Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3065

FOREST FIRE POINT RECOGNITION BASED ON SUPER-RESOLUTION TECHNIQUES

FuLin Li, WenFa Xu*, Zhen Min, XuePeng Wu, ChangYu Xiang, YuDong Liu
School of Information Science and Engineering, Wuchang Shouyi University, Wuhan 430064, Hubei, China.
Corresponding Author: WenFa Xu, Email: xuwenfa@wsyu.edu.cn

Abstract: Forest fires, as a global ecological disaster, pose a serious threat to the stability of environmental systems, biodiversity security, and socio-economic development. In response to the shortcomings of traditional fire point detection methods in small target recognition accuracy, model robustness, and real-time performance, this paper proposes an intelligent forest fire point detection model based on an improved YOLOv8n framework. This model integrates the Fast Super Resolution Convolutional Neural Network feature enhancement module, Enhanced Squeeze Excitation attention mechanism, and an improved Minimum Point Distance Intersection over Union bounding box regression algorithm, aiming to improve the detection ability of early fire points and overall system performance. Through systematic experiments on a self built multi scene forest fire image dataset, the results showed that compared with Faster R-CNN, YOLOv5s, and the standard YOLOv8n model, the proposed method performed well in comprehensive detection performance, with mAP reaching 84.7%, Precision reaching 84.2%, Recall reaching 80.4%, and also possessing high real-time processing capabilities. This study not only provides effective technical support for intelligent monitoring of forest fires, but also proposes a multi module collaborative optimization framework, which provides new theoretical references and practical paths for research and application in the field of small target detection.
Keywords: Forest fire point recognition; Super-resolution; YOLOv8n; Object detection; Deep learning

1 INTRODUCTION

Forests, as one of the most important terrestrial ecosystems on Earth, play an irreplaceable role in global carbon cycling, climate regulation, and biodiversity conservation [1]. However, due to the impact of global climate change, forest fires have been frequent and expanding in scale in recent years. The "Black Summer" forest fires in Australia in 2020 burned over 24 million hectares of forest land, causing nearly 3 billion animals to die or be displaced; In 2021, a forest fire broke out in Liangshan Prefecture, Sichuan Province, China, causing dozens of casualties and damage to tens of thousands of acres of forest land. According to a report released by the United Nations Environment Programme in 2022, the direct economic losses caused by forest fires worldwide exceed $50 billion annually, and ecological losses are difficult to estimate.

How to achieve early monitoring and rapid response to forest fires has gradually become an important research direction for modern intelligent forestry and emergency response. Early intelligent fire monitoring methods mainly relied on the color features, temperature information, and smoke dynamics of remote sensing images for fire point extraction. Algorithms included color thresholding, background subtraction, and thermal infrared recognition [2-3], but these methods generally had problems such as high false alarm rates and strong dependence on the environment [4].

With the rapid development of deep learning technology, object detection algorithms based on convolutional neural networks have been widely explored and applied in the field of forest fire point recognition. Early research mainly used Faster R-CNN [5] and YOLO series models, among which YOLO demonstrated significant engineering applicability in multi-source heterogeneous data scenarios of unmanned aerial vehicle aerial video and fixed monitoring equipment due to its real-time advantage of single-stage detection [6]. In response to the technical difficulties of small scale and weak features of forest fire point targets, researchers have introduced super-resolution reconstruction algorithms such as FSRCNN and EDSR to construct a preprocessing module, effectively improving the spatial resolution of input images and enhancing the model's ability to extract features from small-scale fire points. It is worth noting that in recent years, visual Transformer architecture has shown breakthrough progress in the field of fire detection. Swin Transformer has achieved global context modeling through a hierarchical window attention mechanism, while DETR has significantly improved algorithm robustness under complex background interference based on an end-to-end detection framework.

Although forest fire point recognition technology based on deep learning has made significant progress in image understanding and object detection, it still faces many challenges in practical applications. Firstly, forest fire points often exhibit extremely small early scales, irregular shapes, and blurred edges, which are easily disturbed by complex backgrounds, making it difficult for detection models to accurately capture their features and prone to missed and false detections. Secondly, due to the suddenness and danger of forest fires, there are safety hazards in the process of obtaining real fire image data, and early fire point samples are extremely scarce, resulting in uneven sample distribution in most datasets, especially the lack of diverse small fire point instances, which affects the training quality and generalization ability of the model. In addition, although lightweight object detectors such as YOLOv5 have been applied in fire point recognition, their actual deployment on edge devices such as drones and front-end monitoring terminals is still limited by factors such as computing power, power consumption, and transmission delay, making it difficult to balance high accuracy and real-time performance.

Therefore, improving the ability to detect small targets, building high-quality multi scene datasets, and implementing deployment solutions that balance lightweight and accuracy will be key directions for future fire point recognition research. In order to solve the above problems, this paper proposes a forest fire point recognition method based on YOLOv8 object recognition algorithm and fused with image super-resolution. This method improves the detection accuracy of fire points while being lightweight, providing new ideas for the field of fire point recognition.

2 REAL TIME YOLO OBJECT DETECTION WITH FSRCNN FUSION

YOLOv8 is the latest generation object detection model in the YOLO series launched by the Ultralytics team in 2023. As an unofficial continuation and architectural innovation version of YOLOv5, YOLOv8 has been systematically optimized in model design, training strategy, and inference efficiency. YOLOv8 supports tasks including object detection, image segmentation, and pose estimation, making it an ideal choice for achieving tasks such as object detection. This article has made the following improvements to the YOLOv8 model. The introduction of FSRCNN (Fast Super Solution Convolutional Neural Network) proposed by Dong et al. on ECCV in 2016 [7] for image super-resolution reconstruction has improved the recognition accuracy of small targets; To enhance the feature representation of fire points, we drew inspiration from the eSE attention mechanism module proposed by Gao et al. in 2021 [8] and added a feature optimization module that can explicitly model channel relationships; At the same time, the MPDloU bounding box regression loss function proposed by Li Yang et al. in 2022 [9] was introduced to improve YOLOv8, thereby enhancing the model's ability to detect low resolution fire points and achieving higher accuracy and stronger robustness in forest fire point recognition.

2.1 Hyperfractionation Module

As an improved version of SRCNN, the FSRCNN model not only achieves better reconstruction quality, but also significantly improves computational efficiency. The core innovation of FSRCNN lies in abandoning the strategy of SRCNN first interpolating and then convolving, and instead directly extracting features from low resolution images, and achieving resolution improvement at the end of the network through deconvolution operation, effectively reducing the redundancy of previous calculations. In addition, FSRCNN introduces a five stage structure of "feature extraction reduction nonlinear mapping expansion deconvolution", which significantly reduces computational complexity and improves nonlinear expression ability by adding a 1x1 convolution bottleneck module in the middle layer. The grid structure diagram of the FSRCNN super-resolution model is shown in Figure 1.

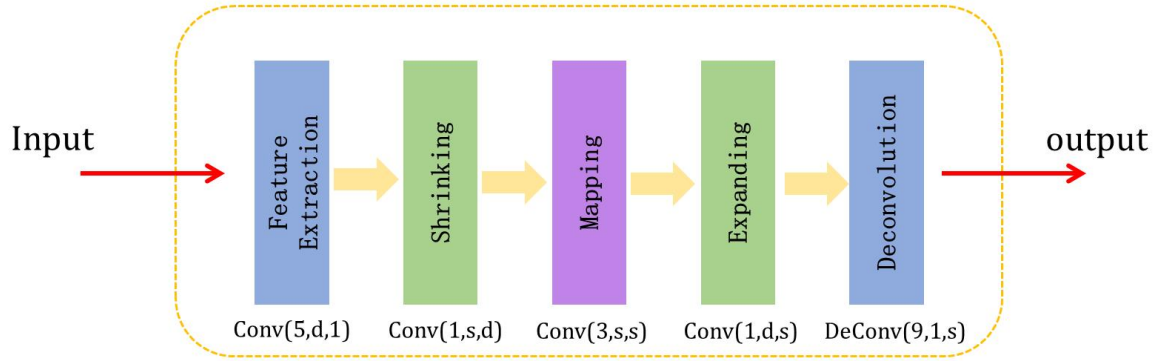

Figure 1 FSRCNN Grid Structure Diagram
Note: Feature Extraction is the feature extraction layer, Shrinking is the feature compression layer, Mapping is the non-linear mapping layer, Expanding is the channel expansion layer, and Deconvolution is the deconvolution/upsampling layer

FSRCNN has lower latency while maintaining good image reconstruction quality, making it particularly suitable for real-time visual applications that are sensitive to computing resources. In subsequent tasks such as object detection and small object recognition, FSRCNN is also widely used in the image preprocessing stage to improve image clarity and detail expression ability, thereby enhancing the detection model's ability to recognize fine-grained objects.
The core idea of FSRCNN is to directly learn end-to-end from low resolution images, using shallow convolutional networks to learn the mapping relationship from low resolution to high-resolution images, avoiding the multi-stage image reconstruction process in traditional super-resolution methods. This model adopts feature mapping reduction processing to reduce computational complexity, and then upsampling to restore high-resolution images, significantly improving computational efficiency and processing speed. Its mathematical expression is:

$$I_{SR} = \int (I_{LR}; \theta) \quad (1)$$

Among them θ is the grid parameter, $\int(\cdot)$ is the network, I_{LR} is the input image, I_{SR} is the input image.
Compared to other models, FSRCNN adopts a more efficient design, including simplified feature extraction and upsampling processes, which significantly accelerates inference speed while maintaining high image reconstruction

quality.

According to Dongetal's research "Accelerating the Super Solution Neural Network"[10], the speed improvement of FSRCNN compared to SRCNN is mainly due to the following reasons: (1) reducing the dimensionality and computational complexity of the input image through feature dimensionality reduction; (2) By optimizing the deconvolution operation in the upsampling stage, the computational cost has been reduced; (3) Through end-to-end training, the redundant calculation process of the model has been reduced. The experimental results in the literature indicate that FSRCNN improves speed several times compared to SRCNN and maintains a high level of performance in image quality.

Therefore, integrating the FSRCNN super-resolution model into the YOLO object recognition model can effectively improve the quality of the input image, enabling the model to perform accurate recognition even in the face of low resolution. At the same time, the lightweight design of FSRCNN can also enable the fusion model to maintain high accuracy while having good response speed.

2.2 eSE Attention Module

ESE attention mechanism (Efficient Squeeze and Excitation Attention Mechanism) is a technique for optimizing the channel attention mechanism in convolutional neural networks, aiming to improve the expression ability and efficiency of the model. The eSE attention mechanism models the dependency relationships between channels to adaptively adjust the weights of different channels, enhancing the network's attention to useful features while reducing the impact of unimportant information. Compared with traditional SE mechanisms, eSE significantly improves computational efficiency and reduces computational complexity while maintaining performance through further optimization of computation and parameter quantity.

The core idea of eSE attention mechanism is to guide the network to focus on important features and suppress irrelevant or redundant information by weighting the channels of the feature map. The biggest improvement of the eSE mechanism lies in its optimization of the traditional SE mechanism. Specifically, in the Excitation stage, eSE uses a more efficient operation, avoiding the computational complexity of the fully connected layer in the traditional SE mechanism.

The specific principle of eSE attention mechanism is shown in Figure 2. The input image is first fed into an average pooling Favg, then into a 1x1 convolutional layer Wc, and finally, output feature maps with different weights are obtained through an h-sigmoid activation function.

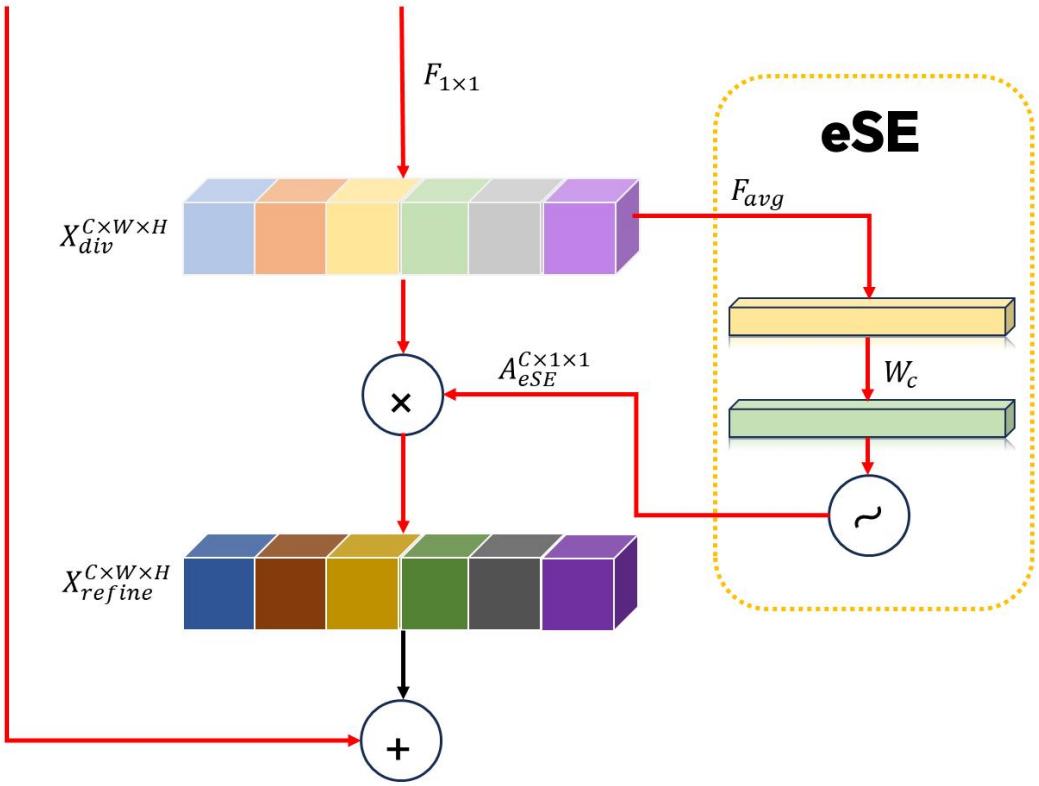

Figure 2 eSE Principle Diagram of Attention Mechanism

In addition, the eSE attention mechanism optimizes the structure of traditional SE modules by enhancing the modeling of inter channel dependencies. The specific mathematical process can be expressed as follows:

2.2.1 Input feature map:

Given an input feature map $X \in R^{C\times W\times H}$, where is the number of channels and is the spatial dimension。

2.2.2 Global average pooling：

Compress the space of each channel and generate channel descriptors：

$$E_{avg}(X_c) = \frac{1}{W\times H}\sum_{i=1}^{W}\sum_{j=1}^{H} X_c(i,j)\ ,\ \ c = 1,2,\ldots,C \quad (2)$$

Output vector： $\boldsymbol{z \in R^{C\times 1\times 1}}$。

2.2.3 Channel weight learning：

Through an uncompressed **1x1** convolutional layer $\boldsymbol{W_c}$（Preserve channel dimension）and h-sigmoid activation function generates attention weights：

$$A_{eSE} = \sigma_h(W_c z),\ \ W_c \in R^{C\times C},\ \ \sigma_h(\cdot) = h - sigmoid \quad (3)$$

Among them, $\boldsymbol{\sigma_h}$ is the improved sigmoid function (such as Hard Sigmoid) enhances nonlinearity and computational efficiency.

2.2.4 Feature recalibration：

Multiply the weight $\boldsymbol{A_{eSE}}$ with the input feature map channel by channel to obtain the refined output：

$$X_{refine} = X \odot A_{eSE} \quad (4)$$

⊙ represents channel wise multiplication.

In summary, the eSE attention mechanism, as an improved channel attention module, enhances the important features of fire points and suppresses redundant information by dynamically recalibrating the fire point images, significantly reducing the computational complexity of the model while maintaining high performance.

2.3 MPD-IoU Bounding Box Regression Loss Function

MPD IoU is a loss function used for bounding box regression in object detection tasks, designed to improve the fit between detection boxes and real boxes, while enhancing the robustness of the model to difficult to locate samples. Figure 3 shows the schematic diagram of the principle of MPD IoU.

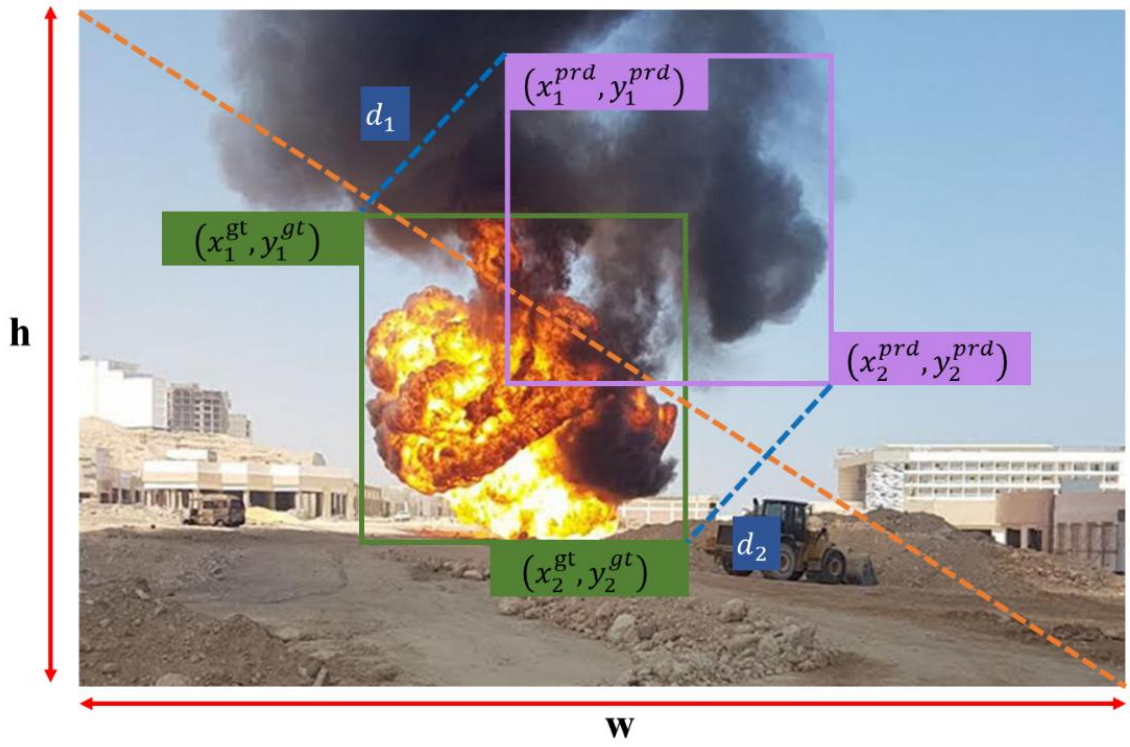

Figure 3 Schematic Diagram of MPD IoU Principle

Note: The purple box represents the prediction box, and the green box represents the target box. (x_1^{gt}, y_1^{gt}) represents the upper left corner coordinate of the annotation box; (x_2^{gt}, y_2^{gt}) represents the coordinates of the upper right corner of the annotation box; (x_2^{gt}, y_2^{gt}) represents the coordinates of the upper right corner of the annotation box; (x_2^{gt}, y_2^{gt}) represents the coordinates of the upper right corner of the annotation box; $\overline{d_1}$ represents the minimum point distance between the annotation box and the prediction box, which is the offset between the diagonals of the two boxes (shown by the blue dashed line in the figure).

Unlike traditional IoU loss functions, MPD IoU integrates multidimensional penalty factors such as center offset and aspect ratio consistency, and combines weighted average IoU with sample dynamic penalty mechanism to significantly

improve the boundary fitting ability for high aspect ratio targets and fuzzy boundary targets. In the detection of small and irregular targets, this loss function has higher gradient sensitivity, effectively alleviating the problem of gradient vanishing during the training process. Meanwhile, MPD loU optimizes bounding box regression by directly minimizing the offset distance between the predicted box, annotated box, and the four corner points of the annotated box, rather than relying on traditional LOU or geometric constraints. The loss function can be expressed as:

$$\zeta_{MPDloU} = 1 - loU + \lambda \cdot \overline{d_1} \quad (5)$$

Among them, loU represents the intersection ratio; $\overline{d_1}$ represents the normalized value of the distance between diagonal points (key innovation point); λ represents balanced weight

Introducing MPD IoU into the fire point detection framework effectively alleviates the gradient degradation and localization error problems of traditional IoU loss in small object detection. The fusion of dynamic penalty and multi-dimensional geometric modeling mechanism significantly improves the fitting accuracy and stability of the model to the bounding box of fire points, especially suitable for complex remote sensing scenes with sparse fire point samples, significant scale differences, and strong noise interference.

In summary, the introduction of MPD IoU can further improve the localization accuracy of regression branches in the fire point detection framework, enhance the model's adaptability to high deformation and multi-scale targets in fire scenes, and achieve more reliable flame and smoke detection results.

3 EXPERIMENT AND RESULT ANALYSIS

3.1 Dataset Establishment

Due to the limited availability of credible open source fire datasets and the lack of annotation in most existing datasets, it is difficult to directly use them for training. Therefore, we constructed a fire dataset. This dataset covers scenarios such as forest fires, indoor fires, traffic fires, building fires, outdoor fires, and candle flames. As shown in Table 1, the dataset contains 10295 images, divided in a 14:3:3 ratio. The training set contains 7206 images, while the validation set and test set each contain 1544 images. The specific distribution is shown in Table 1.

Table 1 Distribution of Fire Point Dataset

category	Data set	Training set	Validation set	Test set
Quantity (in sheets)	10,295	7,206	1,544	1,544

3.2 Experimental Environment

This research experiment was conducted in the following hardware and software environments: the hardware platform used high-performance computing devices equipped with NVIDIA RTX 3090 graphics cards, providing powerful parallel computing capabilities for deep learning tasks. In terms of software environment, the experiment was built on the Windows 10 operating system, using Python 3.8 as the programming language environment, and implementing algorithm design and model training using the PyTorch 2.1.0 deep learning framework. YOLOv8n was used as the core training tool for the object detection model, with an input image size of 512 × 512, a model training period of 300, a batch size of 16, an initial learning rate of 0.01, and a cosine annealing strategy to dynamically adjust the learning rate.

3.3 Evaluation

Adopt the following four indicators：

$$Presicion = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$mAP = \frac{\sum P_A}{N_c} \quad (8)$$

$$mloU = \frac{1}{N}\sum_{i=1}^{N} IoU_i \quad (9)$$

TP, FP, and FN are used to correctly detect, incorrectly detect, and miss detection targets, respectively. They represent the similarity measure between a single predicted result and the true annotation. The average value of all categories is used to evaluate the efficiency of the network. At the same time, the number of image frames processed per second is selected to evaluate the effectiveness of the network in the application scenario of fire point image recognition.

3.4 Ablation experiment

In order to verify the impact of each module on overall performance, this paper conducted a systematic ablation experiment and set the following model combinations：

1. YOLOv8n
2. YOLOv8n+FSRCNN
3. YOLOv8n+FSRCNN+eSE
4. YOLOv8n+FSRCNN+eSE+MPDloU

Table 2 shows the performance of each model on the validation set:

Table 2 Data of ablation experiment results

Number	FSRCNN	eSE	MPDIoU	mAP	Precision	Recall	MIoU	FPS/ms
1	×	×	×	0.78	0.81	0.77	0.72	9.2
2	√	×	×	0.80	0.83	0.79	0.71	12.4
3	√	√	×	0.83	0.84	0.81	0.79	12.7
4	√	√	√	0.84	0.84	0.80	0.81	13.1

Note: √ indicates the introduction of the module, x indicates the non introduction of the module

According to the data in the table above, the introduction of each module has a significant improvement in model performance. Specifically, number 1 only introduces the FSRCNN module, which increases mAP and Precision to 0.78 and 0.81 respectively compared to the baseline, but slightly lowers Recall and MIoU. Number 2 added eSE module on this basis, increasing mAP and Precision to 0.80 and 0.83 respectively, and Recall also slightly increased to 0.79. Number 3 introduces the MPDIoU module on the original basis, further increasing mAP to 0.83 and significantly improving all indicators. In the end, Number 4 integrated all three modules and achieved the best overall performance, with mAP reaching 0.84, Precision and Recall being 0.84 and 0.80, respectively, and MIoU improving to 0.81 while maintaining a high inference speed.

It can be seen that introducing the FSRCNN module alone can effectively improve detection accuracy, but there is a certain trade-off between recall and inference speed. After further introducing the eSE module, the overall performance of the model steadily improved, indicating that the attention mechanism is helpful for feature extraction. After combining the MPDIoU module, not only is the detection accuracy further improved, but the positioning accuracy is also significantly enhanced. When integrating all modules, the model achieved the best performance in all indicators, verifying the effectiveness and superiority of the proposed multi module fusion strategy in improving detection accuracy and maintaining high inference speed.

3.5 Visual Comparison of Model Parameters

To further verify the accuracy improvement effect of the model in this article, a visual parameter comparison was also conducted. The result is shown in Figure 4. The green box represents the prediction results of the YOLOv8n model, the red box represents the prediction results of the improved model in this paper, and the blue arrow indicates the differences in the block diagrams between the previous and subsequent models.

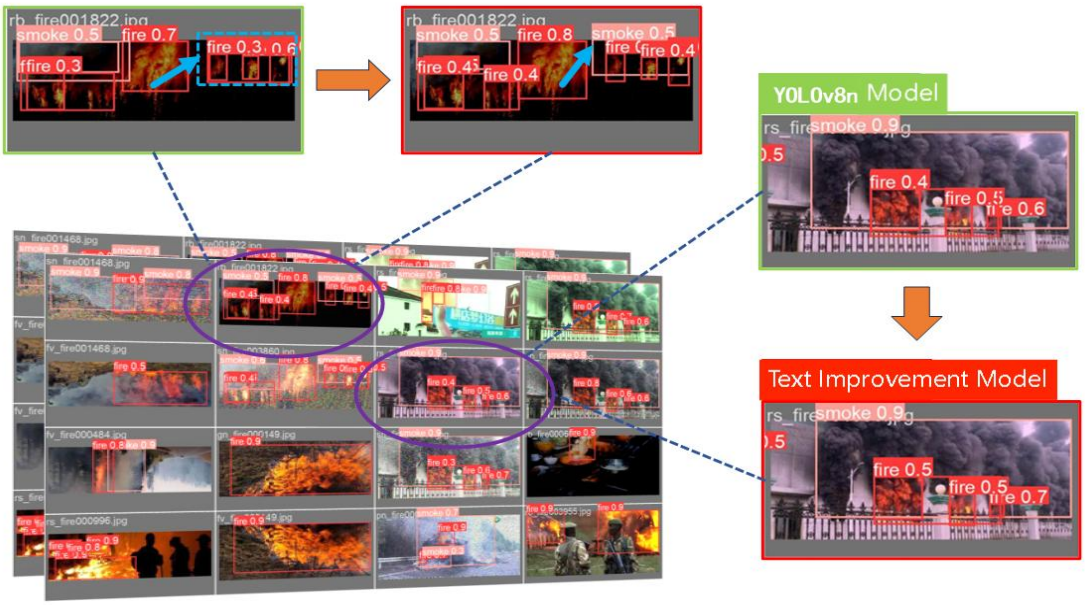

Figure 4 Comparison of Detection Performance between YOLOv8n Model and Text Model

As shown in the figure, YOLOv8n network has the problem of false detection when detecting flames with overlapping and unclear features, and has a high false detection rate for smoke. The improved model can effectively detect the above-mentioned false alarms of flames and smoke, and the confidence level is also higher than that of the YOLOv8n model. The qualitative results further indicate that the detection performance of the improved model is superior to that of the original model, and it has universality on the new dataset.

3.6 Model Comparison Experiment

Compare the performance of the improved model proposed in this article with the following classic object detection models on the same dataset and training configuration. The detailed results are shown in Table 3:

Table 3 Comparison of Fire Point Detection Results of Different Models

Model	mAP	Precision	Recall	FPS
FasterR-CNN	0.76	0.80	0.74	5.2
YOLOv5s	0.77	0.82	0.78	14.5
YOLOv8n	0.78	0.81	0.77	15.8
Model in this article	0.84	0.84	0.80	13.1

From the table, it can be seen that Faster R-CNN performs relatively stably on mAP and Precision, but has low recall and inference speed of only 5.2 FPS, resulting in low inference efficiency. YOLOv5s is slightly better than Faster R-CNN in terms of detection accuracy and inference speed, with better overall performance. YOLOv8n has further improved its inference speed to 15.8 FPS, but the mAP improvement compared to YOLOv5s is only 0.1, while Precision and Recall are slightly lower than YOLOv5s. In contrast, the model proposed in this study achieved the highest values in the three key performance indicators of mAP, Precision, and Recall, and the inference speed also reached 13.1 FPS, which is at a moderate level.

From the results in Table 3, it can be seen that traditional Faster R-CNN is inferior to lightweight detection models in terms of accuracy and inference speed. The YOLO series models exhibit a good speed accuracy balance in fire point detection tasks, especially YOLOv5s and YOLOv8n, which have significant advantages in inference speed. However, the improved model proposed in this study outperforms the comparative model in terms of detection accuracy and maintains a high level of inference speed, verifying that the proposed method has better comprehensive performance and application potential in forest fire point detection tasks.

4 CONCLUSION

This article proposes a novel detection method based on YOLOv8n fusion FSRCNN super-resolution module, eSE attention mechanism, and MPDIoU regression loss to address the challenges of small target recognition, fuzzy features, and high real-time requirements in forest fire point detection tasks. Through sufficient experimental verification on a self built multi scene fire point dataset, the method proposed in this paper outperforms existing mainstream detection frameworks in terms of accuracy, recall, and comprehensive detection performance, and effectively captures fine-grained fire point features while maintaining high inference speed. The ablation experiment showed that the FSRCNN module significantly improved the resolution of small targets, the eSE module enhanced the feature expression ability, and the MPDIoU improved the localization accuracy. The comprehensive comparison results have verified the superiority and applicability of the proposed method in early identification of forest fire points. Future work will further explore lightweight super-resolution methods and self supervised learning mechanisms to enhance detection performance in small sample environments.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This article is funded by the 2024 College Student Innovation and Entrepreneurship Training Program (Project Number: 202412309002)

REFERENCES

[1] Pan YD, Birdsey R, Fang JY, et al. A large and persistent carbon sink in the world's forests. Science, 2011, 333(6045): 988–993.
[2] Li J, Hu F, Du L. Forest Fire Detection Based on Image Processing Technology: A Review. Fire Technology, 2020, 56(5): 1945-1972.
[3] Yuan F, Wan W, Wei H, et al. Review of remote sensing methods for forest fire detection. Remote Sensing, 2018, 10(5): 763.
[4] Koltunov A, Ustin S L. Early Fire Detection Using Non-Linear Multispectral Analysis: International Journal of Wildland Fire, 2007, 16(1): 1-10.
[5] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
[6] Yuan Z, Zhang Y, Liu Y, et al. A Real-Time Forest Fire Detection Algorithm Based on UAV Video: Remote Sensing, 2021, 13(12): 2305.
[7] Dong C, Loy C C, He K, et al. Accelerating the Super-Resolution Convolutional Neural Network: IEEE Transactions on Image Processing, 2016, 25(12): 5534-5545.

[8] Gao Z, He W, Luo H, et al. ESNet: Efficient Squeeze-and-Excitation Network for Channel Attention: IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(6): 2567-2579.
[9] Li Y, Huang T, Liu Y, et al. Bounding box regression method based on improved dynamic IoU. Journal of Software, 2022, 33(11): 4667–4680.
[10] Dong C, Loy C C, He K, et al. Accelerating the Super-Resolution Convolutional Neural Network: IEEE Transactions on Image Processing, 2016, 25(12): 5534-5545.