore, further research is needed on traffic sign detection.

This paper based on the YOLOv8 algorithm proposed a data enhancement method by analyzing the characteristics of small objects, and introduced a new method of feature fusion to improve the accuracy. The improvements are as follows:

- Adding noise and blur processing on the original data augmentation method makes the network pay more attention to fine-grained features, which contribute to detecting small traffic signs.
- In the feature fusion stage, the asymptotic feature pyramid network (AFPV) is used to achieve direct interaction of non-adjacent layers and enhance the expressive ability of key features.

## 2. Algorithm Overview

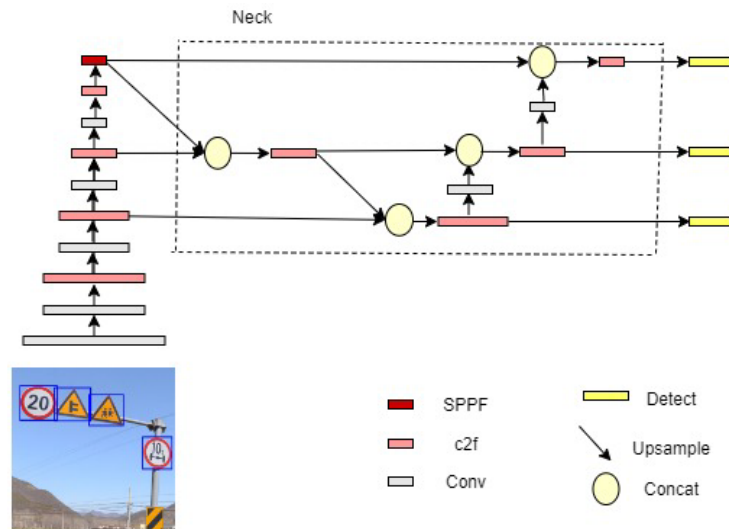
The YOLO family is a classic object detector. Since the first edition of this algorithm was published in 2015, it has achieved a leading efficiency with a single-stage framework and has quickly become a mainstream detection algorithm. Through continuous research and innovations, different versions of YOLO have been proposed. The latest version is the YOLOv8 [7] algorithm, which was open-sourced by Ultralytics in January 2023. This algorithm introduces new features and improvements, becoming the best model in the YOLO family. YOLOv8 includes four parts: Input, Backbone, Neck and Output. The structure of YOLOv8 is shown in **Figure 1**.

### 2.1. YOLOv8 Algorithm Theory

Input mainly includes color perturbation, spatial perturbation, Mosaic, MixUp. Different numbers of pictures are spliced after processing a single one by combined data augmentation, which increases the multi-directional object perspective and also enriches the diversity of image backgrounds.

Backbone mainly includes convolution layer, C2F layer and SPPF layer. The C2F structure is different from the C3 module in YOLOv5 which draws on the idea of ELAN (efficient layer aggregation network) in YOLOv7 [8], which increases the efficiency of gradient propagation and enables the network to quickly converge. The SPPF layer maintains the design in YOLOv5.

Neck adopts a structure combining FPN (feature pyramid network) and PAN (path aggregation network) [9]. The features of adjacent layers are concatenate which are the input to the C2F module. As features are passed from top to bottom



**Figure 1.** YOLOv8 network structure.

and bottom to top, high-level semantic features with underlying features are combined.

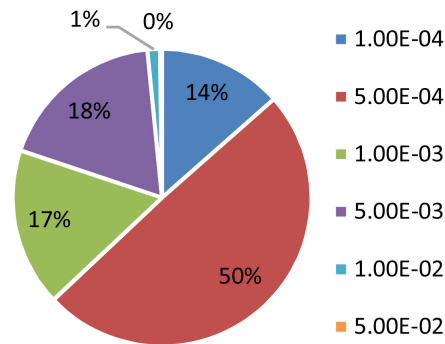
Output realizes the decoupling of detection and classification. The bottom-level features are used to obtain the information of small target objects, and the top-level features are results of large target objects. Each detection layer outputs a result vector, which contains location and corresponding category information.

## 2.2. Improvements in Data Augmentation

DA (data augmentation) is widely used in deep learning data processing, such as image scaling, translations, random rotation, etc. The above methods can get different perspectives of the object. CutMix, Mixup, Mosaic, Copy-Paste, etc. combine different pictures, which can increase the diversity of backgrounds. Research on data enhancement for small target object detection [10] focuses on oversampling or resampling of small target samples, by transforming small target objects and pasting them in different positions in the image to increase the number of small target objects in training.

This paper uses the TT100K [11] dataset, which includes traffic signs produced by Tsinghua University using Tencent Street View panoramas. The pictures are obtained under different weather conditions, and the illumination changes greatly. Usually, the proportion of traffic signs in the picture is small, only 0.2% of the image size. The proportion of the target size in the training pictures of this data set is counted, and the result is shown in **Figure 2**. It can be seen that the proportion of traffic signs is mostly concentrated in the range of 0.01% to 0.05%.

DA, such as motion blur, fog and raindrop noise, is added on the basis of original method to simulate images acquired in different weather and traveling states. The effect after data augmentation is shown in **Figure 3**.



**Figure 2.** The ratio of image size in TT100K training set.

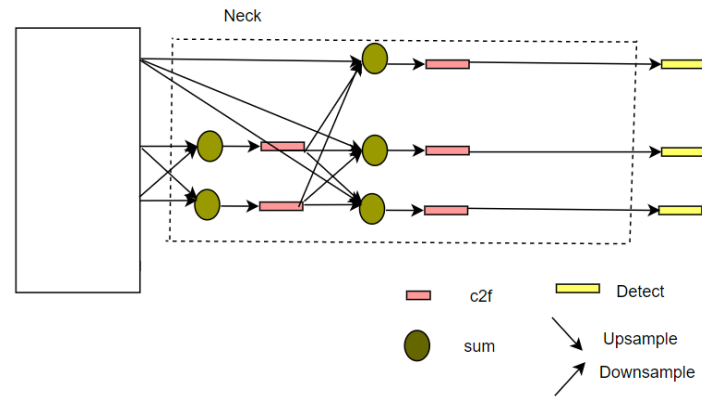


**Figure 3.** Comparison of DA. (a) DA in YOLOv8; (b) Ours.

### 2.3. Improvement of Feature Fusion

Feature pyramid is an important part of neural network, which is used to fuse features to obtain comprehensive description information. Neck in YOLOv8 achieves the consistency of the size of the adjacent feature maps by up-sampling or convolution, and then directly concatenates the channels of the feature maps to complete the preliminary fusion between the features. On the other hand, unidirectional information in concatenate layer can be used, or feature information in another direction can be obtained by using the feature pyramid after the feature fusion in the same direction is completed. Moreover, feature fusion cannot be directly performed between non-adjacent feature layers in the current network. This will lead to the loss of information when the features of non-adjacent layers are fused.

Therefore, this paper uses AFPN [12] to complete the feature fusion in Neck, and its structure is shown in **Figure 4**. This method combines the ideas of the pyramid idea and ASFF (adaptively spatial feature fusion) [13] to realize the fusion of features. Firstly, it is the fusion of adjacent features. By up-sampling or down-sampling between adjacent layers to get the unification of the size of the feature map in the current layer, and then uses the adaptive spatial fusion of features which pays more attention to key layers, as shown in formula 1. Among them,  $\alpha_{ij}^l$  and  $\beta_{ij}^l$  respectively represent the weight of the input during linear



**Figure 4.** AFPN structure.

combination, satisfying  $\alpha_{ij}^l + \beta_{ij}^l = 1$ ,  $x_{ij}^{n \rightarrow l}$  represents the conversion of eigenvectors located at  $(i, j)$  from the  $n$ th layer to the  $l$ th layer. This formula is also applicable to linear combinations between multi-layer features. Secondly, it is the fusion of cross-layer features. Since the feature fusion already contains the adjacent feature information, the language gap between features is reduced when the cross-layer features are fused.

$$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} \quad (1)$$

### 3. Model Training and Testing

#### 3.1. Experimental Environment

This article builds an algorithm running environment based on Docker. The specific training environment is: Ubuntu20.04 system, Nvidia 3080Ti which memory size is 12 G, the Pytorch version is 1.10.0a0, and the CUDA version is 12.0.

#### 3.2. Evaluation Criteria

Recall and mAP (mean average precision) are used commonly to evaluate the model performance. The recall mainly focuses on the ratio of the number of detected targets to the total number of targets. The mAP indicates the average value of the average accuracy rate of each type of target under a certain IOU (intersection over union) condition for all categories.

#### 3.3. Comparative Experiment

This paper compares and analyzes the comprehensive performance of the improved YOLOv8 network. Parameters of training such as learning rate and batch size remain unchanged except for the module improvement, and the number of epochs is 500. The experimental results are shown in **Table 1**, it can be found that compared with the standard network, the improved YOLOv8 has increased mAP by 3.31 percentage points, Recall has increased by 3.59 percentage points.

We also conduct ablation experiments on the influence of different improved parts on the model, and the results are shown in **Table 2**. Among them, “x”

**Table 1.** Comprehensive index test results.

Model	mAP/%	Recall/%
YOLOv8	56.22	49.33
YOLO-ours	59.53	52.92

**Table 2.** Ablation study of data augmentation and AFPN.

YOLOv8	DA-ours	AFPN	mAP/%	Recall/%
√	×	×	56.22	49.33
√	√	×	55.63	51.27
√	×	√	58.04	50.83

means that the corresponding improvement strategy is not used, and “√” means used. The recall of the ablation of DA-ours is improved by 1.94 percentage points, while the mAP is slightly lower than YOLOv8. The AFPN improves map and recall by 1.82 and 1.5 respectively, which enhances feature fusion.

#### 4. Conclusion

Aiming at the problems of missed detection and low detection rate in traffic signs, this paper proposes an improved YOLOv8 algorithm to detect traffic signs, which uses the data augmentation method in a specific scenario to effectively enhance the diversity of the data itself, so that the network can learn effective features and apply AFPN for fusion features. Experiments show that the improved algorithm can reduce missed detection and improve accuracy.

#### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

#### References

- [1] Zhang, Z.W., Gao, Y., Wang, J., et al. (2020) Deep Learning Traffic Sign Recognition System Based on YOLOv3. *Building Electricity*, **39**, 64-68. (In Chinese)
- [2] Li, Z., Zhang, H.H. and Deng, J.Y. (2021) Traffic Sign Detection Algorithm Based on Improved Faster R-CNN. *Chinese Journal of Liquid Crystals and Displays*, **36**, 484-492. (In Chinese) <https://doi.org/10.37188/CJLCD.2020-0195>
- [3] Luo, K.T. and He, Q.L. (2022) Chinese Traffic Sign Detection Based on YOLOv5 Algorithm. *Operational Research and Fuzzy Science*, **12**, 1570-1584. (In Chinese) <https://doi.org/10.12677/ORF.2022.124165>
- [4] Liu, Y.C., Shao, Z.R. and Hoffmann, N. (2021) Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. ArXiv: 2112.05561.
- [5] Chen, Z. and Cheng, Y.Y. (2022) Improved Traffic Sign Detection Based on FCOS Algorithm. *Computer Engineering and Design*, **43**, 3271-3278. (In Chinese)
- [6] Tian, Z., Shen, C.H., Chen, H. and He, T. (2019) FCOS: Fully Convolutional One-Stage

- Object Detection. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 9626-9635.  
<https://doi.org/10.1109/ICCV.2019.00972>
- [7] Jocher, G., Chaurasia, A. and Qiu, J. (2023) YOLO by Ultralytics.  
<https://github.com/ultralytics/ultralytics>
- [8] Wang, C.-Y., Bochkovskiy, A. and Liao, H.-Y.M. (2022) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. ArXiv: 2207.02696.
- [9] Liu, S., Qi, L., Qin, H.F., Shi, J.P. and Jia, J.Y. (2018) Path Aggregation Network for Instance Segmentation. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 18-23 June 2018, 8759-8768.  
<https://doi.org/10.1109/CVPR.2018.00913>
- [10] Li, H.G., Yu, R.N. and Ding, W.R. (2021) Research Development of Small Object Traching Based on Deep Learning. *Acta Aeronautica et Astronautica Sinica*, **42**, 24691-024691. (In Chinese)
- [11] Zhu, Z., Liang, D., Zhang, S., *et al.* (2016) Traffic-Sign Detection and Classification in the Wild. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 27-30 June 2016, 2110-2118.  
<https://doi.org/10.1109/CVPR.2016.232>
- [12] Yang, G.Y., Lei, J., Zhu, Z.K., *et al.* (2023) AFPN: Asymptotic Feature Pyramid Network for Object Detection. ArXiv: 2306.15988.
- [13] Liu, S.T., Huang, D. and Wang, Y.H. (2019) Learning Spatial Fusion for Single-Shot Object Detection. ArXiv: 1911.09516.