



OPEN ACCESS

EDITED BY

Sushant K. Singh,
CAIES Foundation, India

REVIEWED BY

Hariharan Shanmugasundaram,
Vardhaman College of Engineering, India
Zhong Wang,
Hefei Normal University, China

*CORRESPONDENCE

Xing Zhang,
✉ 2023740048@qhu.edu.cn

RECEIVED 17 September 2024

ACCEPTED 18 December 2024

PUBLISHED 07 January 2025

CITATION

Li C, Du Y, Zhang X and Wu P (2025) YOLOGX:
an improved forest fire detection algorithm
based on YOLOv8.

Front. Environ. Sci. 12:1486212.

doi: 10.3389/fenvs.2024.1486212

COPYRIGHT

© 2025 Li, Du, Zhang and Wu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

YOLOGX: an improved forest fire detection algorithm based on YOLOv8

Caixiong Li, Yue Du, Xing Zhang* and Peng Wu

School of Computer and Information Science, Qinghai Institute of Technology, Xining, China

To tackle issues, including environmental sensitivity, inadequate fire source recognition, and inefficient feature extraction in existing forest fire detection algorithms, we developed a high-precision algorithm, YOLOGX. YOLOGX integrates three pivotal technologies: First, the GD mechanism fuses and extracts features from multi-scale information, significantly enhancing the detection capability for fire targets of varying sizes. Second, the SE-ResNeXt module is integrated into the detection head, optimizing feature extraction capability, reducing the number of parameters, and improving detection accuracy and efficiency. Finally, the proposed Focal-SIoU loss function replaces the original loss function, effectively reducing directional errors by combining angle, distance, shape, and IoU losses, thus optimizing the model training process. YOLOGX was evaluated on the D-Fire dataset, achieving a mAP@0.5 of 80.92% and a detection speed of 115 FPS, surpassing most existing classical detection algorithms and specialized fire detection models. These enhancements establish YOLOGX as a robust and efficient solution for forest fire detection, providing significant improvements in accuracy and reliability.

KEYWORDS

forest fire detection, YOLOv8, GD mechanism, SE-ResNeXt module, focal-SIoU loss function

1 Introduction

Forest fires, as highly destructive natural disasters, have profound impacts on ecosystems and human activities. These fires devastate extensive forest ecosystems, leading to biodiversity loss and soil degradation, while also disrupting ecological balance (Bowman et al., 2011). Moreover, forest fires damage surrounding buildings, crops, and infrastructure, causing substantial negative impacts on local economies (Robinne and Secretariat, 2021). According to statistics, economic damages from forest fires amount to billions of dollars annually (Thomas et al., 2017). Therefore, researching and developing effective forest fire prevention and control technologies is a critical focus in the fields of environmental science and economics.

Current forest fire detection methods encounter several critical challenges. Firstly, fires often occur in remote areas with vast geographical ranges and scarce human resources, resulting in inefficiencies in manual detection (Allison et al., 2016). Secondly, the rapid propagation of fires means that if they are not detected promptly, they can quickly escalate, causing greater destruction (Martell, 2007). Furthermore, fires usually occur in complex natural environments such as mountains, jungles, and wilderness, where terrain, vegetation, and weather conditions create substantial challenges for detection (Attri et al., 2020). Early

detection is crucial for identifying and reducing fire response times, which can prevent fires from getting out of control (Mcnamee et al., 2023). Thus, developing efficient and reliable fire detection technologies is essential for protecting the ecological environment and mitigating economic damage (Carta et al., 2023).

Traditional forest fire detection methods rely on manually selected features, such as frequency domain, color, shape, and texture. Specific methods include Toreyin et al. employed frequency domain features to detect smoke and flames (Toreyin and Cetin, 2009; Töreyn et al., 2007; Toreyin et al., 2006). Besbes et al. utilized color features to detect smoke (Besbes and Benazza-Benyahia, 2016). Gomes et al. combined color, frequency domain, and temporal features for smoke detection (Gomes et al., 2014). Yuanbin proposed a smoke detection technique integrating temporal, color, texture, and shape features (Yuanbin, 2016). Hossain et al. utilized artificial neural networks (ANN) and local binary patterns (LBP) to detect forest fires and smoke (Hossain et al., 2020). Wang et al. integrated LBP, LBPV, and support vector machines (SVM) for fire and smoke detection (Wang et al., 2016). Although these techniques perform well in real-time efficiency and detection accuracy, they heavily rely on manually selected features, limiting their generalizability and practical applicability.

In recent years, research has shifted from traditional manual feature selection to deep learning-based approaches. These methods are generally classified into two categories: two-stage methods and single-stage methods. Two-stage methods first generate candidate regions, followed by classification and localization of these regions. Prominent examples of two-stage methods include Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2015). Barmpoutis et al. proposed a fire detection method combining Faster R-CNN and multi-dimensional texture analysis, which enhances robustness to false alarms in diverse environmental conditions (Barmpoutis et al., 2019). Zhang et al. developed a wildfire smoke detection system trained on synthetic smoke images (Zhang et al., 2018). Vayadande et al. enhanced Faster R-CNN performance in early wildfire smoke detection through preprocessing and augmenting the training dataset (Vayadande et al., 2018). Cheknane et al. combined deep learning and traditional image processing techniques to propose a two-stage fire detection framework yielding notable improvements in accuracy and speed (Cheknane et al., 2024). Xiao and Wang enhanced Faster R-CNN with skip pooling and context information fusion, performing well in complex environments (Xiao et al., 2020). Chetoui et al. fine-tuned Faster R-CNN and trained it on a large dataset of fire and smoke images, achieving high precision (Chetoui and Akhloufi, 2024). Wang et al. optimized the region proposal network and feature extraction techniques, achieving real-time flame and smoke detection (Wang et al., 2022). Khan et al. proposed the DeepSmoke model, improving outdoor smoke detection accuracy through multi-scale feature extraction (Khan et al., 2021). Ibraheam et al. compared various R-CNN models and found that Faster R-CNN has advantages in speed and accuracy, making it suitable for fire detection applications (Ibraheam et al., 2021). While two-stage methods have high detection accuracy, they are usually slower in real-time performance and rely on significant computational resources and complex network structures, limiting their application in resource-constrained devices.

Single-stage algorithms directly predict the category and location of targets through regression-based object detection networks. Prominent examples of single-stage algorithms include SSD (Liu et al., 2016) and the YOLO series (Redmon et al., 2016; Redmon and Farhadi, 2017; 2018; Bochkovskiy et al., 2020; Li et al., 2022; Wang et al., 2023). Zhao et al. proposed the Fire-YOLO algorithm, achieving real-time detection of small fire targets in complex scenarios (Zhao et al., 2022). Yun et al. developed the FFYOLO model, enhancing classification accuracy and reducing parameters through the CPDA module and MCDH detection head (Yun et al., 2024). Xu et al. integrated ConvNeXtV2 and ConvFormer networks with YOLOv7, significantly improving forest fire detection accuracy (Xu et al., 2024). Talaat et al. proposed the YOLO-SF model, combining instance segmentation and YOLOv7-Tiny to enhance fire detection accuracy (Talaat and ZainEldin, 2023). Cao et al. combined LBP-CNN and YOLOv5 for fire and smoke detection in various environments (Cao et al., 2023). Sun et al. proposed a YOLOv8-based method for fire and smoke detection in IoT monitoring systems (Zhang, 2024). Li et al. introduced an improved YOLOv5-IFFDM model, enhancing fire and smoke detection accuracy by adding attention mechanisms and optimizing loss functions (Li and Lian, 2023). Gonçalves et al. discussed the use of YOLOv7x and YOLOv8s for smoke and wildfire detection (Gonçalves et al., 2024). Choutri et al. researched drone-based fire detection methods using enhanced YOLO algorithms, integrating geolocation capabilities for real-time monitoring (Choutri et al., 2023). Single-stage algorithms excel in real-time performance and efficient computational resource utilization, but they may slightly lack detection accuracy.

While current deep learning technologies demonstrate high detection accuracy in forest fire detection, they encounter challenges such as sensitivity to environmental conditions, inadequate fire source recognition, and inefficient feature extraction. We proposed a forest fire detection algorithm, YOLOGX, based on an enhanced YOLOv8. By integrating the Gather and Distribution (GD) mechanism, modifying the detection head, and optimizing the loss function, the proposed algorithm significantly improves feature extraction capabilities and detection accuracy while reducing the false alarm rate. The primary contributions of this paper are summarized as follows.

- By incorporating the Gather-Distribute (GD) mechanism within the neck section, we establish low-level and high-level GD branches to merge local and global information effectively, significantly enhancing the detection of fire targets of various sizes.
- By integrating the SE-ResNeXt module into the detection head, we employ group convolutions to decrease the number of parameters and utilize the SE module to enhance feature extraction capabilities. This module improves model performance while reducing computational costs.
- By introducing the Focal-SIoU loss function, which integrates angle, distance, shape, and IoU losses, we effectively address the shortcomings of traditional loss functions in managing angle errors, thereby substantially enhancing the model's capability to differentiate between anchor boxes of varying quality.

TABLE 1 The completed dataset.

Dataset	Number of images	Number of targets	Number of smokes	Number of fires
Train	18,686	42,275	18,968	23,307
Test	2,338	5,257	2,342	2,915



FIGURE 1
Forest fire dataset examples.

- The experimental results indicate that the enhanced YOLOv8 model demonstrates a mean Average Precision (mAP) of 80.92% on the D-Fire dataset, operating at a detection speed of 115 frames per second (FPS). The model's complexity stands at 12.3 GFLOPs, with parameters reduced to 6.2 M. This performance outperforms those of most existing classical detection algorithms and specialized fire detection algorithms.

reflection, which enhance its challenge and representativeness for real-world applications.

All labeled images were divided into train and test sets in an 8:2 ratio. To augment the train set, we applied data augmentation techniques such as flipping and cropping, resulting in 18,686 training samples, while the test set remained at 2,338 samples. Detailed information about the dataset is provided in Table 1, and examples from the forest fire dataset are shown in Figure 1.

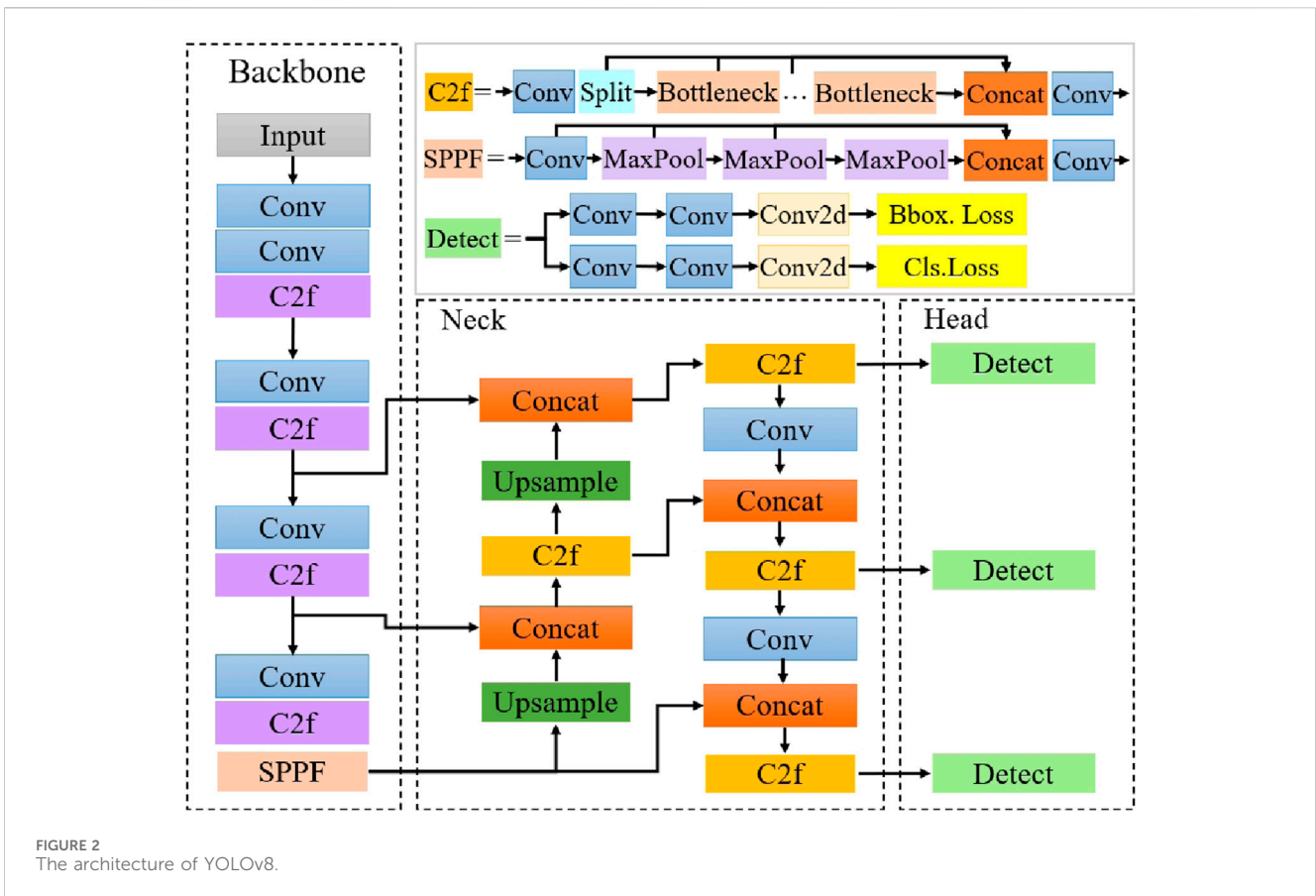
2 Materials and methods

2.1 Dataset

We evaluated our forest fire detection algorithm on the D-Fire dataset (Gaia, 2022). The dataset comprises 21,527 images categorized into four classes: fire (1,164 images), smoke (5,867 images), fire and smoke (4,658 images), and neither fire nor smoke (9,838 images). The dataset presents various shapes, textures, intensities, sizes, and colors of smoke and fire. It encompasses complex environmental factors, including insect occlusion, raindrops, variations in lighting conditions, haze, clouds, and sunlight

2.2 Standard YOLOv8 model

The YOLO model has achieved significant success in the domain of computer vision. YOLOv8 (Ultralytics, 2023), released by Ultralytics in January 2023, significantly enhances detection accuracy and speed compared with its predecessors, YOLOv5 (Ultralytics, 2020) and YOLOv7 (Wang et al., 2023). The YOLOv8 architecture comprises a backbone network, a neck network, and a detection head. The backbone network extracts multi-scale features from RGB images through convolution operations, which provides essential information for subsequent processing. The neck network fuses features extracted by the



backbone network, typically utilizing a Feature Pyramid Network (FPN) to aggregate low-level features into higher-level representations, thereby enhancing the model’s ability to detect objects at different scales. Finally, the detection head predicts object classes, employing three sets of detection heads to select and detect images of different sizes, thereby improving detection precision. Figure 2 illustrates the architecture of the standard YOLOv8 network.

2.3 Improved YOLOv8 model

YOLOGX, an enhanced YOLOv8 model, is utilized for forest fire detection. First, the Gather and Distribution (GD) mechanism is integrated into the neck section, enhancing the detection capability for multi-scale fire targets through low-level and high-level GD branches, along with feature alignment, information fusion, and information injection modules. Second, the detection head incorporates the SE-ResNeXt module (Xie et al., 2017; Hu et al., 2018), which reduces the number of parameters and enhances feature extraction efficiency through group convolution and SE modules. Finally, the Focal-SIoU loss function (Gevorgyan, 2022; Lin et al., 2017) is introduced, addressing angle, distance, shape, and IoU losses, thereby improving the model’s ability to differentiate between anchor boxes of varying quality. These improvements are illustrated within the red dashed boxes in Figure 3.

2.3.1 Gather-Distribute mechanism

The size of fire targets in forest fire detection varies due to differences in distance and range of fire areas. To enhance detection capabilities for fire targets of varying sizes, we introduced the Gather-Distribute (GD) mechanism (Wang et al., 2024), which replaces the neck of YOLOv8. The GD mechanism comprises the Feature Alignment Module (FAM), Information Fusion Module (IFM), and Information Injection Module (IIM), as shown in Figure 4. This mechanism facilitates efficient interaction and fusion of information from different levels to acquire global information, which is then injected into various feature levels. Without significantly increasing latency, the GD mechanism enhances the information fusion capability of the neck part and improves the model’s ability to detect targets across varying sizes.

The GD mechanism consists of two branches: Low-GD and High-GD. The Low-GD branch processes larger-sized fire images by adjusting feature map sizes and adopting fusion strategies to enhance local details of the features, as shown in Figure 5.

In the Low-GD branch, feature maps B2, B3, B4, and B5 first enter the low FAM, where they are adjusted to the same spatial resolution through average pooling operations and then connected in the channel dimension. The formula is expressed in Equation 1:

$$F_{Low_FAM_align} = align_{FAM}([B2, B3, B4, B5]) \tag{1}$$

The output of FAM is processed by multi-layer RepBlock information extraction and then divided into two parts by

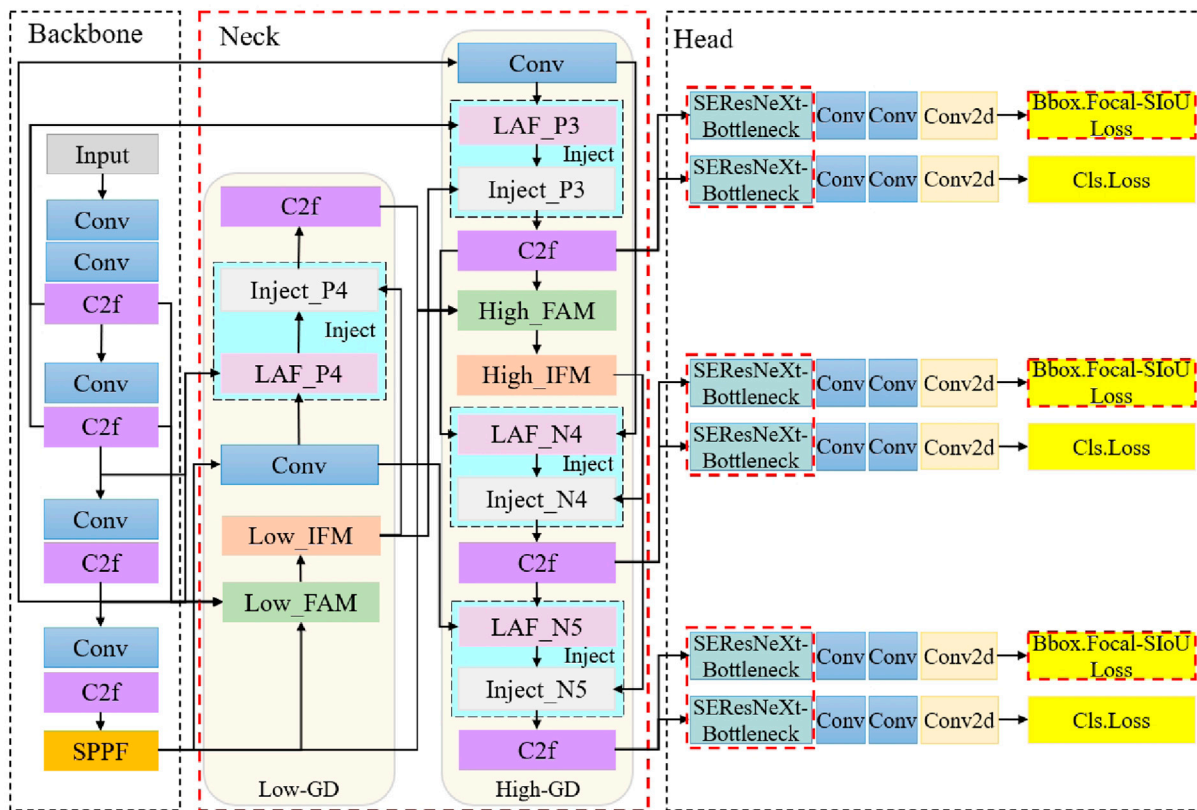


FIGURE 3 The architecture of the forest fire detection network based on YOLOGX.

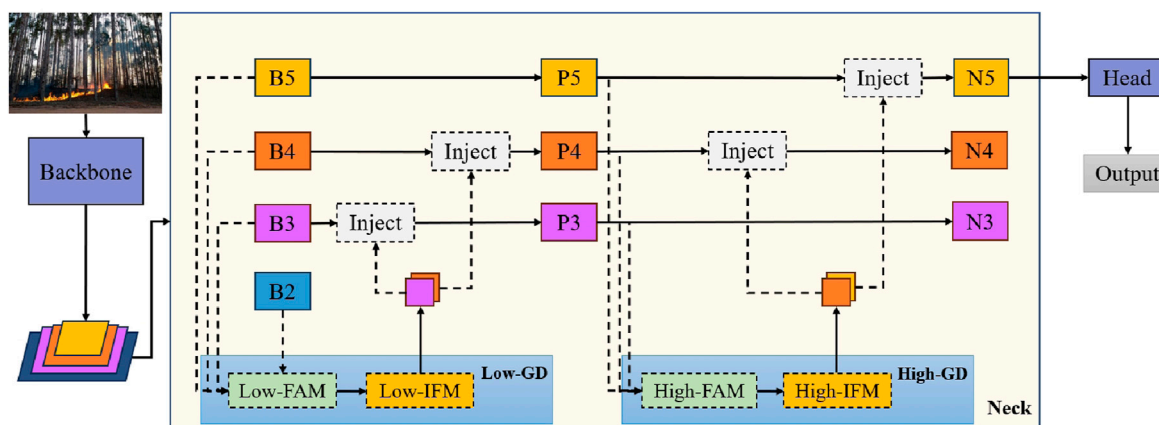


FIGURE 4 The GD mechanism.

channel, resulting in *Inject_P3* and *Inject_P4*. The formula is expressed in Equation 2:

$$[Inject_P3, Inject_P4] = Split(RepBlock(F_{align})) \quad (2)$$

High-GD targets smaller-sized fire images by employing multi-head attention mechanisms and feed-forward networks for feature

fusion to improve the model’s accuracy in recognizing small-scale fires, as shown in Figure 6. In the High-GD architecture, a Transformer extracts and fuses features, utilizing its global modeling capability. The feature maps P3, P4, and P5, output by the information injection module, further undergo semantic information fusion. High FAM concatenates feature maps of

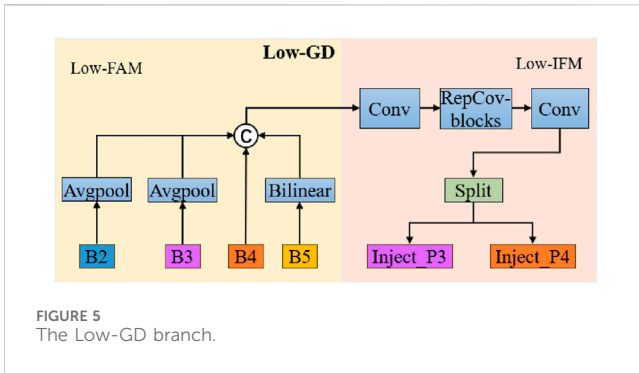


FIGURE 5 The Low-GD branch.

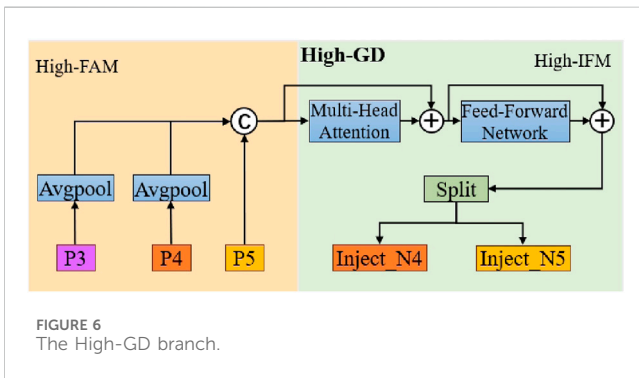


FIGURE 6 The High-GD branch.

different scales, achieving multi-scale information alignment. The formula is expressed in Equation 3:

$$F_{High_FAM_align} = align_{FAM}([P3, P4, P5]) \quad (3)$$

High IFM employs the Transformer module for global modeling and fusion of global feature information. First, the Multi-Head Self-Attention (MHSA) module calculates global attention; then, the feed-forward network (FFN) is introduced to acquire advanced features, ultimately producing *Inject_N4* and *Inject_N5*. The formula is expressed in Equation 4:

$$[Inject_N4, Inject_N5] = Split(FFN(MHSA(F_{align}))) \quad (4)$$

The information injection module injects global information into various feature levels to enhance the model’s ability to detect fire targets. This module applies attention mechanisms to weight global information, highlighting key features and suppressing unimportant

information, then injects the weighted global information into local features. Through addition or concatenation operations, the global information is effectively merged with local features, enhancing the model’s comprehension and recognition of fire targets, as shown in Figure 7.

2.3.2 SE-ResNeXt detection head

In forest fire detection, network structure design is critical for both model performance and computational efficiency. Traditional convolutional layers are parameter-heavy and computationally intensive, limiting their applicability in resource-constrained environments. Introducing the SE-ResNeXt module into the YOLOv8 detection head can enhance feature extraction efficiency and fire detection performance while maintaining low computational costs and parameters, thus achieving efficient and stable forest fire detection.

The ResNeXt network (Xie et al., 2017), a variant of ResNet, reduces model parameters and improves training speed through the use of group convolution, as illustrated in Figure 8. Let k denote the size of the convolution kernel, C the number of input feature matrix channels, and n the number of output feature matrix channels. The number of parameters in a traditional convolution is $k \times k \times C \times n$, whereas for group convolution it is $k \times k \times C \times n \times (1/g)$, where g denotes the number of groups, typically set to 32. Group convolution significantly reduces the number of model parameters.

The SE-ResNeXt network extends the ResNeXt network by incorporating the SE module, as shown in Figure 9. The SE module (Hu et al., 2018) reduces the feature matrix dimensions from $H \times W$ to 1 through global average pooling (GAP). It then extracts inter-channel dependencies using a bottleneck structure composed of two fully connected layers, which output the same number of weights as the input features. The first fully connected layer reduces the feature dimension to 1/16 of the input, followed by a ReLU activation function, after which another fully connected layer restores the dimension. The Sigmoid function produces normalized weights, which are then applied to each channel’s features through scaling. The SE module is connected after the ResNeXt bottleneck, thereby forming the SE-ResNeXt module.

Integrating the SE-ResNeXt module into the YOLOv8 detection head leverages its capability to enhance the modeling of inter-channel relationships, thereby improving fire detection performance. Figure 10 illustrates the architecture of the SE-ResNeXt detection head. Initially, the input feature map X undergoes 1×1 convolution to adjust the channel dimensions, followed by batch normalization and ReLU activation.

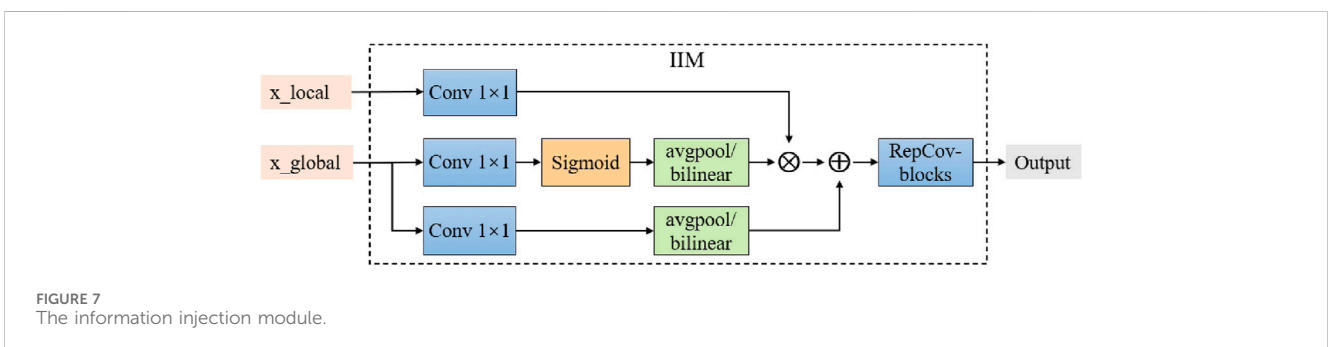


FIGURE 7 The information injection module.

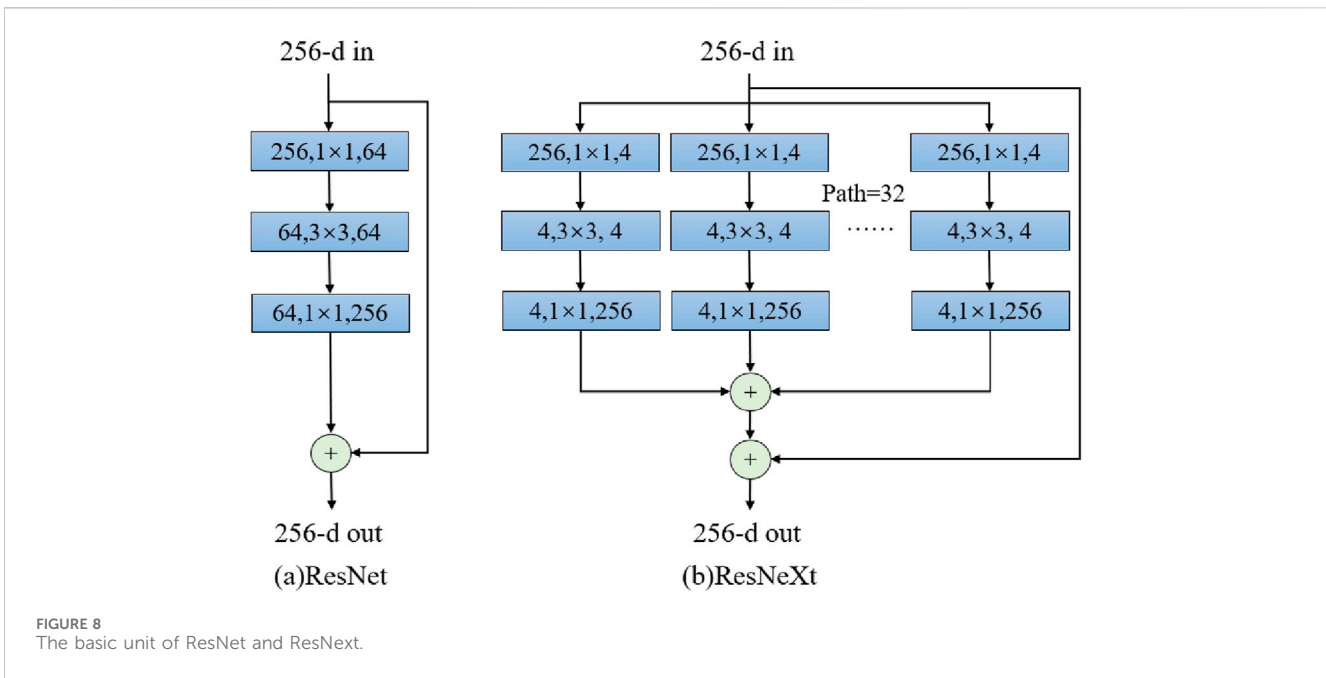


FIGURE 8 The basic unit of ResNet and ResNext.

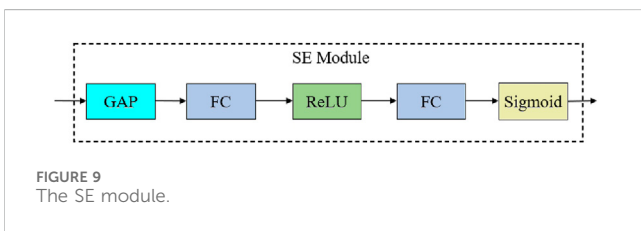


FIGURE 9 The SE module.

Subsequently, the feature map undergoes 3×3 group convolution (with 32 groups) and 1×1 convolution for feature extraction, followed by feature fusion. The group convolution expands the network while reducing the number of parameters. Here, BatchNorm prevents ReLU activation saturation, accelerates network convergence, and mitigates overfitting. The SE module is integrated after the ResNeXt bottleneck, utilizing global average pooling (GAP) for the Squeeze operation. Two fully connected layers then extract inter-channel dependencies. Finally, the Excitation operation, i.e., the Sigmoid function, is applied to obtain normalized weights, which are then used for each channel feature through scaling. The processed features are then combined with the network input X to serve as input for the next module. This architecture, by integrating residual connections with the SE module, significantly enhances feature extraction capabilities and improves model training efficiency.

In summary, the SE-ResNeXt-enhanced YOLOv8 detection head not only effectively extracts multi-dimensional features and improves detection accuracy but also maintains low computational costs and parameters, thereby achieving efficient and stable forest fire detection.

2.3.3 Focal-SIoU loss

In forest fire detection, loss function design is critical, as it significantly influences detection efficiency and model performance.

Traditional loss functions typically emphasize core metrics in bounding box regression, such as distance between bounding boxes, overlap area, and aspect ratio. However, these loss functions often overlook angular inconsistency between the predicted and ground truth boxes. This omission of directional error can cause the predicted box to fail to accurately enclose the ground truth box during training, thereby reducing training speed and convergence, ultimately affecting overall detection performance. To address this challenge, we introduced the Focal-SIoU loss function, specifically optimized for forest fire detection in the improved YOLOv8 model. This function corrects angular error while optimizing the training process.

The Focal-SIoU loss function comprises four components: angular loss, distance loss, shape loss, and IoU loss. The calculation of angular cost is represented by Equations 5–8. The left subfigure of Figure 11 illustrates that Λ depends on α , where α denotes the relative angle between two boxes.

$$\Lambda = 1 - 2\sin^2\left(\arcsin(x) - \frac{\pi}{4}\right) \tag{5}$$

$$x = \frac{c_h}{\sigma} = \sin\alpha \tag{6}$$

$$\sigma = \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2} \tag{7}$$

$$c_h = \max(b_{c_y}^{gt}, b_{c_y}) - \min(b_{c_y}^{gt}, b_{c_y}) \tag{8}$$

The calculation process of distance cost is described by Equations 9, 10. The right subfigure of Figure 11 shows that the distance cost Δ depends on ρ_x and ρ_y , which quantify the distance differences between the predicted and ground truth boxes. The weight of the distance cost is controlled by Δ , using γ_1 to balance the losses of ρ_x and ρ_y .

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma_1 \rho_t}) \tag{9}$$

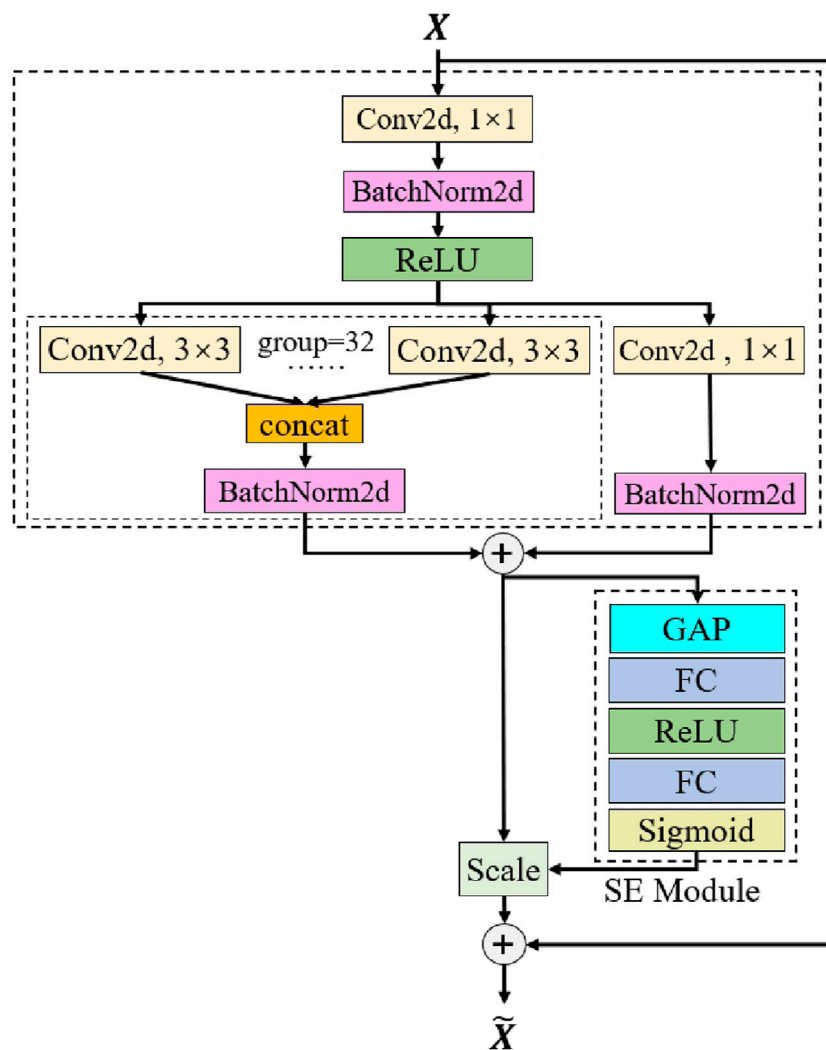


FIGURE 10 The SE-ResNeXt detection head.

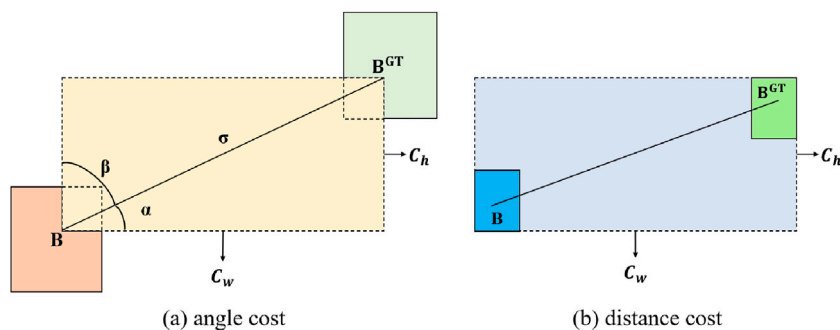


FIGURE 11 The angle cost and distance cost.

TABLE 2 The experimental environment settings.

Experimental environment	Details
Operating System	Ubuntu 20.04
Deep Learning Framework	PyTorch 2.0.0
Programming Language	Python 3.8
CUDA Version	CUDA 11.8
CPU	Intel(R) Xeon(R) Platinum 8352 V CPU @ 2.10 GHz
GPU	RTX 4090 (24 GB)
Host memory	120 GB

$$\rho_x = \frac{b_{c_x}^{gt} - b_{c_x}}{c_w}, \rho_y = \frac{b_{c_y}^{gt} - b_{c_y}}{c_h}, \gamma_1 = 2 - \Lambda \quad (10)$$

The shape cost calculation process is represented by Equations 11, 12. ω_w and ω_h denote the relative changes in width and height. The parameter θ is a hyperparameter used to control the sensitivity of the shape loss to relative changes in the bounding box width and height. We set $\theta = 4$.

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta \quad (11)$$

$$\omega_w = \frac{|\omega - \omega^{gt}|}{\max(\omega, \omega^{gt})}, \omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (12)$$

The calculation process of IoU cost is described by Equation 13.

$$L_{IoU_{Cost}} = 1 - IoU \quad (13)$$

Finally, by combining these losses, the expression of the SIoU loss (Gevorgyan, 2022) is given by Equation 14:

$$L_{SIoU} = 1 - IoU + (\Lambda + \Omega)/2 \quad (14)$$

To address the issue of imbalanced training samples, this study introduces the Focal loss (Lin et al., 2017) is described by Equation 15. By combining SIoU loss with Focal loss, we developed the Focal-SIoU loss function, which differentiates between high-quality and low-quality anchor boxes, thereby improving model stability and detection performance. The parameter γ_2 is the focusing parameter in the Focal-SIoU loss function, designed to balance the contribution of easy and hard samples to the loss. We set $\gamma_2 = 2$.

$$L_{Focal-SIoU} = IoU^{\gamma_2} L_{SIoU} \quad (15)$$

In this way, the Focal-SIoU loss function not only addresses directional mismatch but also balances the impact of high-quality and low-quality anchor boxes, thereby enhancing the model's overall adaptability and effectiveness in fire recognition tasks.

2.4 Experimental setup

2.4.1 Experimental environment and parameter settings

The experiments were conducted using PyTorch 2.0.0 and Python 3.8, with all models trained on Ubuntu 20.04. A detailed description of the experimental environment is provided in Table 2.

During training, stochastic gradient descent (SGD) optimization was employed during training, with an initial learning rate of 0.01, a patience value of 5,000, and a momentum factor of 0.937. The input image size was set to 640×640 , with a batch size of 16, and training spanned 300 epochs.

2.4.2 Evaluation metrics

To evaluate the model's performance, we utilized various metrics, including precision, recall, mean Average Precision at 50% (mAP50), mean Average Precision from 50% to 95% (mAP50-95), frames per second (FPS), the number of parameters (Params), and giga floating-point operations per second (GFLOPs). The calculation methods for these metrics are provided in Equations 16–19.

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$AP = \int_0^1 Precision(r) dr \quad (18)$$

Here, r represents recall, and $Precision(r)$ denotes the precision corresponding to the recall value r .

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (19)$$

3 Results

3.1 Comparative experiment of loss function improvement

Table 3 presents the experimental results for different loss functions. The parameters (6.2 M) and computational complexity (12.3 GFLOPs) remained constant across all methods, indicating that the performance improvements were primarily due to enhancements in the loss functions rather than changes in model complexity. The Focal-SIoU achieved the highest mAP@0.5 of 80.92%, while the MPDIoU excelled in mAP@0.5–0.95. If the primary focus is on detection performance at high IoU thresholds, MPDIoU would be the better choice. However, for overall detection accuracy, Focal-SIoU is more suitable. For forest

TABLE 3 The experimental results of loss function improvement.

Method	mAP@0.5	mAP@0.5–0.95	Params	GFLOPs
CIoU	80.03	46.42	6.2 M	12.3
SIoU (Gevorgyan, 2022)	79.81	47.04	6.2 M	12.3
EIoU (Zhang et al., 2022)	80.04	46.93	6.2 M	12.3
WIoU (Tong et al., 2023)	80.22	46.86	6.2 M	12.3
MPDIoU (Siliang and Yong, 2023)	80.17	47.13	6.2 M	12.3
Focal-EIoU (Zhang et al., 2022)	80.71	46.91	6.2 M	12.3
Focal-SIoU	80.92	46.44	6.2 M	12.3

The highest scores are presented in bold.

TABLE 4 Ablation experiment results.

YOLOv8	GD	SE-ResNeXt detection head	Focal-SIoU	Precision	Recall	mAP@0.5	mAP@0.5–0.95	FPS	GFLOPs	Params
✓				78.80	72.50	79.46	46.61	197	8.2	3.0 M
✓	✓			80.30	73.00	79.82	46.92	139	12.0	6.0 M
✓		✓		78.20	72.70	79.55	46.59	127	8.6	3.2 M
✓			✓	79.10	71.90	79.36	46.37	123	8.2	3.0 M
✓	✓	✓		80.20	73.00	80.03	46.42	115	12.3	6.2 M
✓	✓	✓	✓	77.20	75.70	80.92	46.44	115	12.3	6.2 M

The highest scores are presented in bold.

fire detection, where overall detection accuracy is paramount, we ultimately selected Focal-SIoU.

3.2 Ablation experiment

Table 4 presents the ablation experiment results of various components, including the GD mechanism, SE-ResNeXt detection head, and Focal-SIoU, when integrated with the YOLOv8 model. The baseline YOLOv8 model demonstrates balanced performance with a precision of 78.80%, a recall of 72.50%, and a high frame rate of 197 FPS. After incorporating the GD module, precision increased to 80.30% and recall improved to 73.00%; however, the parameters increased clearly. Introducing the SE-ResNeXt detection head slightly decreased precision to 78.20% and slightly increased recall to 72.70%. After adding Focal-SIoU, precision and recall were adjusted to 79.10% and 71.90%, respectively. Combining the GD mechanism and SE-ResNeXt detection head obviously improved precision to 80.20% and recall to 73.00%; however, parameters increased to 6.2 M, and frames per second reduced to 115. Finally, combining all components resulted in the best performance in terms of recall (75.70%) and mAP@0.5 (80.92%), although precision slightly decreased to 77.20%, with the frame rate remaining at 115 FPS.

The mAP@0.5 improved from 79.46% to 80.92%, representing a 1.46% increase. Precision is critical in forest fire detection, and even a 1% improvement in mAP@0.5 is significant. This enhancement means that the model can detect fires more rapidly

and accurately, enabling timely interventions that can mitigate fire-related damages.

Although the improvements in mAP@0.5 and recall led to a slight decrease in mAP@0.5–0.95 (from 46.61% to 46.44%) and a reduction in FPS (from 197 FPS to 115 FPS), we deem this tradeoff to be reasonable. First, in the accuracy-driven application of forest fire detection, accurately identifying fire threats takes precedence over achieving higher FPS, and 115 FPS is adequate for real-time detection. Second, the decrease in mAP@0.5–0.95 is merely 0.17%, which has minimal impact on practical detection. Additionally, while the GFLOPs increased from 8.2 to 12.3 and parameters grew from 3.0 M to 6.2 M, these increases were implemented to enhance the model's ability to detect small targets and handle complex scenarios. These changes remain within acceptable limits, and contemporary hardware, such as GPUs and embedded systems, can efficiently handle this level of computational complexity. In brief, the improvement in mAP@0.5 is pivotal for the early detection and prevention of forest fires. While the model's complexity has increased, the improvement in detection accuracy distinctly outweighs the computational costs.

Figure 12 illustrates the training curves of YOLOv8 and improved YOLOv8. The results indicate that the improved YOLOv8 surpasses the baseline model in both precision and recall, with more pronounced improvements observed during the early training stages. Furthermore, the enhanced model shows improvements in mAP@0.5 and mAP@0.5–0.95, suggesting that the optimization modules contribute to enhancements in both the model's feature extraction capabilities and convergence.

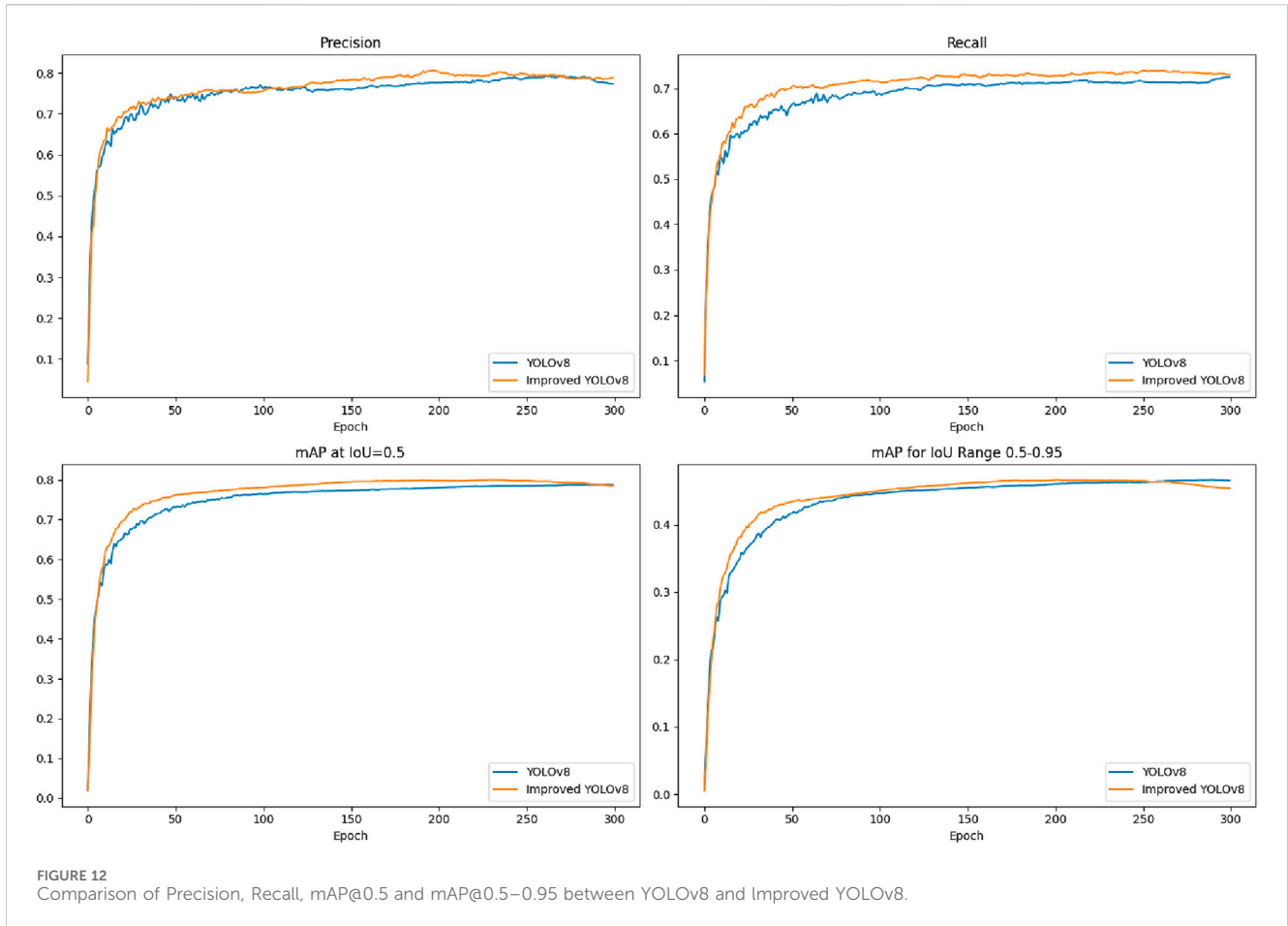


TABLE 5 The comparison of performance with other DL methods.

Method	mAP@0.5	Params	GFLOPs	FPS
Faster-RCNN (Ren et al., 2015)	64.78	28.5 M	471.58	47
EfficientDet (Tan et al., 2020)	71.00	3.79 M	4.7	90
YOLOv4-tiny (Bochkovskiy et al., 2020)	62.00	5.9 M	16	197
YOLOv4 (Bochkovskiy et al., 2020)	73.00	65 M	142	40
YOLOv5 (Ultralytics, 2020)	67.10	7.2 M	16.2	120
YOLOv7 (Meituan, 2022)	80.20	37.2 M	105.1	159
YOLOv8 (Ultralytics, 2023)	79.46	3.0 M	8.2	197
RT-DETR-L (Zhao et al., 2024)	72.18	32.8 M	108.0	108
RT-DETRv2-S (Lv et al., 2024)	80.60	20.1 M	\	207
YOLOGX (ours)	80.92	6.2 M	12.3	115

The highest scores are presented in bold.

3.3 Comparative experiment with other methods

Table 5 provides a comparison of YOLOGX with other deep learning methods on the D-Fire dataset. YOLOGX achieves the best

performance in terms of mAP@0.5, reaching 80.92%, while maintaining high real-time performance (115 FPS), thus striking a favorable balance between accuracy and efficiency. Although its parameters and computational complexity (6.2 M parameters, 12.3 GFLOPs) are slightly higher than some methods, the

TABLE 6 The comparison of performance with state-of-the-art studies on the D-Fire dataset.

Method	Year	Accuracy	Precision	Recall	mAP@0.5	Params
Pruned-KD (Wang et al., 2021)	2021	82.40	79.85	46.02	62.71	\
PicoDet (Yu et al., 2021)	2021	88.26	84.32	51.27	68.95	\
Fire-YOLO (Zhao et al., 2022)	2022	88.21	84.12	51.58	68.88	\
LFNet (Xiaoxue et al., 2023)	2023	97.92	87.68	53.35	71.15	\
ELDNet (Guo et al., 2024)	2024	\	70.60	\	74.00	7.77 M
YOLOGX (ours)	2024	\	77.20	75.70	80.92	6.2 M

The highest scores are presented in bold.



FIGURE 13 Comparison of forest fire detection results before (up) and after (down) improvement.

decrease in frame rate is minimal compared to its accuracy, making it well-suited for real-time detection.

In contrast, Faster-RCNN, YOLOv4, YOLOv7, RT-DETR-L, and RT-DETRv2-S have more parameters than YOLOGX. EfficientDet has lower mAP@0.5 and FPS. YOLOv4-tiny and YOLOv5 exhibit high FPS, but lower mAP@0.5. YOLOGX outperforms other methods in both accuracy and real-time capability, making it particularly suitable for forest fire detection, which requires relatively high accuracy and high real-time performance.

To comprehensively evaluate our method, we compared YOLOGX with the latest research methods including Pruned-KD, PicoDet, Fire-YOLO, LFNet, and ELDNet on the D-Fire dataset, as shown in Table 6. Under the same experimental settings, YOLOGX achieved a mAP@0.5 of 80.92%, significantly higher than those of other methods (Pruned-KD 62.71%, PicoDet 68.95%, Fire-YOLO 68.88%, LFNet 71.15%, ELDNet 74.00%). Additionally, YOLOGX demonstrated superior precision and recall, with values of 77.20% and 75.70%, respectively. YOLOGX also excelled in detection speed, with 6.2 M parameters. Overall, YOLOGX strikes a good balance between accuracy and efficiency, making it particularly suitable for forest fire detection applications requiring high precision.

3.4 Visualization analysis

Figure 13 compares the performance of the YOLOv8 and YOLOGX algorithms in fire and smoke detection. The visual results indicate that the YOLOGX algorithm demonstrates significant advantages in detecting fire and smoke. Firstly, YOLOGX successfully detects more fire targets that YOLOv8 fails to recognize, demonstrating its higher sensitivity to fire targets in complex and concealed environments. Secondly, YOLOGX excels in detecting small fire targets, outperforming YOLOv8 in capturing and marking small fire points. Additionally, in the same flame detection tasks, YOLOGX shows higher detection confidence, indicating greater reliability and accuracy. These improvements are attributed to the integration of the GD mechanism, SE-ResNeXt detection head, and Focal-SIoU loss function in YOLOGX, which enhance its effectiveness in fire detection.

4 Discussion

We introduced YOLOGX, an enhanced YOLOv8 model designed for real-time forest fire detection. The model incorporated the Gather and Distribute (GD) feature aggregation

module, SE-ResNeXt detection head, and Focal-SIoU loss function, significantly enhancing detection capabilities. Experimental results demonstrate that the improved model surpasses the original in recall and mean Average Precision (mAP), particularly excelling in detecting small targets. These results underscore the effectiveness and robustness of the enhanced model in forest fire detection, highlighting its potential for practical applications.

Despite the superior performance of YOLOGX, some limitations remain. First, acquiring high-quality fire image datasets is challenging due to the complex and time-consuming annotation processes, which limits training data sources and affects detection performance. Second, diverse forest backgrounds and varying weather conditions can easily result in false positives and missed detections. The complex characteristics of smoke and fire, particularly in low-light conditions, exacerbate detection difficulty. Additionally, high real-time requirements and limited computational resources necessitate model optimization to balance performance and efficiency.

Future research should focus on several aspects to further enhance model performance. Data augmentation techniques, particularly Generative Adversarial Networks (GANs) and other large image generation models (Feng et al., 2024; Liang et al., 2024; Xin et al., 2024), hold substantial promise. These technologies enable the synthesis of diverse training data, thereby enhancing model generalization, especially in scenarios where data acquisition is constrained. To address environmental complexity, constructing forest background models and integrating multi-scale feature extraction with meteorological data fusion techniques can enhance fire and smoke detection accuracy. To address the complexity of fire characteristics, employing temporal analysis methods (e.g., LSTM) and multimodal data (e.g., infrared images, thermal imaging) can further improve detection performance. To optimize computational efficiency, model compression techniques and lightweight models (e.g., MobileNet, EfficientNet) can reduce computational load and improve inference speed.

5 Conclusion

This study introduced YOLOGX, an enhanced YOLOv8 forest fire detection algorithm designed to improve feature extraction and detection accuracy in complex natural environments. The model incorporated the Gather and Distribution (GD) mechanism, including low-GD and high-GD branches, which integrate feature alignment, information fusion, and information injection modules to enhance the fusion capabilities of multi-scale features. The detection head integrated the SE-ResNeXt module, which reduces parameters through group convolution and SE modules, enhancing the modeling of feature channel relationships and reducing computational cost. Additionally, we proposed the Focal-SIoU loss function, which combines angle, distance, shape, and IoU losses, and employs Focal loss to address training sample imbalance, thereby improving the model's anchor box discrimination ability. Experimental results demonstrate that the YOLOGX model achieves a mAP of 80.92% on the D-Fire dataset,

with a detection speed of 115 FPS, a computational complexity of 12.3 GFLOPs, and parameters reduced to 6.2 M, outperforming most existing detection algorithms.

Future research will concentrate on several key areas: developing and applying data augmentation techniques to increase data diversity; integrating multimodal detection methods, including infrared images and thermal imaging, to improve detection performance; and exploring model compression techniques to reduce computational demands and enhance inference speed. We anticipate that with ongoing optimization and experimental validation, YOLOGX will assume a broader role in fire prevention and disaster management, providing robust technical support for societal safety and protection.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material

Author contributions

CL: Methodology, Software, Writing—original draft, Writing—review and editing. YD: Validation, Visualization, Writing—review and editing. XZ: Conceptualization, Funding acquisition, Resources, Writing—original draft, Writing—review and editing. PW: Data curation, Project administration, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Natural Science Foundation of Qinghai Province, China (No. 2023-QLGKLYCZX-017).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Allison, R. S., Johnston, J. M., Craig, G., and Jennings, S. (2016). Airborne optical and thermal remote sensing for wildfire detection and monitoring. *Sensors* 16, 1310. doi:10.3390/s16081310
- Attri, V., Dhiman, R., and Sarvade, S. (2020). A review on status, implications and recent trends of forest fire management. *Archives Agric. Environ. Sci.* 5, 592–602. doi:10.26832/24566632.2020.0504024
- Barmoutis, P., Dimitropoulos, K., Kaza, K., and Grammalidis, N. (2019). “Fire detection from images using faster r-cnn and multidimensional texture analysis,” in ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP) (IEEE), 8301–8305.
- Besbes, O., and Benazza-Benyahia, A. (2016). “A novel video-based smoke detection method based on color invariants,” in 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP) (IEEE), 1911–1915.
- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.
- Bowman, D. M., Balch, J., Artaxo, P., Bond, W. J., Cochrane, M. A., D’antonio, C. M., et al. (2011). The human dimension of fire regimes on earth. *J. Biogeogr.* 38, 2223–2236. doi:10.1111/j.1365-2699.2011.02595.x
- Cao, X., Su, Y., Geng, X., and Wang, Y. (2023). Yolo-sf: yolo for fire segmentation detection. *IEEE Access* 11, 111079–111092. doi:10.1109/access.2023.3322143
- Carta, F., Zidda, C., Putzu, M., Loru, D., Anedda, M., and Giusto, D. (2023). Advancements in forest fire prevention: a comprehensive survey. *Sensors* 23, 6635. doi:10.3390/s23146635
- Cheknane, M., Bendouma, T., and Boudouh, S. S. (2024). Advancing fire detection: two-stage deep learning with hybrid feature extraction using faster r-cnn approach. *Signal. Image Video Process.* 18, 5503–5510. doi:10.1007/s11760-024-03250-w
- Chetoui, M., and Akhloufi, M. A. (2024). Fire and smoke detection using fine-tuned yolov8 and yolov7 deep models. *Fire* 7, 135. doi:10.3390/fire7040135
- Choutri, K., Lagha, M., Meshoul, S., Batouche, M., Bouzidi, F., and Charef, W. (2023). Fire detection and geo-localization using uav’s aerial images and yolo-based models. *Appl. Sci.* 13, 11548. doi:10.3390/app132011548
- Gaia (2022). D-fire: an image dataset for fire and smoke detection. Available at: <https://github.com/gaias/DFireDataset/> (Accessed February 1, 2024).
- Feng, S., Huang, Y., and Zhang, N. (2024). An improved yolov8 obb model for ship detection through stable diffusion data augmentation. *Sensors* 24, 5850. doi:10.3390/s24175850
- Gevorgyan, Z. (2022). Siou loss: more powerful learning for bounding box regression. *arXiv Prepr. arXiv:2205*. doi:10.48550/arXiv.2205.12740
- Girshick, R. (2015). “Fast r-cnn,” in Proceedings of the IEEE international conference on computer vision, 1440–1448.
- Gomes, P., Santana, P., and Barata, J. (2014). A vision-based approach to fire detection. *Int. J. Adv. Robotic Syst.* 11, 149. doi:10.5772/58821
- Gonçalves, L. A. O., Ghali, R., and Akhloufi, M. A. (2024). Yolo-based models for smoke and wildfire detection in ground and aerial images. *Fire* 7, 140. doi:10.3390/fire7040140
- Guo, X., Cao, Y., and Hu, T. (2024). An efficient and lightweight detection model for forest smoke recognition. *Forests* 15, 210. doi:10.3390/fl15010210
- Hossain, F. A., Zhang, Y. M., and Tonima, M. A. (2020). Forest fire flame and smoke detection from uav-captured images using fire-specific color features and multi-color space local binary pattern. *J. Unmanned Veh. Syst.* 8, 285–309. doi:10.1139/jvus-2020-0009
- Hu, J., Shen, L., and Sun, G. (2018). “Squeeze-and-excitation networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 7132–7141.
- Ibraheem, M., Li, K. F., Gebali, F., and Sielecki, L. E. (2021). A performance comparison and enhancement of animal species detection in images with various r-cnn models. *AI* 2, 552–577. doi:10.3390/ai2040034
- Khan, S., Muhammad, K., Hussain, T., Del Ser, J., Cuzzolin, F., Bhattacharyya, S., et al. (2021). DeepSmoke: deep learning model for smoke detection and segmentation in outdoor environments. *Expert Syst. Appl.* 182, 115125. doi:10.1016/j.eswa.2021.115125
- Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., et al. (2022). Yolov6: a single-stage object detection framework for industrial applications. arXiv Prepr. arXiv:2209.02976. doi:10.48550/arXiv.2209.02976
- Li, J., and Lian, X. (2023). Research on forest fire detection algorithm based on improved yolov5. *Mach. Learn. Knowl. Extr.* 5, 725–745. doi:10.3390/make5030039
- Liang, Y., Feng, S., Zhang, Y., Xue, F., Shen, F., and Guo, J. (2024). A stable diffusion enhanced yolov5 model for metal stamped part defect detection based on improved network structure. *J. Manuf. Process.* 111, 21–31. doi:10.1016/j.jmapro.2023.12.064
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). “Focal loss for dense object detection,” in Proceedings of the IEEE international conference on computer vision, 2980–2988.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). “Ssd: single shot multibox detector,” in Computer vision—ECCV 2016: 14th European conference, Amsterdam, The Netherlands, October 11–14, 2016, proceedings, Part I 14 (Springer), 21–37.
- Lv, W., Zhao, Y., Chang, Q., Huang, K., Wang, G., and Liu, Y. (2024). Rt-detr2: improved baseline with bag-of-freebies for real-time detection transformer. arXiv preprint arXiv:2407.17140.
- Martell, D. L. (2007). “Forest fire management,” in *Handbook of operations research in natural resources* (Springer), 489–509.
- Mcnamee, M., Mcnamee, R., Meacham, B., and Amon, F. (2023). *Environmental benefits of rapid fire detection*.
- Meituan (2022). Available at: <https://github.com/WongKinYiu/yolov7> (Accessed May 1, 2024).
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). “You only look once: unified, real-time object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 779–788.
- Redmon, J., and Farhadi, A. (2017). “Yolo9000: better, faster, stronger,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 7263–7271.
- Redmon, J., and Farhadi, A. (2018). Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks. *Adv. neural Inf. Process. Syst.* 28. doi:10.1109/TPAMI.2016.2577031
- Robinne, F.-N., and Secretariat, F. (2021). “Impacts of disasters on forests,” in *Particular forest fires. UNFFS Background paper*.
- Siliang, M., and Yong, X. (2023). Mpdjou: a loss for efficient and accurate bounding box regression. arXiv Prepr. arXiv:2307.07662. doi:10.48550/arXiv.2307.07662
- Talaat, F. M., and ZainEldin, H. (2023). An improved fire detection approach based on yolo-v8 for smart cities. *Neural Comput. Appl.* 35, 20939–20954. doi:10.1007/s00521-023-08809-1
- Tan, M., Pang, R., and Le, Q. V. (2020). “Efficientdet: scalable and efficient object detection,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 10781–10790.
- Thomas, D., Butry, D., Gilbert, S., Webb, D., and Fung, J. (2017). *The Costs and Losses of Wildfires, Special Publication (NIST SP)*. Gaithersburg, MD: National Institute of Standards and Technology. doi:10.6028/NIST.SP.1215 (Accessed December 29, 2024).
- Tong, Z., Chen, Y., Xu, Z., and Yu, R. (2023). Wise-iou: bounding box regression loss with dynamic focusing mechanism. arXiv Prepr. arXiv:2301.10051. doi:10.48550/arXiv.2301.10051
- Toreyin, B. U., and Cetin, A. E. (2009). “Wildfire detection using lms based active learning,” in 2009 IEEE international Conference on acoustics, speech and signal processing (IEEE), 1461–1464.
- Töreyn, B. U., Cinbiş, R. G., Dedeoğlu, Y., and Çetin, A. E. (2007). Fire detection in infrared video using wavelet analysis. *Opt. Eng.* 46, 067204. doi:10.1117/1.2748752
- Töreyn, B. U., Dedeoğlu, Y., and Cetin, A. E. (2006). “Contour based smoke detection in video using wavelets,” in 2006 14th European signal processing conference (IEEE), 1–5.
- Ultralytics (2020). Ultralytics-yolov5. Available at: <https://github.com/ultralytics/YOLOv5> (Accessed May 1, 2024).
- Ultralytics (2023). Ultralytics-yolov8. Available at: <https://github.com/ultralytics/ultralytics> (Accessed February 5, 2024).
- Vayadande, K., Gurav, R., Patil, S., Chavan, S., Patil, V., and Thorat, A. (2018). “Wildfire smoke detection using faster r-cnn,” in International conference on emerging trends in communication, computing and electronics (Springer), 141–164.
- Wang, C., He, W., Nie, Y., Guo, J., Liu, C., Wang, Y., et al. (2024). Gold-yolo: efficient object detector via gather-and-distribute mechanism. *Adv. Neural Inf. Process. Syst.* 36.
- Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2023). “Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 7464–7475.
- Wang, S., Zhao, J., Ta, N., Zhao, X., Xiao, M., and Wei, H. (2021). A real-time deep learning forest fire monitoring algorithm based on an improved pruned+ kd model. *J. Real-Time Image Process.* 18, 2319–2329. doi:10.1007/s11554-021-01124-9
- Wang, Y., Hua, C., Ding, W., and Wu, R. (2022). Real-time detection of flame and smoke using an improved yolov4 network. *Signal, Image Video Process.* 16, 1109–1116. doi:10.1007/s11760-021-02060-8
- Wang, Y., Wu, A., Zhang, J., Zhao, M., Li, W., and Dong, N. (2016). “Fire smoke detection based on texture features and optical flow vector of contour,” in 2016 12th world congress on intelligent control and automation (WCICA) (IEEE), 2879–2883.

- Xiao, Y., Wang, X., Zhang, P., Meng, F., and Shao, F. (2020). Object detection based on faster r-cnn algorithm with skip pooling and fusion of contextual information. *Sensors* 20, 5490. doi:10.3390/s20195490
- Xiaoxue, Z., Yu, W., Siyuan, W., and Bangyong, S. (2023). An improved lightweight fire detection algorithm based on cascade sparse query. *Opto-Electronic Eng.* 50, 230216–230221. doi:10.12086/oe.2023.230216
- Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). “Aggregated residual transformations for deep neural networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 1492–1500.
- Xin, C., Hartel, A., and Kasneci, E. (2024). Dart: an automated end-to-end object detection pipeline with data diversification, open-vocabulary bounding box annotation, pseudo-label review, and model training. *Expert Syst. Appl.* 258, 125124. doi:10.1016/j.eswa.2024.125124
- Xu, Y., Li, J., Zhang, L., Liu, H., and Zhang, F. (2024). Cntcb-yolov7: an effective forest fire detection model based on convnextv2 and cbam. *Fire* 7, 54. doi:10.3390/fire7020054
- Yu, G., Chang, Q., Lv, W., Xu, C., Cui, C., Ji, W., et al. (2021). Pp-picodet: a better real-time object detector on mobile devices. arXiv Prepr. arXiv:2111.00902.
- Yuanbin, W. (2016). “Smoke recognition based on machine vision,” in 2016 international symposium on computer, consumer and control (IS3C) (IEEE), 668–671.
- Yun, B., Zheng, Y., Lin, Z., and Li, T. (2024). Ffyolo: a lightweight forest fire detection model based on yolov8. *Fire* 7, 93. doi:10.3390/fire7030093
- Zhang, D. (2024). A yolo-based approach for fire and smoke detection in iot surveillance systems. *Int. J. Adv. Comput. Sci. and Appl.* 15. doi:10.14569/ijacsa.2024.0150109
- Zhang, Q.-X., Lin, G.-H., Zhang, Y.-M., Xu, G., and Wang, J.-J. (2018). Wildland forest fire smoke detection based on faster r-cnn using synthetic smoke images. *Procedia Eng.* 211, 441–446. doi:10.1016/j.proeng.2017.12.034
- Zhang, Y.-F., Ren, W., Zhang, Z., Jia, Z., Wang, L., and Tan, T. (2022). Focal and efficient iou loss for accurate bounding box regression. *Neurocomputing* 506, 146–157. doi:10.1016/j.neucom.2022.07.042
- Zhao, L., Zhi, L., Zhao, C., and Zheng, W. (2022). Fire-yolo: a small target object detection method for fire inspection. *Sustainability* 14, 4930. doi:10.3390/su14094930
- Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., et al. (2024). “Detrs beat yolos on real-time object detection,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 16965–16974.