Proceedings of Machine Learning Research vol 278:1–11, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

Enhanced YOLOv8-Based Lightweight Algorithm for Flame Detection

Shichen Duan 1447293032@QQ.COM
College of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

Jun zhou* ZHOUJUN@SUES.EDU.CN
College of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China
**Corresponding author*

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

This paper presents an enhanced real-time lightweight fire flame detection algorithm based on improved YOLOv8n. To address the challenges of flame detection in complex scenarios, the algorithm integrates the RVB Block and EMA attention mechanism into the YOLOv8n backbone, enhancing its ability to capture fire features accurately. Additionally, a lightweight Slim-Neck structure is introduced to reduce computational complexity and parameters, facilitating embedded deployment. The proposed WiseIoU loss function further accelerates convergence and optimizes bounding box loss. Experiments demonstrate that the improved algorithm achieves a precision rate of 97.7%, a mAP@50 of 98% and a recall rate of 94.4%, with a 16% reduction in parameters and a 1.7 reduction in GFLOPs. The algorithm's lightweight nature and high accuracy provide strong technical support for early fire warning and control.

Keywords: Lightweight, RVB Block, EMA, Flame detection, Slim-Neck, WIOU.

1. Introduction

Fire, as a sudden and catastrophic disaster, is characterized by rapid spread and devastating destructive power, often causing severe casualties and property damage within a short period. Consequently, early fire detection and warning systems hold significant importance in the field of fire safety. Traditional fire detection methods primarily rely on physical devices such as smoke sensors and temperature sensors. While these approaches can achieve fire detection to a certain extent, they are susceptible to environmental interference in complex scenarios and often struggle to precisely localize the fire source. With the rapid advancement of computer vision technology, image or video-based fire detection methods have gradually emerged as a research focus.

In recent years, deep learning-based object detection algorithms have achieved remarkable progress, among which the YOLO (You Only Look Once) series has garnered significant attention due to its efficiency and real-time performance. The YOLOv8 algorithm achieves a better balance between detection accuracy and speed, making it well-suited for real-time object detection tasks. However, flame detection, as a specialized object detection task, presents numerous challenges: the shape, color, and size of flames exhibit considerable variation across different scenarios, thereby increasing the difficulty of flame detection. Moreover, in practical applications, fire detection systems are often required to be deployed on resource-constrained embedded devices (such as drones and surveillance cameras), which imposes stricter demands on the lightweight design of the algorithm.

Lightweight object detection algorithms have emerged as a prominent research focus in the field of computer vision in recent years. Numerous researchers have significantly reduced the computational burden of models through approaches such as network architecture refinement, incorporation of efficient modules, and loss function optimization. For instance, Wang et al. (2019a) proposed YOLOv4-tiny, which substantially decreases both model parameters and computational costs by simplifying the network structure and adopting depthwise separable convolutions while maintaining competitive detection accuracy. Long et al. (2020) developed PP-YOLO, which enhances model efficiency and precision by introducing a path aggregation network (PAN) and lightweight attention mechanism based on YOLOv3. Furthermore, Daoud et al. (2024) presented YOLOv8, which achieves an improved balance between detection accuracy and speed through optimized backbone and Neck modules, representing a significant advancement in lightweight object detection research.

In the domain of lightweight module design, attention mechanisms and efficient convolution operations have been extensively utilized. For instance,Wang et al. (2019b) proposed ECA-Net, which enhances feature extraction capability without increasing computational overhead through an efficient channel attention mechanism. Han et al. (2019) introduced GhostNet, employing cheap linear transformations to generate redundant feature maps, thereby significantly reducing the model's computational complexity. These lightweight designs provide crucial technical support for the practical deployment of object detection algorithms. Yuan et al. (2024) developed a lightweight detection algorithm based on YOLOv8. By incorporating a lightweight Slim-Neck module, the approach reduces computational costs while maintaining model accuracy. Additionally, through loss function optimization, the model's robustness is further improved, demonstrating superior performance in resource-constrained environments.

In the field of flame detection, lightweight design has similarly garnered extensive research attention. Numerous researchers have developed various efficient flame detection solutions by integrating lightweight object detection algorithms. For instance, Zheng et al. (2023) proposed a fire detection algorithm based on MobileNetV3, which significantly reduces computational load through the incorporation of a lightweight backbone network and attention mechanism, achieving real-time detection on embedded devices.Furthermore, Wang et al. (2024) introduced forest fire object detection algorithm based on improved yolov5s. By employing depthwise separable convolutions and channel pruning techniques, the approach effectively reduces both model parameters and computational complexity while maintaining high detection accuracy. These advancements demonstrate the effectiveness of lightweight architectures in practical flame detection applications.

Although the above research has made significant progress in lightweight flame detection, the detection accuracy and real-time performance in complex scenarios still need to be further improved. Based on the YOLOv8 algorithm and considering the characteristics of flame detection tasks, this paper proposes an improved real-time lightweight solution, aiming to provide technical support for the practical deployment of flame detection systems.

2. Improvements in YOLOv8

2.1. C2f-RVB-EMA Module

The design of the C2f module is inspired by the Cross Stage Partial (CSP) structure, which processes feature maps by splitting them into two parts. One part is directly transmitted to the next layer, while the other part undergoes convolutional operations before being merged with the directly transmitted portion. This design reduces computational load while preserving richer feature information. As shown in Figure 1a, the C2f module consists of multiple Bottleneck structures and standard convolutional layers. During training, the numerous convolutional operations within the module may lead to high feature similarity across different channels, potentially resulting in feature redundancy issues. To address this problem, this paper introduces the lightweight RepViT Block (RVB) from RepViT (Wang et al., 2023), whose structure is illustrated in Figure 1b. The RVB module effectively alleviates feature redundancy while maintaining computational efficiency through structural reparameterization and optimized feature representation.

RepViT reduces computational complexity by decreasing the channel expansion ratio while compensating for potential performance degradation through increased network width. Compared to the multiple convolutional operations in the C2f module, RepViT employs simpler early-stage convolutions as input processing modules, which not only lowers computational complexity but also improves optimization stability. As illustrated in Figure 1c, replacing the original C2f module in YOLOv8 with the proposed C2f-RVB Block allows the model to leverage multi-branch structures during training to enhance learning capability. During inference, structural reparameterization merges these branches into a single path, significantly reducing computational overhead and memory consumption. This modification maintains model performance while improving efficiency in both training and deployment phases.

Considering the research focus on flame detection, where flame morphology exhibits significant variability and is susceptible to interference from factors such as wind speed, direction, and background illumination. Therefore, we introduce the EMA (Ouyang et al., 2023) (Efficient Multi-Scale) attention mechanism, whose structure is shown in Figure 1d. This attention mechanism emphasizes preserving channel-wise information by partitioning the channel dimension into multiple sub-features while ensuring uniform distribution of spatial-semantic features across feature groups. Specifically, in addition to globally encoding information to recalibrate channel weights in each parallel branch, it further enhances feature fusion through cross-dimensional interaction between the two parallel branches. To better utilize the EMA module for smoothing feature weight functions and mitigating noise interference, we have embedded the EMA attention layer within the RepViT Block, resulting in the integrated C2f-RVB-EMA module whose detailed architecture is presented in Figure 1e.

2.2. Slim-Neck Module

The YOLOv8 algorithm employs numerous standard convolutional layers and C2f modules to enhance detection accuracy, which consequently leads to decreased inference speed and increased model parameters. The neck network, serving as the intermediary component connecting the backbone and head networks, plays a crucial role in feature integration. To reduce model complexity while maintaining detection precision, we introduce a Slim-Neck structure designed to effectively fuse multi-scale feature maps extracted from the backbone network.

This study replaces the conventional Conv modules with GSConv (Li et al., 2022) modules in the Slim-Neck structure, as illustrated in Figure 2a, effectively reducing model complexity and parameters while maintaining detection accuracy. The proposed module intelligently fuses information generated by traditional convolution with depthwise separable convolution through a shuffling operation. The specific process is as follows: First, the number of input channels is C1, and the number of channels is halved to C2/2 through standard convolution; then, the number of channels is kept unchanged through depthwise separable convolution. Subsequently, the result of the standard

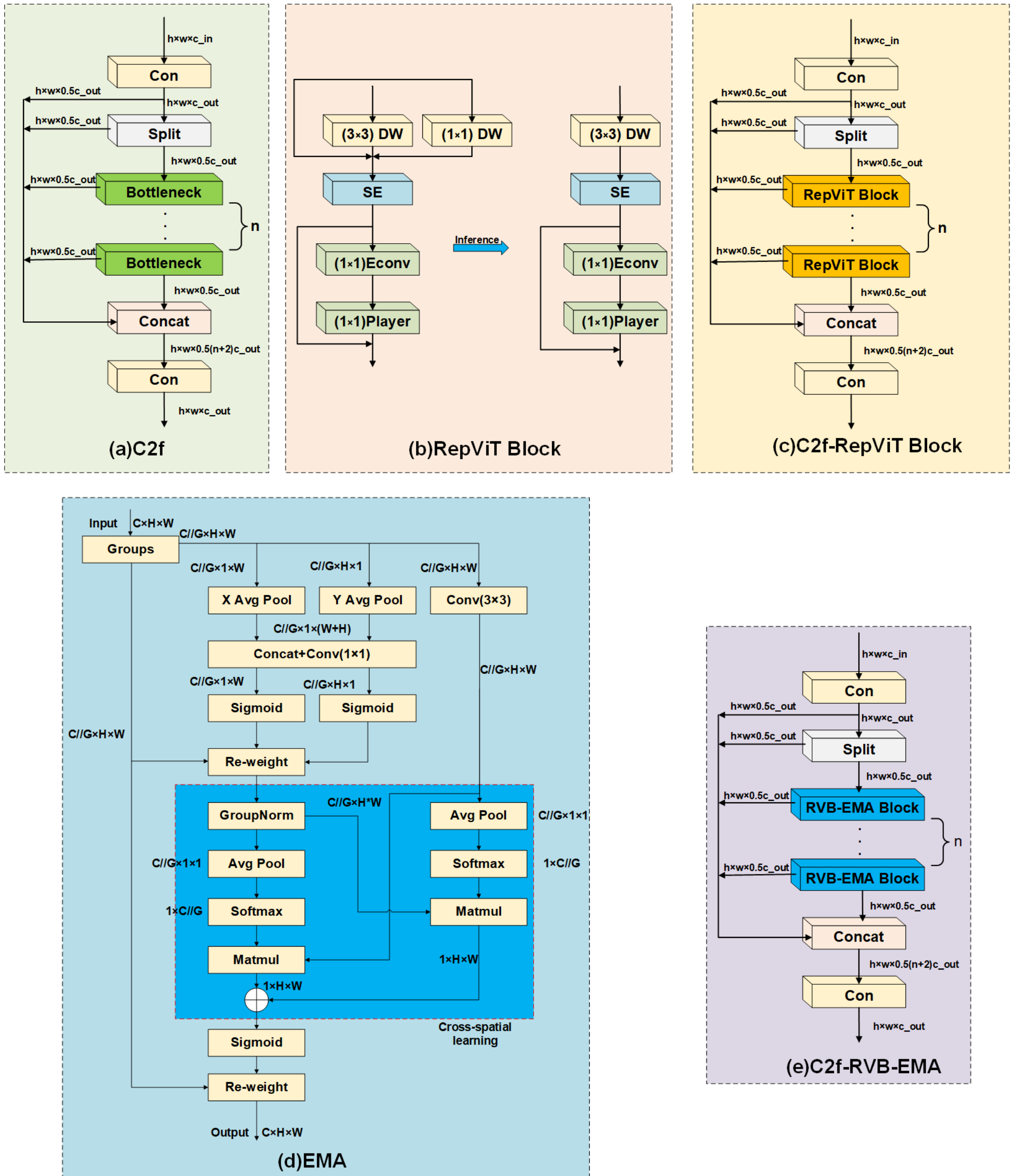

Figure 1: (a) C2f module structure, (b) RepViT Block structure, (c) C2f-RepViT Block structure, (d) EMA attention mechanism structure, (e) C2f-RVB-EMA module structure.

convolution is cascaded and shuffled with the output of the depthwise separable convolution. Finally, when performing the shuffle process, the channel information is shuffled uniformly to ensure the effective retention of multi-channel information.

Building upon GSConv, this study further incorporates the VoV-GSCSP module, whose architecture is illustrated in Figure 2b. The VoV-GSCSP module employs a one-shot aggregation approach to design the cross-stage partial network, replacing the original CSP module in the neck network. This reconstructed YOLOv8 neck architecture demonstrates enhanced capability for flame feature extraction. The introduction of the Slim-Neck structure not only reduces the complexity of the model, but also speeds up the inference speed of the model.

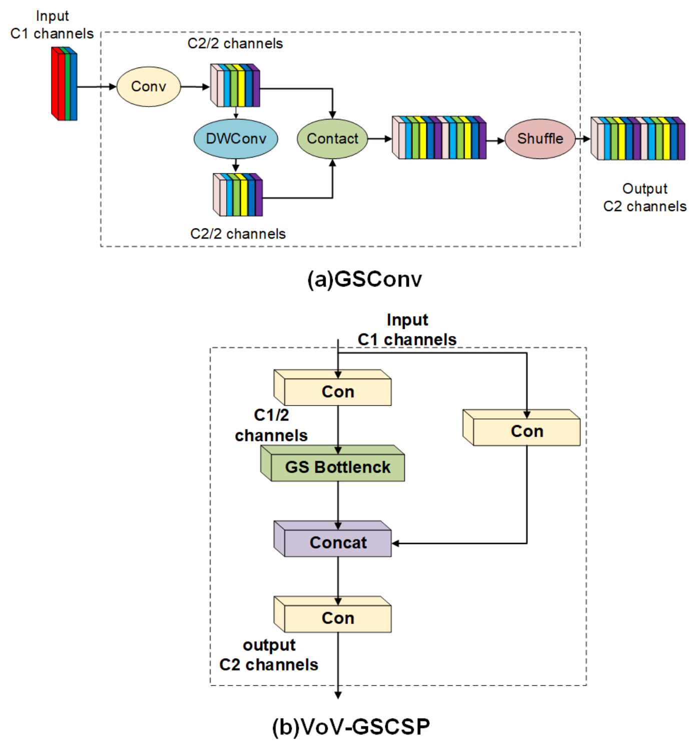

Figure 2: (a)GSConv structure, (b) VoV-GSCSP module structure.

2.3. WIoU Loss Function

This study proposes improvements to the original CIoU loss function. CIoU comprehensively considers factors such as distance, overlapping area, center point offset, and aspect ratio in bounding box regression. Compared with DIoU, which cannot be distinguished when the center points coincide, CIoU has better convergence performance. However, the CIoU loss function also has certain shortcomings. First, it does not consider the problem of direction mismatch between the ground truth box and the prediction box. In addition, low-quality regression samples have a greater impact on the loss function, while high-quality regression samples are difficult to further optimize. These two factors lead to slow convergence speed and low efficiency during model training, and insufficient accuracy of the prediction box.

To address the limitations of the CIoU loss function, this study adopts the Wise-IoU v3 (WIoU v3) loss function (Tong et al., 2023) as a superior alternative. When dealing with datasets containing lower annotation quality, WIoU v3 reduces emphasis on high-quality anchor boxes while increasing attention to medium-quality anchors, thereby improving overall detection performance. WIoU v3 introduces a non-monotonic focusing coefficient based on β. A non-monotonic focusing coefficient is constructed through β and applied to WIoU v1. The specific formula is shown in formula (4). This mechanism assigns smaller gradient gains to anchors with lower outlier degrees (typically higher quality), thereby shifting the regression focus toward medium-quality anchors. Conversely, anchors exhibiting higher outlier degrees receive reduced gradient gains, effectively mitigating the negative impact of potentially harmful gradients from low-quality samples.

$$\mathcal{L}_{\text{WIoUv }3} = r\mathcal{L}_{\text{WIoUv }1}, r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \tag{1}$$

The WIoU v3 loss function demonstrates superior computational efficiency compared to CIoU by eliminating the aspect ratio calculation in its design. Furthermore, WIoU v3 incorporates a dynamic non-monotonic focusing mechanism that evaluates anchor quality through "outlier degree" rather than traditional IoU metrics, while integrating a gradient gain allocation strategy to further optimize model performance. The outlier degree of an anchor box is quantitatively expressed as the ratio between $\mathcal{L}^*_{IoU}$ and $\overline{\mathcal{L}_{\text{IOU}}}$, as formally defined in Equation (5).

$$\beta = \frac{\mathcal{L}^*_{\text{IoU}}}{\overline{\mathcal{L}_{\text{IoU}}}} \in [0, +\infty) \tag{2}$$

The WiseIoU loss function effectively addresses the limitations of CIoU by optimizing bounding box regression, thereby significantly improving the network's convergence efficiency.

3. Experiment Preparation

3.1. Training Environment and Parameters

The experiments in this study were conducted using Python programming language, PyTorch 2.1 and CUDA 11.8 framework. During the training process, the hyperparameters were carefully adjusted according to the research objectives, as specified in Table 1.

Table 1: Training hyperparameter specifications.

Epochs	300
Batch Size	32
Initial learning rate	0.01
Momentum	0.937
Optimization algorithm	SGD

3.2. Flame Dataset

Given the limited availability of public flame datasets, this study employed web crawling techniques and field experiments to collect flame images across diverse scenarios, including urban structures and indoor/outdoor environments. Due to the constrained sample size obtained, data augmentation methods were systematically applied to expand the existing dataset. The enhanced dataset comprising 4,910 images was partitioned in a 7:2:1 ratio, allocating 3,437 images for training, 982 for validation, and 491 for testing purposes.

3.3. Evaluation Metrics

To evaluate the detection performance of the proposed algorithm improvements, this study employs mean Average Precision (mAP@50), Precision (P), and Recall (R) as primary evaluation metrics. Considering the introduced lightweight design, additional metrics including model parameters and GFLOPs are also incorporated for comprehensive assessment. The computational formulas for these metrics are presented in Equations (3-5):

$$mAP = \frac{1}{N}\sum_{i=1}^{n} AP_i \tag{3}$$

Here, mAP@50 refers to the mean Average Precision calculated when the Intersection over Union (IoU) threshold is set to 0.5. Specifically, it computes the Average Precision (AP) for each class (representing the proportion of correct predictions among all samples) and then averages these values across all categories. A higher mAP value indicates superior network performance.

$$\text{Precision } = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Recall } = \frac{TP}{TP + FN} \tag{5}$$

Among them, TP (True Positives) represents the number of correctly predicted positive samples, FN (False Negatives) denotes the number of incorrectly predicted negative samples, and FP (False Positives) indicates the number of non-target samples erroneously predicted as flames. Precision (P) quantifies the proportion of correctly identified positives among all predicted positives, while Recall (R) measures the fraction of actual positives that are correctly identified. Regarding computational efficiency, lower GFLOPs values indicate reduced computational complexity, which typically corresponds to decreased parameter counts.

4. Experiment Results

4.1. Ablation Experiment

To validate the contribution of the proposed modules to model performance, this study conducted five ablation experiments for comparative analysis. By systematically adding or removing key modules, we evaluated each module's impact on detection accuracy, model complexity, and parameters. All experiments were performed under identical training and testing conditions, and the results are shown in Table 2.

In the ablation experiments, Model A employed the RepViT Block (C2f-RVB) from RepViT to replace the original C2f module in YOLOv8. While this substitution resulted in a slight degradation in detection performance, it successfully achieved network lightweighting. Building upon Model A, Model B further incorporated the EMA attention mechanism to enhance feature extraction capability. This enhanced architecture maintained comparable detection performance to the baseline YOLOv8 while preserving its lightweight characteristics. Model C introduced the lightweight Slim-Neck structure to address the performance trade-off observed in the C2f-RVB module. By replacing YOLOv8's original neck network with this optimized architecture, the modified algorithm not only retained its lightweight advantages but also demonstrated improved detection performance. Model D, combined both the C2f-RVB-EMA module and Slim-Neck structure. This integrated approach effectively balanced the competing objectives of model lightweighting and detection accuracy, leveraging the complementary strengths of each component to achieve optimal overall performance. Finally, Model E further optimizes the architecture by adopting the WIoU loss function to replace the original CIoU, aiming to minimize bounding box regression errors, accelerate convergence speed, and enhance the localization accuracy of detected targets.

Table 2: Results of ablation experiments.

Model	C2f-RVB	C2f-RVB-EMA	Slim-Neck	WIoU	P/%	R/%	mAP@50/%	Parameters	GFLOPs
YOLOv8n					97.4	92.9	97	3005843	8.1
A	✓				96.9	92.2	96.7	2282979	6.3
B		✓			97	92.9	97	2286947	6.4
C			✓		97.1	93.5	97.1	2795859	7.3
D		✓	✓		97.5	94.1	97.9	2519651	6.4
E	✓	✓	✓		**97.7**	**94.4**	**98**	**2519651**	**6.4**

As evidenced by Table 2, the improved YOLOv8 algorithm demonstrates superior detection performance compared to the original YOLOv8 while achieving significant lightweighting. Specifically, the model exhibits a 1% increase in mean Average Precision (mAP@50), 0.3% improvement in Precision (P), and 1.5% enhancement in Recall. Regarding computational efficiency, the modified architecture reduces parameter count by 16% and decreases GFLOPs by 1.7. Comparative results in Figure 3 further confirm that the enhanced YOLOv8 outperforms its baseline counterpart in both mAP@50 and loss value metrics.

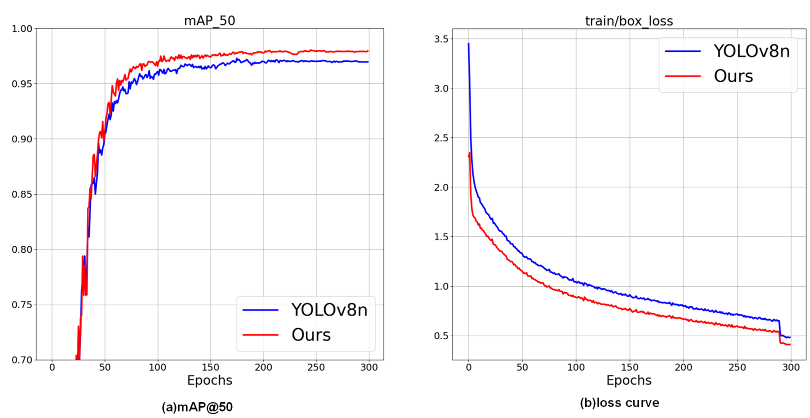

Figure 3: Comparative results of ablation experiments (a)mAP@50 comparison chart, (b)loss curve comparison chart.

It can be seen that the improved YOLOv8 has the following improvements. First, the detection accuracy is improved. Second, the false detection rate is reduced, and small targets are better identified. Finally, while the accuracy is improved to a certain extent, the detection frame is made more accurate. These comparative performance enhancements are visually demonstrated in Figure 4.

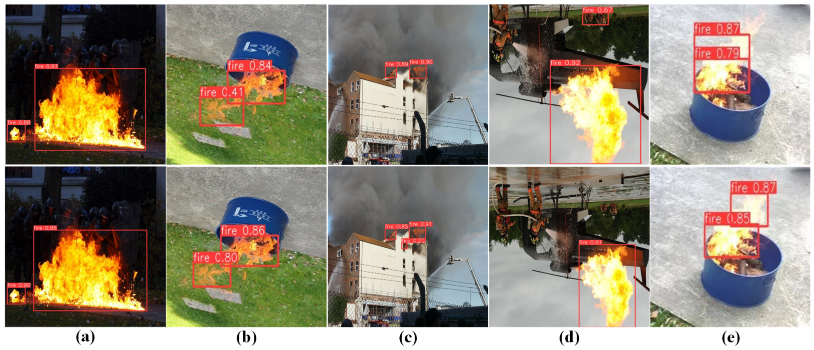

Figure 4: Detection performance comparison between YOLOv8 and the improved YOLOv8 (Row 1: Original YOLOv8 results; Row 2: Improved YOLOv8 results). (a)(b) Detection accuracy; (c)(d) False detection rate and small object detection; (e) Bounding box precision.

4.2. Comparative Experiment

To comprehensively evaluate the performance of the improved lightweight YOLOv8 algorithm, this study conducted comparative experiments with several mainstream object detection algorithms. The experimental results are presented in Table 3.

Table 3: Performance comparison with mainstream algorithms.

Model	P/%	R/%	mAP@50/%	Parameters	GFLOPs
SSD	82.3	72.8	78.8	26151824	62.7
YOLOv3-tiny	94.6	91.1	96.1	8666692	12.9
YOLOv5n	95.7	92.9	96.5	2649200	7.1
YOLOv6n	96.5	91.5	96.1	4233843	11.8
YOLOv7	95.7	95.4	97.2	37196556	105.1
YOLOv8n	97.4	92.9	97	3005843	8.1
YOLOv9-t	96.2	94.1	97.2	1970979	7.6
YOLOv10n	96.2	92.6	96.9	2694806	8.2
Ours	**97.7**	**94.4**	**98**	**2519651**	**6.4**

The study initially compares the proposed YOLOv8 algorithm with the classic one-stage SSD (Liu et al., 2015) algorithm, demonstrating superior performance across all evaluation metrics. Subsequent comparisons with YOLO series variants reveal significant findings: while YOLOv3-tiny, YOLOv6n, and YOLOv7 exhibit substantially increased parameter counts and GFLOPs, they fail to deliver corresponding performance improvements. Furthermore, although YOLOv5n, YOLOv9-t and YOLOv10n achieve model lightweighting compared to the original YOLOv8n, these architectures compromise detection performance. In contrast, our improved lightweight YOLOv8 algorithm not only maintains reduced computational complexity but also enhances detection accuracy. The optimized YOLOv8 demonstrates exceptional performance across all evaluation dimensions, with a compact model size of merely 5.15MB after lightweighting - meeting stringent deployment requirements on resource-constrained devices. These comprehensive experiments confirm the proposed model's superior effectiveness for flame detection applications.

5. Conclusion

This paper proposes an improved lightweight flame detection algorithm based on YOLOv8, aiming to enhance both detection accuracy and real-time performance. The key modifications include replacing the original C2f module in the backbone network with a more efficient RepViT Block integrated with EMA attention mechanism, which reduces noise interference while improving feature extraction capability. For lightweight processing, the Slim-Neck structure is adopted in the neck network, achieving model compression without sacrificing detection accuracy and facilitating deployment on resource-constrained devices. Furthermore, the WiseIoU loss function is introduced to accelerate convergence and optimize bounding box regression specifically tailored for flame detection. The improved YOLOv8 algorithm demonstrates superior detection performance on experimental datasets, particularly showing gains in mAP@50 while maintaining reduced parameters and GFLOPs. Compared with current mainstream detection algorithms, the enhanced YOLOv8 not only exhibits stronger feature representation capabilities but also offers compact model size and high processing speed, meeting the requirements for real-time fire detection in practical scenarios. Additionally, considering the complexity and variability of real-world environments, subsequent work will focus on enhancing the algorithm's robustness to ensure reliable performance across diverse application scenarios.

References

Zeineb Daoud, Amal Ben Hamida, and Chokri Ben Amar. Fire object detection and tracking based on deep learning model and kalman filter. *Arabian Journal for Science and Engineering*, 49(3): 3651–3669, 2024. doi: 10.1007/s13369-023-08127-7.

Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. Ghostnet: More features from cheap operations. *CoRR*, abs/1911.11907, 2019.

Hulin Li, Jun Li, Hanbing Wei, and et al. Slim-neck by gsconv: a lightweight-design for real-time detector architectures. *J. Real Time Image Process.*, 21:62, 2022.

Wei Liu, Dragomir Anguelov, Dumitru Erhan, and et al. SSD: single shot multibox detector. *CoRR*, abs/1512.02325, 2015.

Xiang Long, Kaipeng Deng, Guanzhong Wang, and et al. PP-YOLO: an effective and efficient implementation of object detector. *CoRR*, abs/2007.12099, 2020.

Daliang Ouyang, Su He, Jian Zhan, and et al. Efficient multi-scale attention module with cross-spatial learning. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023.

Zanjia Tong, Yuhang Chen, Zewei Xu, and Rong Yu. Wise-iou: Bounding box regression loss with dynamic focusing mechanism. *ArXiv*, abs/2301.10051, 2023.

Ao Wang, Hui Chen, Zijia Lin, Hengjun Pu, and Guiguang Ding. Rep vit: Revisiting mobile cnn from vit perspective. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15909–15920, 2023.

Chien-Yao Wang, Hong-Yuan Mark Liao, I-Hau Yeh, and et al. Cspnet: A new backbone that can enhance learning capability of CNN. *CoRR*, abs/1911.11929, 2019a.

Jianan Wang, Changzhong Wang, Weiping Ding, and Cheng Li. Yolov5s-ace: Forest fire object detection algorithm based on improved yolov5s. *Fire Technology*, 60(6):4023–4043, 2024. doi: 10.1007/s10694-024-01619-4.

Qilong Wang, Banggu Wu, Pengfei Zhu, and et al. Eca-net: Efficient channel attention for deep convolutional neural networks. *CoRR*, abs/1910.03151, 2019b.

Zijian Yuan, Pengwei Shao, Jinran Li, and et al. Yolov8-acu: improved yolov8-pose for facial acupoint detection. *Frontiers in Neurorobotics*, Volume 18 - 2024, 2024. doi: 10.3389/fnbot.2024.1355857.

Hongtao Zheng, Junchen Duan, Yu Dong, and Yan Liu. Real-time fire detection algorithms running on small embedded devices based on mobilenetv3 and yolov4. *Fire Ecology*, 19, 05 2023. doi: 10.1186/s42408-023-00189-0.