Road Sign Detection in Extreme Weather Conditions

Jianhui Zhang BNVKKDIU@163.COM
National and Local Joint Engineering Research Center for Intelligent Vehicle Road Collaboration and Safety Technology, School of Automobile and Transportation, Tianjin University of Technology and Education, Tianjin, China

Kun Wei* YANGQUANWEIKUN@163.COM
National and Local Joint Engineering Research Center for Intelligent Vehicle Road Collaboration and Safety Technology, School of Automobile and Transportation, Tianjin University of Technology and Education, Tianjin, China

Wenqing Wei EEINQING@163.COM
National and Local Joint Engineering Research Center for Intelligent Vehicle Road Collaboration and Safety Technology, School of Automobile and Transportation, Tianjin University of Technology and Education, Tianjin, China

Murat Muzdybayev 854626117@QQ.COM
International Engineering College, East Kazakhstan Technical University, Ockemeh, Republic of Kazakhstan; Protozanov str., 69, Ust-Kamenogorsk, 070024 East Kazakhstan Region, Kazakhstan
**Corresponding author*

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

In road traffic sign detection, the low detection precision of road traffic signs in the detection screen is attributed to their small proportion and adverse weather conditions, such as fog, snow, and nighttime. To enhance the detection precision of road signs in extreme weather, this paper proposes an algorithm based on a lightweight improvement of YOLOv8n, referred to as FRPP-YOLOv8n (FFA-Net-RFAConv-PSA-P2-YOLOv8n). Firstly, the improved lightweight FFA-Net module is incorporated to dehaze the images. Secondly, RFAConv (Receptive-Field Attention convolutional operation) is introduced to enhance network performance. The PSA (Partial self-attention) mechanism is employed to improve detection precision, and finally, a small target detection layer is added to enhance the detection precision of small targets. The experimental results indicate that the improved algorithm achieves 53.4% mAP50-95 and 82.1% mAP50 on the CCTSDB2021 traffic sign dataset, which is an increase of 4.5% for both metrics compared to the original algorithm. Additionally, it maintains a high precision rate of 91.3%, representing a 4% improvement over the original algorithm.

Keywords: Yolov8n; traffic signs; FFA-Net; RFAConv; PSA; P2

1. Introduction

Traffic signs play a crucial role in the domain of autonomous driving by conveying essential information to vehicles, guiding their direction, and influencing aspects such as path planning and speed regulation, thereby enhancing driving safety. With the proliferation of autonomous vehicles and the gradual implementation of assisted driving technologies, the challenges posed by vehicle cameras and sensors under adverse weather conditions—such as fog, snow, and nighttime—are garnering increasing attention. At the same time, the percentage of traffic sign images is small, and

the false detection rate is high. Therefore, accurately recognizing road traffic signs in poor weather conditions has become a significant research direction.

Khan et al. (2018) proposed a new Intelligent Traffic Sign Recognition (ITSR) system with light preprocessing. The system employs the Dark Area Sensitive Tone Mapping (DASTM) technique, which enhances the illumination of only the darker areas of the image, with minimal impact on the brighter areas. DASTM is combined with an optimized version of the TS detector of YOLOv3. Experiments show that the system achieves high detection precision.

Yao et al. (2022) addresses the shortcomings of YOLOv4-Tiny in traffic sign detection by proposing an adaptive feature pyramid network (AFPN) to better fuse feature layers of different scales. Additionally, two receptive field blocks (RFB) are added to the backbone network to enhance feature extraction for small targets.

Ahmed et al. (2021) proposed the modular framework DFR-TSD for challenging conditions such as rain, snow, fog, and lens smudges. The core of the network is the Enhance-Net enhancement network, which combines challenge classifiers, region enhancement strategies, and detection loss constraints to focus on the local enhancement of traffic sign regions. On the CURE-TSD dataset, DFR-TSD significantly outperforms end-to-end models such as Faster R-CNN and RetinaNet.

Compared to two-stage target detection algorithms, single-stage target detection algorithms feature simpler models and faster processing speeds, making them increasingly the preferred choice for target detection. Consequently, this paper proposes four key enhancements based on the YOLOv8n model: (1) Incorporating an improved lightweight defogging mechanism, FFA-Net, to enhance image clarity and thereby improve detection precision. (2) Introducing RFAConv to boost network performance. (3) Implementing PSA to enhance both channel and spatial attention mechanisms, further improving detection precision. (4) Adding a small target detection layer to further increase the detection precision of small targets. The improved algorithm is illustrated in Figure 1.

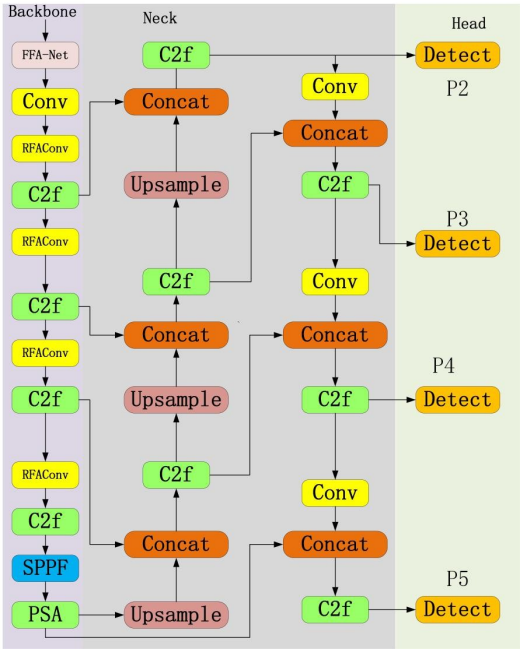

Figure 1: FRPP-YOLOv8n structural figure

2. FFA-Net

We introduce the end-to-end feature fusion network, FFA-Net (Qin et al., 2020). The mechanism for fog removal operates by inputting a blurred image, which is processed through a shallow feature extraction component. This output is then fed into a group of N architectures that utilize multiple skip connections. The outputs from these N architectures are subsequently passed through the Feature Attention (FA) module, where the features are fused together. The fused features are then directed to the reconstruction component and the global residual learning structure to yield a fog-free output. As illustrated in Figure 2, FFA-Net comprises three core sections: the Feature Attention module FA, the basic block structure, and the Group architecture designed with multiple skip connections. Notably, the FA module integrates both channel and pixel attention mechanisms, allowing for flexibility and versatility in targeting various features and pixels. The FA attention module is comprised of Channel Attention (CA) and Pixel Attention (PA), as depicted in Figure 3. Channel Attention focuses on different channel features, while Pixel Attention hones in on distinct pixels, thereby enhancing the ability to distinguish and process diverse types of image information. This adaptability in processing various information types also improves the representation capability of the CNN.

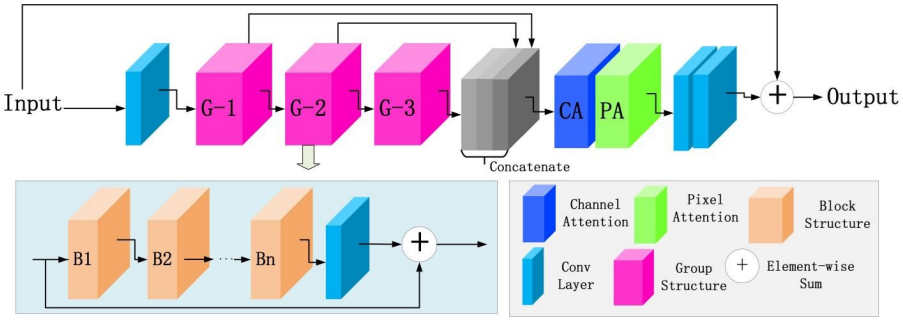

Figure 2: Detailed structure diagram of FFA-Net module

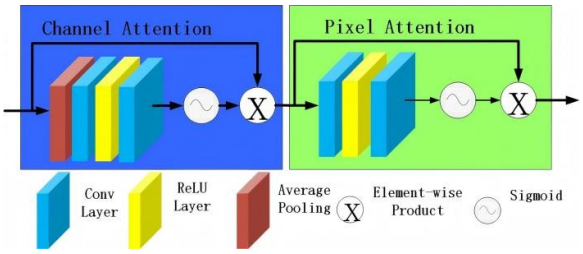

Figure 3: FA module

The basic module primarily consists of a localized residual set and a feature attention module FA. Local residual learning enables less important information, such as mist or low-frequency regions, to be bypassed through multiple local residual connections, allowing the main network to concentrate on relevant information. This structure can further enhance network performance and training stability, as illustrated in Figure 4.

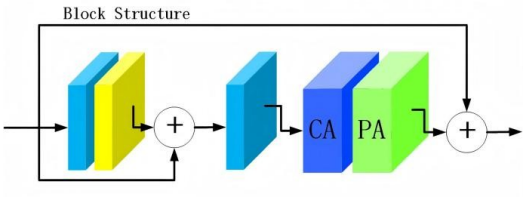

Figure 4: Basic Block Structure

Finally, a Group architecture with multiple levels of jump connections is proposed. Each Group architecture comprises multiple Block structures connected in series, integrated with a jump connection structure. The successive B-structure increases the model's depth and expressive capability, while the jump connection structure allows the modules to circumvent challenging training scenarios. This approach enables the network to learn adaptive weights for significant features, such as thick fog characteristics, and to retain the shallow information of the image, ensuring that the de-fogged image maintains the non-haze features of the original image as much as possible. To optimize the module's efficiency, we utilize only one Block in each Group architecture. Moreover, compared to the original model, the improved lightweight model sets the channel dimension of the feature map to 8, with the channel compression ratio in the CA layer reduced to 1/4 of the channel dimension. In contrast to the original model's feature channel count of 64, the channel compression ratio in the CA layer is now 1/16 of the channel dimension, thereby decreasing computational complexity.

3. RFAConv

The performance of convolutional neural networks is constrained by the standard convolutional operation. This limitation arises because the convolutional operation relies on shared parameters, rendering it insensitive to the variations in information caused by positional changes. RFAConv (Zhang et al., 2023) introduces an innovative convolutional operation that addresses the issue of parameter sharing in traditional convolutional neural networks. It emphasizes the importance of distinct features within the receptive-field slider and prioritizes spatial features, thereby enhancing the network's sensitivity to these spatial characteristics. By incorporating the Receptive-Field Attention (RFA) mechanism, RFAConv dynamically generates attention weights for each receptive field, significantly improving network performance. Furthermore, RFAConv has been designed to remain lightweight, requiring only a minimal increase in parameters and computational overhead while substantially enhancing network performance, as illustrated in Figure 5.

4. PSA

Self-attention has been widely utilized in various vision tasks due to its exceptional global modeling capabilities. However, the self-attention module exhibits high computational complexity and memory consumption. To address this issue, and considering the prevalence of attentional head redundancy (Xu et al., 2023), this paper introduces the efficient PSA module in Yolov10 (Wang et al., 2024). This module features a two-path architecture that enhances the spatial-channel feature representation of convolutional neural networks, effectively capturing both local structural patterns and

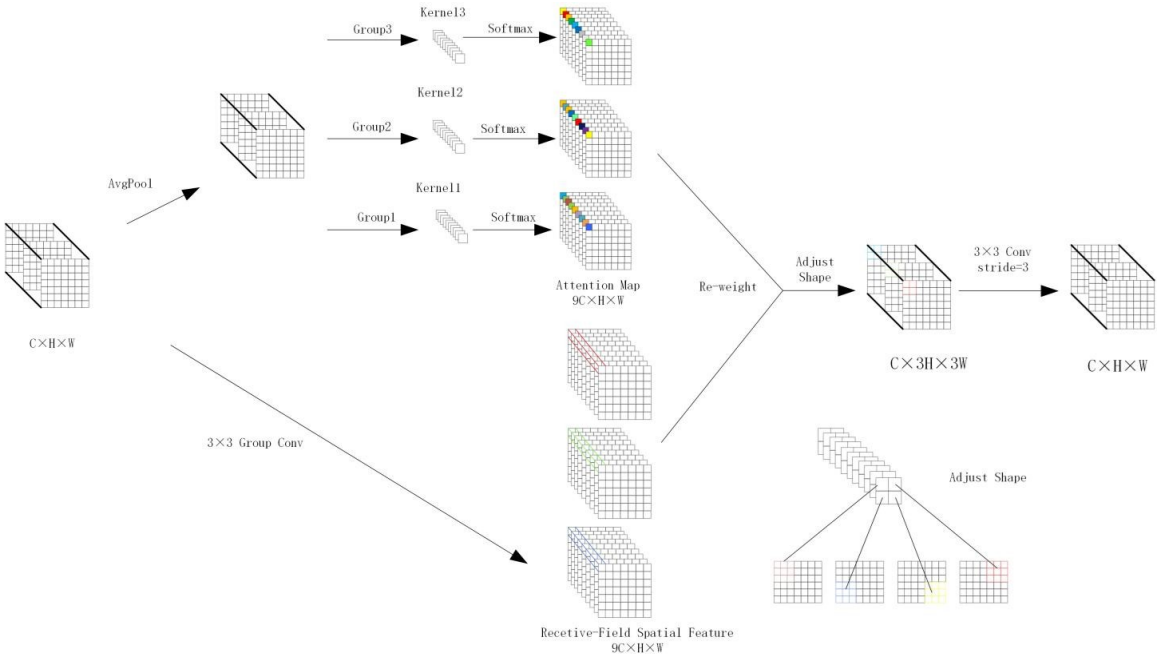

Figure 5: RFAConv module structure diagram

global contextual dependencies through parallel processing streams. Its primary working principle involves uniformly dividing the cross-channel features after a 1x1 convolution into two parts; one part is processed through the NPSA module, which comprises a multi-head self-attention module (MHSA) and a feed-forward network (FFN), and then this part is connected and fused with the other part via a 1×1 convolution, as illustrated in Figure 6. Furthermore, the PSA attention module follows (Graham et al., 2021) by assigning the dimensions of the query and key to be half of the values in the MHSA, while the PSA employs BatchNorm (Ioffe and Szegedy, 2015) to facilitate rapid inference.

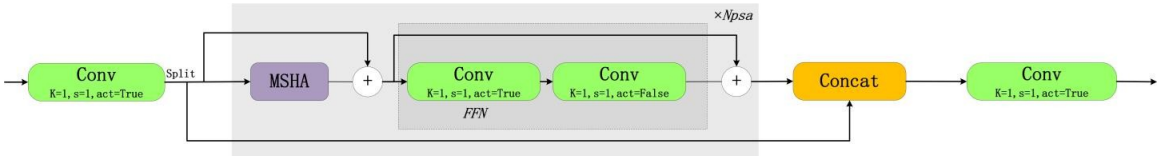

Figure 6: PSA module structure diagram

The PSA combines MHSA and FFN, leveraging their synergy to enhance the capture of local and global features. The MHSA allows for dynamic adjustment of the number of heads and the dimension of each head according to the dimensions of the input features, thereby facilitating adaptation to varying task requirements. Additionally, the PSA module, through the integration of MHSA and FFN, is capable of dynamically modeling the complex interactions between input features, making it more adaptable than some lightweight mechanisms with fixed weights.

Traffic signs are typically small targets in images, and the MHSA of PSA can effectively capture the long-range dependencies of these small targets, thereby enabling more accurate detection

and recognition. In real-world scenarios, traffic signs exhibit considerable diversity (e.g., varying scales, angles, and lighting conditions), and the dynamic feature modeling capability of PSA allows for better adaptation to these variations, enhancing the robustness of the model. Additionally, the parallel design of PSA modules, which includes both MHSA and FFN, further improves feature representation capabilities and enhances detection precision.

5. P2

The feature fusion module of YOLOv8n integrates feature maps from various levels, producing three distinct scales of feature maps: 20×20 for detecting targets larger than 32×32, 40×40 for those larger than 16×16, and 80×80 for targets larger than 8×8. These scales correspond to large, medium, and small targets, respectively. However, due to the characteristics of road traffic signs, which occupy a small proportion of the screen and contain many small targets, the original feature fusion does not effectively utilize the feature information from shallow small targets. As the depth of feature extraction increases, the fine-grained features of these small targets are often lost. Furthermore, with the expansion of the receptive field, the detection performance of the detector head for small targets significantly declines. To enhance the precision of small target detection, this paper introduces a pathway for fusing shallow feature maps into the neck and incorporates 160×160 feature maps for detecting targets larger than 4×4. Additionally, the branch P2 of the tiny target detection head, which can output larger-sized feature maps, is included. YOLOv8n consists of three output layers (P3, P4, and P5) for the object detection model; to improve small target detection capability, a fourth output layer, P2, has been added, and the structure has been adjusted as illustrated in Figure 1.

6. Experiments

To ensure the experimental data is both realistic and effective, this paper conducts 150 training rounds to achieve convergence. The experimental parameters and environmental settings are detailed in Table 1. Additionally, this study utilizes the traffic sign dataset CCTSDB2021 (Zhang et al., 2022), produced by relevant scholars and teams from Changsha University of Science and Technology, for experimentation. This dataset contains 16,356 training images and 1,500 test images. The training set is randomly divided into a training set and a validation set in a 9:1 ratio. The dataset is primarily categorized into three types: prohibitory signs (prohibitory), warning signs (warning), and mandatory signs (mandatory).

Table 1: Experimental parameters and environment

Parameters	Value	Parameters	Value	Parameters	Value
Initial learning rate	0.01	System	Windows10	Optimizer	SGD
Weight decay	0.0005	Python	3.10	Epoch	150
Momentum	0.937	Torch	2.2.2	Gpu	NVIDIA GeForce RTX 4090
Batch size	8	Batch size (test)	16		

7. Ablation Experiments

In this experiment, the metrics P, R, mAP50, mAP50-95, and Size will serve as the basis for evaluating the results. To verify the effectiveness of the proposed improvement strategy, tests will be conducted on the public dataset CCTSDB2021 test set. The experimental results are presented in Table 2. The ablation study indicates that the improved algorithm proposed in this paper achieves scores of 91.3%, 75.3%, 82.1%, and 53.4% in P, R, mAP50, and mAP50-95, respectively, all ranking first. This represents improvements of 4%, 5.8%, 4.5%, and 4.5% compared to the original algorithm. The proposed algorithm demonstrates notable performance across several evaluation metrics when compared to the original algorithm. Furthermore, it maintains high detection precision without significantly increasing the model size, resulting in a lightweight model.

Table 2: Ablation experiments

Groups	P	R	mAP50	mAP50-95	Size
Yolov8n	0.873	0.695	0.776	0.489	6.2Mb
Yolov8n-FFA-Net	0.897	0.751	0.811	0.517	6.3Mb
Yolov8n-PSA	0.89	0.692	0.772	0.475	6.8Mb
Yolov8n-RFAConv	0.884	0.713	0.792	0.498	6.3Mb
Yolov8n-P2	0.875	0.718	0.791	0.506	6.2Mb
Yolov8n-FFA-Net-PSA-P2-RFAConv	0.913	0.753	0.821	0.534	6.8Mb

8. Comparison experiments

To demonstrate the advantages of the improved algorithm presented in this paper, a comparison is made with Yolov10n. As illustrated in Table 3, the results from the comparative tests indicate that the experiment proposed herein achieves favorable detection outcomes. Furthermore, the improved algorithm demonstrates superior performance relative to Yolov10n.

Table 3: Comparison experiments

Groups	P	R	mAP50	mAP50-95	Size
Yolov8n	0.873	0.695	0.776	0.489	6.2Mb
Yolov10n	0.869	0.673	0.751	0.468	5.8Mb
Yolov8n-FFA-Net-PSA-P2-RFAConv	0.913	0.753	0.821	0.534	6.8Mb

9. Conclusion

In this paper, we propose four improvement strategies based on the YOLOv8n algorithm for the task of detecting road traffic signs under adverse weather conditions. Firstly, we introduce an improved lightweight defogging module, FFA-Net, designed to address extreme weather conditions such as fog, snow, and nighttime visibility challenges. Secondly, we detail RFAConv, a modification aimed at enhancing network performance. Additionally, we present PSA, a partially self-attention mechanism, and discuss the P2 detection layer and detection head, which significantly improve

the precision of detecting small targets. Subsequently, we conduct ablation experiments to integrate these four enhancements into the YOLOv8n model, followed by experimental analysis on the CCTSDB2021 dataset. All modifications have resulted in a significant improvement in precision.

Acknowledgments

This Research is supported by 2021 Tianjin Higher Education Science and Technology Development Fund Project (Tianjin Municipal Education Commission Project) with granted No 2021KJ021.

References

S. Ahmed, U. Kamal, and M. K. Hasan. Dfr-tsd: A deep learning based framework for robust traffic sign detection under challenging weather conditions. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):5150–5162, 2021. doi: 10.1109/TITS.2020.3048878.

B. Graham, A. El-Nouby, H. Touvron, P. Stock, A. Joulin, H. Jégou, and M. Douze. Levit: a vision transformer in convnet's clothing for faster inference. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12259–12269, 2021.

S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015.

J. A. Khan, D. Yeo, and H. Shin. New dark area sensitive tone mapping for deep learning based traffic sign recognition. *Sensors*, 18(11):3776, 2018. doi: 10.3390/s18113776.

X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11908–11915, 2020. doi: 10.1609/aaai.v34i07.6865.

A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, and J. Han. Yolov10: Real-time end-to-end object detection. In *Advances in Neural Information Processing Systems*, volume 37, pages 107984–108011, 2024. doi: 10.48550/arXiv.2405.14458.

H. Xu, Z. Zhou, D. He, F. Li, and J. Wang. Vision transformer with attention map hallucination and ffn compaction. arXiv preprint arXiv:2306.10875, 2023.

Y. Yao, L. Han, C. Du, X. Xu, and X. Jiang. Traffic sign detection algorithm based on improved yolov4-tiny. *Signal Processing: Image Communication*, 107:116783, 2022. doi: 10.1016/j.image.2022.116783.

J. Zhang, X. Zou, L. D. Kuang, J. Wang, R. S. Sherratt, and X. Yu. Cctsdb 2021: a more comprehensive traffic sign detection benchmark. *Human-centric Computing and Information Sciences*, 12, 2022. doi: 10.22967/HCIS.2022.12.023.

X. Zhang, C. Liu, D. Yang, T. Song, Y. Ye, K. Li, and Y. Song. Rfaconv: Innovating spatial attention and standard convolutional operation. arXiv preprint arXiv:2304.03198, 2023.