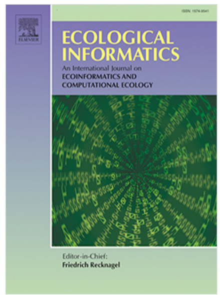
Journal Pre-proof



YOLO-MP: A lightweight forest fire detection model

Hongwei Zhu, Weiwei Ling, Huabiao Yan, Xinghai Zhong, Feng Liao

PII: S1574-9541(25)00525-4
DOI: https://doi.org/10.1016/j.ecoinf.2025.103516
Reference: ECOINF 103516

To appear in: *Ecological Informatics*

Received date : 1 July 2025
Revised date : 12 October 2025
Accepted date : 7 November 2025

Please cite this article as: H. Zhu, W. Ling, H. Yan et al., YOLO-MP: A lightweight forest fire detection model. *Ecological Informatics* (2025), doi: https://doi.org/10.1016/j.ecoinf.2025.103516.

YOLO-MP: A Lightweight Forest Fire Detection Model

Hongwei Zhu[a], Weiwei Ling[b,c,*], Huabiao Yan[a,*], Xinghai Zhong[a] and Feng Liao[a]

[a]College of Science, Jiangxi University of Science and Technology, Ganzhou, 341000, China
[b]Jiangxi College of Applied Technology, Ganzhou, 341000, China
[c]Key Laboratory of Ionic Rare Earth Resources and Environment, Ministry of Natural Resources, Ganzhou, 341000, China

ARTICLE INFO

Keywords:
Forest fire detection
Lightweight network
FPSConv
GERB
W-EIoU

ABSTRACT

Traditional methods for forest fire detection face challenges, including low efficiency, high cost, and significant susceptibility to environmental factors. Furthermore, existing deep learning approaches exhibit deficiencies in feature extraction and model lightweighting capabilities. To address these challenges, this paper proposes a lightweight forest fire detection model, designated YOLO-MP. This model first introduces a novel lightweight backbone network, Ghost-HGNetV2 (GHGNet), which effectively reduces the parameter count and computational cost, suppresses noise, and concurrently enhances the model's feature extraction capabilities. Subsequently, the Feature Pyramid Shared Convolution (FPSConv) module is incorporated for multiscale feature extraction, enabling the efficient capture of characteristic information related to forest fire targets. Following this, the Gradient Enhanced Reparameterized Block (GERB) is introduced to improve the efficiency of the lightweight model and enhance its gradient propagation capabilities. Additionally, a new loss function, Wise-Efficient IoU (W-EIoU), is designed to improve the model's learning capacity and generalization performance across samples of varying quality. Finally, the experimental results demonstrate that YOLO-MP achieves improvements over the baseline model in recall, mAP50, and mAP50-95 by 2.76%, 1.52%, and 1.18%, respectively. The model comprises only 2.07M parameters, representing a 31% reduction compared to the baseline, and requires 6.02 GFLOPs, a decrease of 26% from the baseline.

1. Introduction

In recent years, forest fires have frequently occurred worldwide. Characterized by their devastating destructiveness and the difficulty of containment, these fires have caused significant damage to socio-economic systems, human safety, and the environment. Consequently, forest fire detection has become critically important. Current detection methods primarily rely on manual inspections, sensor networks (Aslan et al., 2012), forest surveillance videos, and remote sensing satellites (Coen and Schroeder, 2013; Rodriguez-Jimenez et al., 2023). Manual inspections not only demand substantial human resources but also exhibit low efficiency. While sensor-based networks are effective for confined indoor spaces, their deployment in expansive forest areas faces challenges such as installation difficulties, high maintenance costs, expensive hardware, and susceptibility to false alarms. Surveillance cameras in forested regions offer limited coverage and restricted visibility, impeding large-scale monitoring. Although remote sensing satellites enable broad spatial coverage, their effectiveness is compromised by weather conditions and limitations in the accurate identification of early-stage fire zones.

With the continuous development of image processing technologies, researchers have increasingly explored machine learning for fire detection. For instance, Yang et al. (2023) addresses limitations of existing machine learning-based forest fire detection methods, which often ignore prior knowledge on the differing costs of missed detections and false alarms, and suffer from frequent misses and insufficient real-time performance. It proposes a pixel-level precision method, the Preferred Vector Machine (PreVM). By introducing an L_0-norm constraint for the fire class, PreVM ensures a high detection rate. Furthermore, an L_1-norm PreVM based on kernel functions is proposed to accelerate training and reduce false alarms. Borges and Izquierdo (2010) proposed a detection framework combining fire color features, skewness, and area size with a Bayesian classifier. Baek et al. (2023) presents a real-time fire detection method leveraging wavelet transforms and a multi-modeling framework. Its core idea is to extract features that represent the temporal dynamics of sensor signals using wavelet multiresolution properties. The optimal discriminative features are selected through a sequential forward floating search method tailored to fire types. Subsequently, a wavelet-based multi-model nearest neighbor detection model is constructed to accommodate different fire scenarios. Çelik and

*Corresponding author
E-mail addresses: lingweiwei@jxyy.edu.cn (W. Ling); yanhuabiao@jxust.edu.cn (H. Yan)

Demirel (2009) introduced a rule-based generic color model for flame pixel classification, which leverages the YCbCr color space to achieve effective separation of chrominance and luminance, thus improving detection performance. Ko et al. (2009) implemented a two-stage approach: initial candidate flame region detection via correlation algorithms, followed by final classification using a binary support vector machine (SVM) classifier with a radial basis function (RBF) kernel. Although these machine learning-based methods demonstrate promising results in forest fire detection, they heavily rely on manual feature extraction processes, where detection performance depends directly on the quality of handcrafted features. Furthermore, traditional image processing approaches remain vulnerable to environmental interference, which exacerbates feature extraction challenges. These approaches also suffer from high false positive rates and frequent misclassification in complex scenes.

In addition, some researchers have begun to apply meteorological and forest environmental data in combination with machine learning algorithms for forest fire prediction. For example, Sarkar et al. (2024) proposed several machine learning-based methods to identify forest fire-prone areas. By integrating satellite remote sensing, GIS technologies, and multiple environmental variables, they developed a high-precision prediction model that enables fine-scale fire risk zoning in northeastern India. Ahajjam et al. (2025) introduced a spatiotemporal clustering and ensemble machine learning model to predict wildfires in Alaska. By fusing dynamic features and mining spatiotemporal patterns, the model enhances the accuracy of predictions for wildfire occurrence and spread. Bhadoria et al. (2021) introduced a forest fire prediction model based on Random Vector Forest Regression (RVFR). By combining the strengths of Support Vector Machines and Random Forest regression, and introducing an enhancement strategy, their model addressed the limitations of traditional single models in capturing the coupled effects of multi-source environmental factors and achieved improved prediction performance. However, such approaches often rely on high-precision meteorological or environmental data from forest regions. Their generalizability is limited, and they typically demand significant computational resources. Moreover, their accuracy in predicting long-term or extreme fire events remains relatively unsatisfactory.

With the increasing attention given to deep learning in forest fire detection, many researchers have developed various neural network-based models. Li et al. (2022) proposed a model, ALFRNet, for forest fire smoke detection, which integrates the Internet of Things (IoT) with an improved YOLOv3 framework. The model enhances multi-scale feature extraction capabilities and employs Cluster-NMS (CNMS) to refine detection box localization for smoke with blurred edges, thereby improving detection efficiency and robustness. Lin et al. (2023) proposed the TCA-YOLO model, which integrates a Transformer encoder with Convolutional Neural Networks (CNN) to enhance global feature extraction for wildfire targets. Zhang et al. (2022a) introduced the FT-ResNet50 model, employing transfer learning combined with focal loss and Mish activation functions to improve the accuracy and generalizability of UAV-based forest fire recognition. Yang et al. (2024) integrated Depthwise Separable Convolutions into YOLOv8 and optimized gradient allocation using a dynamic non-monotonic mechanism, enhancing adaptability to complex flame features. Abramov et al. (2024) developed an early fire detection method based on semantic segmentation and an improved CNN to identify fires and smoke across diverse environments and weather conditions, although it requires substantial hardware resources. El-Madafri et al. (2023) proposed a multi-task learning framework incorporating multi-class confusion elements to enhance discriminative capabilities and reduce false alarms. Sathishkumar et al. (2023) adopted transfer learning with a Learning without Forgetting (LwF) strategy, enabling the model to retain classification performance on original tasks while learning new ones. These methods have made some progress, but challenges persist. Existing lightweight models often lack optimization for resource-constrained devices. During feature extraction, models exhibit low sensitivity in capturing critical information and insufficient noise suppression (Yang et al., 2025b), leading to a tendency to learn irrelevant features. In complex scenarios, forest fire targets are prone to false alarms and missed detections, resulting in suboptimal detection accuracy.

To address these limitations, this study proposes a YOLO-MP model that achieves higher detection accuracy while minimizing computational complexity and mitigating the issues of false positives and false negatives commonly observed in forest fire detection. The key contributions are outlined as follows:

(1) A novel lightweight backbone network, GHGNet, is proposed. Unlike HGNetV2 in RT-DETR (Zhao et al., 2024), GHGNet embeds Ghost operations within multi-stage bottlenecks and introduces cross-stage feature reuse and dynamic weighting mechanisms. This redesigned backbone maintains lightweight characteristics while significantly enhancing feature diversity, effectively addressing the challenges of feature extraction in forest fire detection.

(2) An FPSConv module is designed to replace the traditional Spatial Pyramid Pooling-Fast (SPPF) structure. FPSConv incorporates a shared-weight multi-scale Dilated Convolution mechanism, which expands the receptive

field through different dilation rates. This design enhances the model's contextual perception and multi-scale feature extraction capabilities.

(3) A lightweight module, GERB, is proposed by integrating the concepts of Cross Stage Partial Network (CSPNet) (Wang et al., 2020) and Extended Layer Aggregation Network (ELAN) (Wang et al., 2023). Unlike the existing C2f module, GERB introduces training-phase redundant paths and inference-phase reparameterization fusion, together with a configurable channel scaling factor. This allows the model to dynamically balance accuracy and computational efficiency. The design achieves efficient feature fusion and representation while maintaining network compactness, meeting the requirements of ecological monitoring.

(4) A new loss function, W-EIoU, is proposed by combining the advantages of Wise IoU v3 (WIoUv3) and Efficient IoU (EIoU). This function improves the stability of bounding-box regression and enhances localization accuracy across multiple scales. It is more adaptable to flame targets with irregular shapes and varying sizes, significantly improving the model's robustness in complex natural environments.

Forest fires are not only a natural hazard but also a serious ecological issue, causing large-scale ecological damage, biodiversity loss, and significant carbon emissions. Therefore, achieving earlier and more reliable forest fire detection in complex natural environments has become an urgent requirement for ecological conservation and sustainable forest management. This is crucial for reducing extensive forest loss, mitigating carbon emissions, and protecting biodiversity. The YOLO-MP model proposed in this study performs well in terms of both accuracy and lightweight design, while also providing a practical and scalable tool for ecological monitoring. Integrating innovations in computer vision with the needs of ecological conservation simultaneously promotes technological advancement and supports environmental sustainability.

The remainder of this paper is organized as follows: Section 2 describes the materials and methods, including detailed explanations of each module; Section 3 presents the experimental setup, ablation studies, comparative results, and performance statistics; Section 4 discusses the significance of the proposed method and its ecological applications; Section 5 analyzes the limitations of this study and outlines future work; Section 6 concludes the paper.

2. Materials and methods

2.1. YOLO-MP model structure

Utilizing YOLOv8n (Jocher et al., 2023a) as the baseline model, this paper proposes a forest fire detection model named YOLO-MP (the network architecture is depicted in Fig. 1). The following modifications were implemented in this model: Firstly, the HGNetV2 backbone network from RT-DETR was enhanced to create the GHGNet network, which reduces the number of parameters and enhances the model's feature extraction capabilities. Secondly, to address the issue of an insufficient receptive field in the baseline model, the FPSConv module was introduced. This module enables the model to better capture contextual information; compared to the pooling operations in SPPF, FPSConv can extract more fine-grained features. Subsequently, the GERB module was incorporated into the neck section of the network, replacing the C2f module. This module further reduces the parameter count and computational load. Concurrently, during the training process, it enables the learning of richer feature representations, thereby enhancing the model's generalization capability. Finally, the W-EIoU loss function was employed for the YOLO-MP model. W-EIoU provides a more precise measure of the discrepancy between the model's predictions and the ground-truth values and demonstrates greater sensitivity in handling small targets.

2.2. Improved backbone network for feature extraction

The backbone network serves as the foundation of the model. In YOLOv8, the backbone utilizes an enhanced CSPDarknet53 (Redmon and Farhadi, 2018) architecture, which incorporates cross-stage partial connections to optimize inter-layer information flow and improve detection accuracy. A robust feature extraction network significantly enhances recognition performance. HGNetV2, developed by the Baidu research team, improves upon HGNetV1 with structural refinements, demonstrating superior performance in object detection tasks. Compared to YOLOv8's backbone, HGNetV2 achieves a more lightweight design through efficient convolutional modules, reducing computational costs and parameter counts. Additionally, HGNetV2 employs stacked HGBlocks for hierarchical feature extraction, facilitating fusion between low-level and high-level features. The GHGNet used in the YOLO-MP model is an improved version of HGNetV2, as illustrated in Fig. 2.

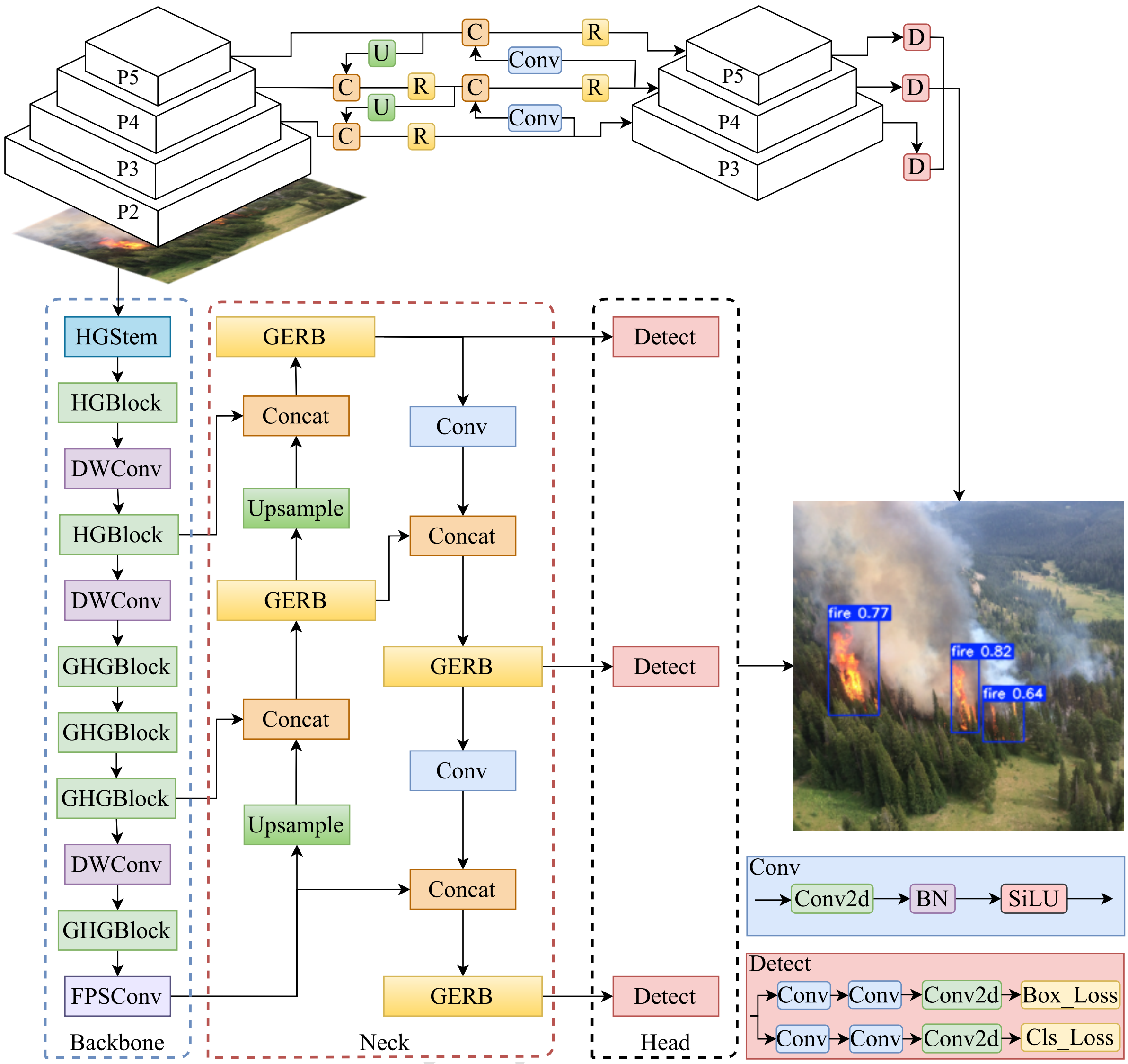


Fig. 1. Network structure diagram of YOLO-MP model.

The GHGNet initially extracts features through the HGStem (structure shown in Fig. 2(b)), skipping the P1 layer and directly constructing the feature pyramid starting from P2. The high-resolution features of the P1 layer provide limited gain for small object detection while incurring a higher computational cost. Subsequently, multiple stages are employed for feature transformation and processing. In these stages, the HGBlock (structure shown in Fig. 2(c)) and the GHGBlock (structure shown in Fig. 2(e)) are employed to refine and enhance the features. Depthwise Convolution (DWConv; structure shown in Fig. 2(d)) (Chen et al., 2023) is applied for downsampling operations, offering higher computational efficiency. The GHGBlock integrates the Ghost module philosophy with hierarchical feature aggregation, forming a lightweight convolutional neural network module. It reduces computational complexity while enhancing feature representation capability through dynamic feature reuse, multi-branch feature transformation, and cross-layer information interaction.

The process of generating feature maps by GhostConv (Han et al., 2020) can be divided into three simple steps: first, a set of primary feature maps containing effective feature information is generated through a convolution operation. Subsequently, a series of 'cheap' linear operations is applied to each primary feature map to generate multiple Ghost feature maps. Finally, the primary feature maps and the Ghost feature maps from the second step are concatenated to produce the final output. In summary, GhostConv generates less redundant feature representations by linearly combining low-rank feature maps, significantly reducing the computational waste caused by excessive inter-channel correlation in standard convolution. Fig. 3 shows schematic diagrams of feature map generation for Standard Convolution and GhostConv. Here, (a) illustrates the feature map generation for Standard Convolution, while (b) illustrates the feature map generation for GhostConv, with ‘p’ denoting the linear operation.

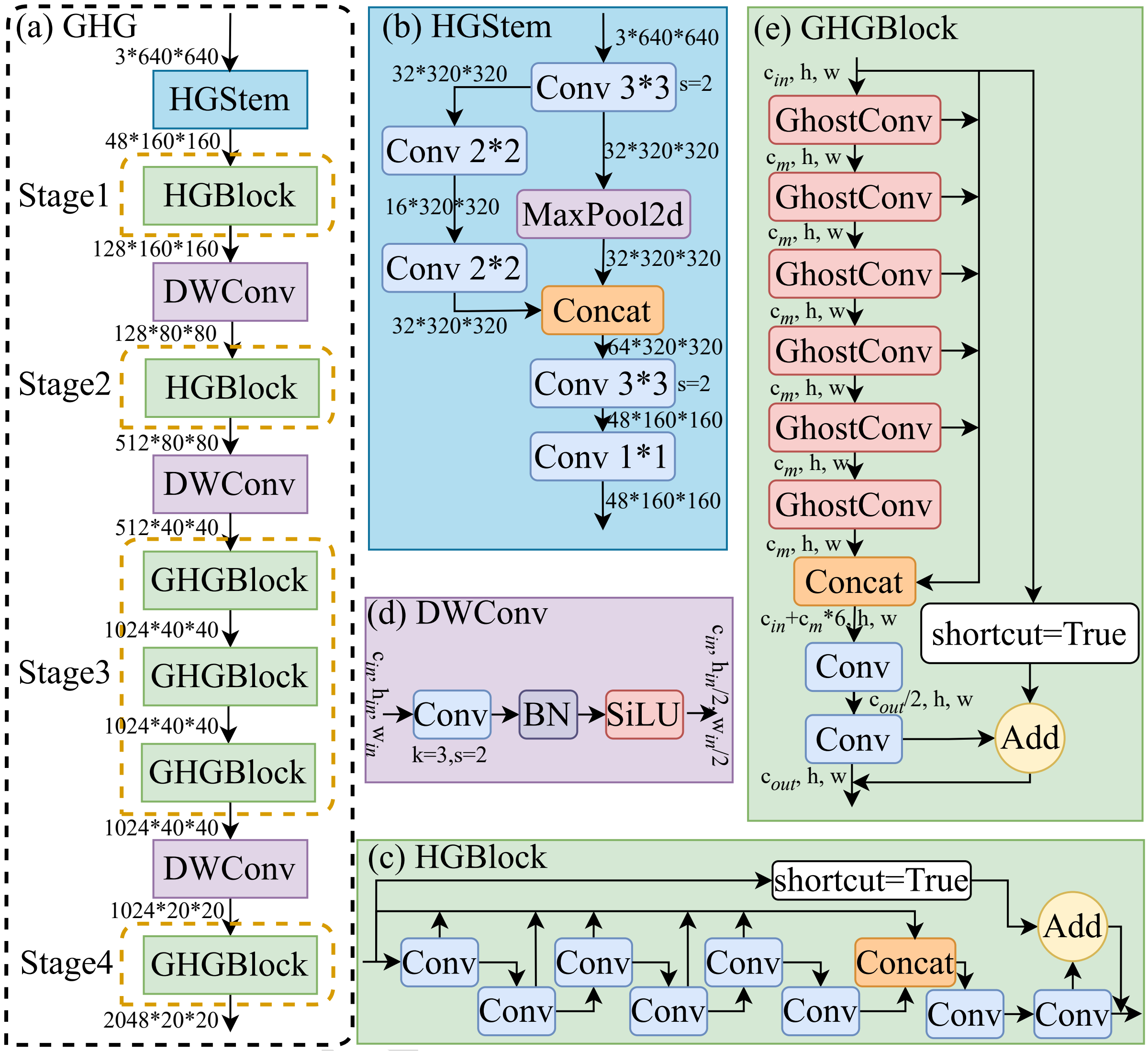


Fig. 2. GHGNet Structure. (a) illustrates the overall structure of GHGNet; (b) shows a schematic of the HGStem; (c) shows a schematic of the HGBlock; (d) shows a schematic of the DWConv; and (e) shows a schematic of the GHGBlock. The input size is given as an example of 3*640*640. Here, k denotes the kernel size, s the stride, c_{in} the number of input channels, c_m the number of intermediate channels, and c_{out} the number of output channels. The parameter dimensions of HGBlock follow those of GHGBlock. For panels (c) and (e), when shortcut=True, $c_{out} = c_{in}$; when shortcut=False, $c_{out} = 2 * c_{in}$.

Although GhostConv reduces model parameters and computational costs through cost-effective linear operations, excessive usage can compromise detection accuracy. To address this trade-off, GHGNet retains Standard Convolutional layers in Stage 1 and Stage 2 for foundational feature extraction, striking a balance between precision and computational efficiency. In comparison to the backbone network of YOLOv8, GHGNet enables efficient feature extraction and downsampling in the initial stage by effectively combining convolutional kernels of different sizes. This rapidly reduces feature map dimensions and increases channel numbers, preparing for subsequent processing. Moreover, its multi-stage hierarchical design is particularly well-suited for handling complex scenarios like forest fires.

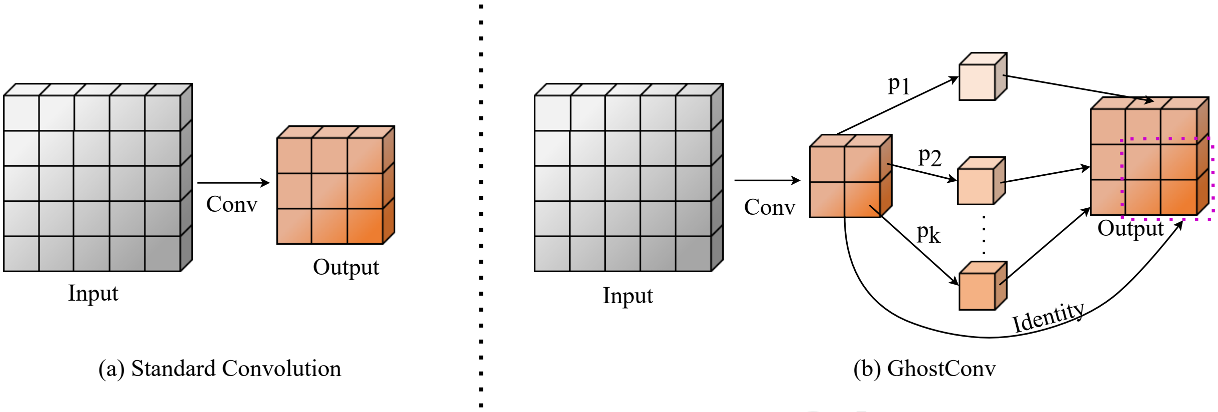


Fig. 3. Schematic diagram of feature map generated by Standard Convolution and GhostConv. (The p in subfigure (b) represents a linear operation.)

It is important to emphasize that all experiments were conducted on exactly the same dataset, without introducing any additional or higher-quality data. The term "parameter reduction" refers solely to the lightweight design of the model architecture, rather than a reduction in input information. The performance advantage of GHGNet arises from its higher parameter utilization efficiency. Through the use of GhostConv for low-redundancy feature generation, GHGBlock for multi-branch cross-layer fusion, and the rational skipping of the P1 layer, the model reduces the number of learnable parameters while simultaneously enhancing its ability to discriminate critical fire-related features. Consequently, YOLO-MP achieves more accurate detection results under the same input conditions, while maintaining low computational cost.

2.3. *FPSConv module*

The SPPF is commonly used for multi-scale feature extraction. Compared to the traditional SPP layer, SPPF offers improved detection accuracy and speed. However, it has limitations. SPPF primarily employs pooling operations with differently-sized pooling kernels to extract multi-scale features. This method is relatively fixed, with predefined kernel sizes, resulting in limited receptive field flexibility and partial spatial information loss during pooling. To address these issues, the YOLO-MP model introduces the FPSConv module to replace SPPF, as shown in Fig. 4. FPSConv uses Dilated Convolutions (Strubell et al., 2017) with different dilation rates for multi-scale feature extraction. It flexibly adjusts dilation rate parameters to control the receptive field size, enabling the extraction of features at different scales. Unlike pooling operations, Dilated Convolutions do not significantly reduce spatial resolution, thus better preserving the original image's spatial information.

The FPSConv module processes input features through a sequence of operations (as formulated in Eq. (1)). Initially, the input undergoes a 1×1 convolutional layer for preliminary feature transformation and channel adjustment. Subsequently, shared 3×3 convolutions with varying dilation rates are applied to capture multi-scale contextual features. These convolutions utilize hierarchical receptive fields by adjusting dilation parameters while sharing kernel weights across scales, thereby minimizing computational overhead. All extracted features are then concatenated to integrate multi-scale information. Finally, a 1×1 convolutional layer fuses the concatenated features across channels, optimizing dimensionality and refining the aggregated representations for subsequent network layers. This design ensures efficient multi-scale feature extraction while maintaining spatial fidelity through Dilated Convolutions, effectively mitigating information loss commonly encountered in traditional pooling-based approaches.

$$Y = C_{1\times1}\left(\bigoplus_{i=1}^{n} C_{3\times3}^{d_i}\left(C_{1\times1}(X)\right)\right), \; n = 1, 3, 5 \tag{1}$$

where Y represents the output, C represents the convolution operation, $\bigoplus$ represents the concatenation operation along the channel dimension, d_i represents the dilation rate of i, and X represents the input.

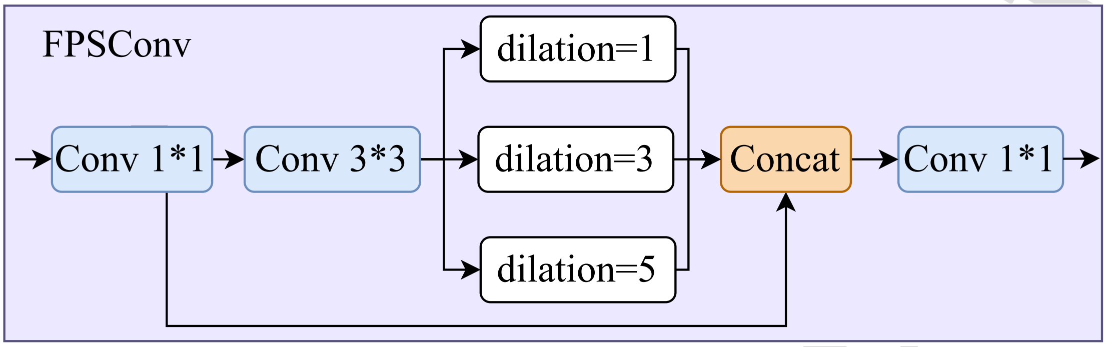
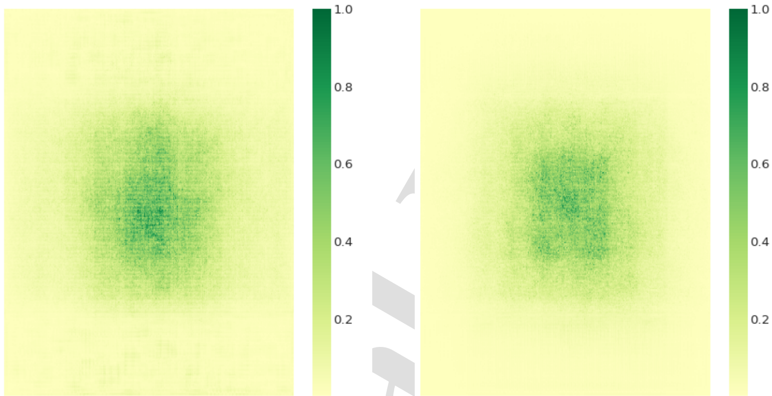


Fig. 4. FPSConv module structure.



(a) Using FPSConv's receptive field. (b) Using SPPF's receptive field.

Fig. 5. Visualization of receptive field.

The receptive field size, which determines the input image information range captured by the network, is enlarged by FPSConv. A larger receptive field enables the network to capture more global and extensive information. To better understand FPSConv's capabilities, we visualize its receptive field following the method in Ding et al. (2022), as shown in Fig. 5. The figure clearly shows that (a) outperforms (b) in both small-range green density (for more precise target localization) and large-range green diffusion (for multi-scale feature extraction and global information capture).

2.4. GERB module

The C2f module is primarily used for feature-map transformation and fusion of features at different hierarchical levels. C2f adopts the CSPNet design philosophy by introducing cross-stage connections to enhance the model's learning capacity. To further reduce the model's parameter count and computational cost, thereby better accommodating devices with limited computational resources, the YOLO-MP model incorporates the GERB module. This module combines the design principles of CSPNet and ELAN, employing a more efficient feature-transformation and fusion strategy to boost model performance. The architecture of the GERB module is illustrated in Fig. 6.

The module first expands the input channel dimension via a 1×1 convolution, followed by splitting the feature maps into two branches. One branch undergoes a series of complex convolutional operations, while the other is preserved. Subsequently, all feature maps from both branches are concatenated along the channel dimension and processed through a 1×1 convolution to generate the final output. The GERB module discards the original C2f bottleneck layer. To mitigate the resulting accuracy degradation, a RepConv (Ding et al., 2021) is introduced in one branch. By employing a multi-branch convolutional structure, this design enhances feature-transformation capability and accelerates inference. The architecture of RepConv is shown in Fig. 7.

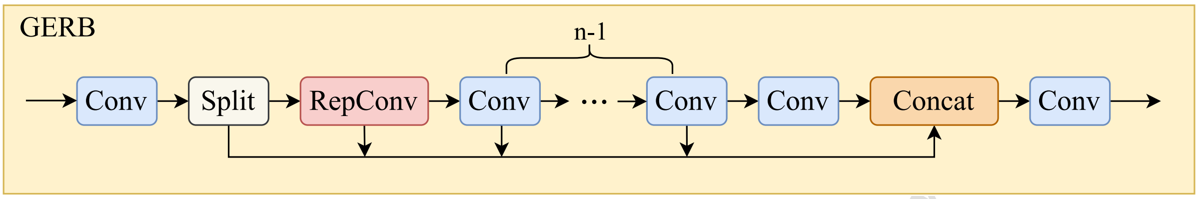


Fig. 6. GERB module structure.

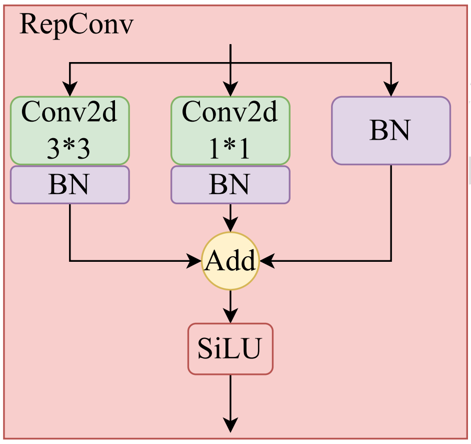


Fig. 7. RepConv Structure.

The RepConv module employs a multi-branch architecture, where the branches operate in parallel to extract distinct feature representations from the input. RepConv offers a novel approach to model lightweighting through its design philosophy of "complexifying training while simplifying inference." It is worth noting that the principle of “complex training but simplified inference” is not the overall training philosophy of the proposed model, but rather a specific design concept of the RepConv module. By adopting this strategy, RepConv maintains diverse representational capacity during training while being re-parameterized into a simple convolutional structure during inference, thereby improving efficiency without sacrificing accuracy. Its working principle is described in Eq. (2).

$$Y = \sigma\left(BN\left(W_{3\times3} * X + b_{3\times3}\right) + BN\left(W_{1\times1} * X + b_{1\times1}\right) + BN(X)\right) \tag{2}$$

where Y represents the output, X represents the input, $W_{3\times3}$ and $b_{3\times3}$ represent the weights and biases of the 3×3 convolution, $W_{1\times1}$ and $b_{1\times1}$ represent the weights and biases of the 1×1 convolution, $*$ represents the convolution operation, BN represents the normalization operation, and σ represents the activation function.

The GERB module addresses the issue of information loss that commonly occurs in bottleneck layers when processing complex feature representations. By incorporating RepConv, which exhibits enhanced feature extraction capabilities and improved gradient flow, this design mitigates the risks of gradient vanishing or exploding during training. Specifically, during the training phase, multiple parallel branches within the RepConv learn distinct feature representations from the input data. During inference, these branches are fused into a single equivalent convolutional layer, simplifying the network structure. On the one hand, the model can perform inference more efficiently. On the other hand, the structural simplification does not cause a significant drop in performance, and the model still maintains high detection accuracy.

2.5. W-EIoU loss

YOLOv8 utilizes Distribution Focal Loss (DFL) and Complete IoU (CIoU) Loss (Zheng et al., 2020) for bounding box regression. The CIoU loss integrates overlap area, centroid distance, and aspect ratio to compute the regression loss, which facilitates faster convergence. However, CIoU has several deficiencies: it lacks a balancing mechanism between easy and hard samples, inaccurately penalizes aspect ratio differences, and introduces additional computational burden due to the use of inverse trigonometric functions.

EIoU Loss (Zhang et al., 2022b) improves upon CIoU in two main aspects. First, EIoU computes the width and height losses as independent terms, which directly penalize the discrepancies between the predicted box and the ground-truth box in terms of width and height. Second, the structure of EIoU's loss function is relatively simpler, which reduces computational complexity.

To further enhance adaptability in complex fire detection scenarios, we propose W-EIoU, which introduces two critical innovations. First, it enhances the penalty mechanism by squaring and summing the relative differences in center point coordinates, thus providing more precise error feedback. Second, a dynamic non-monotonic focusing mechanism is introduced, which adaptively adjusts the loss weights to focus more effectively on hard samples. This mechanism is more effective in addressing complex forest fire detection scenarios, particularly those characterized by occlusion and diverse target scales. The penalty term of the CIoU loss is formulated as shown in Eqs. (3)–(6).

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{3}$$

$$L_{IoU} = 1 - IoU \tag{4}$$

$$\alpha = \frac{v}{L_{IoU} + v} \tag{5}$$

$$L_{CIoU} = L_{IoU} + \frac{\rho^2(b, b^{gt})}{c_w^2 + c_h^2} + \alpha v \tag{6}$$

where v represents the difference in aspect ratio between the predicted bounding box and the ground-truth bounding box; h and h^{gt} denote the height of the predicted bounding box and the height of the ground-truth bounding box, respectively; w and w^{gt} represent the width of the predicted bounding box and the width of the ground-truth bounding box, respectively; IoU is a metric used to measure the degree of overlap between two bounding boxes; $\rho^2(b, b^{gt})$ indicates the squared distance between the centers of the predicted bounding box and the ground-truth bounding box; α represents the weighting function; c_w and c_h correspond to the width and height of the smallest enclosing region that can simultaneously cover both the predicted bounding box and the ground-truth bounding box.

The penalty term of the EIoU loss is formulated as shown in Eq. (7).

$$L_{EIoU} = L_{IoU} + \frac{\rho^2(b, b^{gt})}{c_w^2 + c_h^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \tag{7}$$

where $\rho^2(w, w^{gt})$ denotes the squared difference between the width of the predicted bounding box w and the width of the ground-truth bounding box w^{gt}, and $\rho^2(h, h^{gt})$ represents the squared difference between the height of the predicted bounding box h and the height of the ground-truth bounding box h^{gt}.

The penalty term of the W-EIoU loss is formulated as shown in Eqs. (8)–(10).

$$\beta = \frac{L_{IoU}^{\#}}{L_{IoU}} \in [0, +\infty) \tag{8}$$

$$\phi = \frac{\beta}{\mu\alpha^{\beta-\mu}} \tag{9}$$

$$L_{W-EIoU} = \left(L_{IoU} + \frac{\rho^2(b, b^{gt})}{c_w^2 + c_h^2} + \left(\frac{x - x_t}{c_w}\right)^2 + \left(\frac{y - y_t}{c_h}\right)^2 \right) \times \phi \tag{10}$$

where the outlier degree β of the anchor box quality is defined as the ratio of the current sample's IoU loss value ($L^{\#}_{IoU}$) to the average IoU loss value (L_{IoU}) of all samples; μ and α are hyperparameters that control the degree of focalization; x and y represent the horizontal and vertical coordinates of the predicted bounding box's center point, respectively; x_t and y_t denote the horizontal and vertical coordinates of the ground-truth bounding box's center point, respectively.

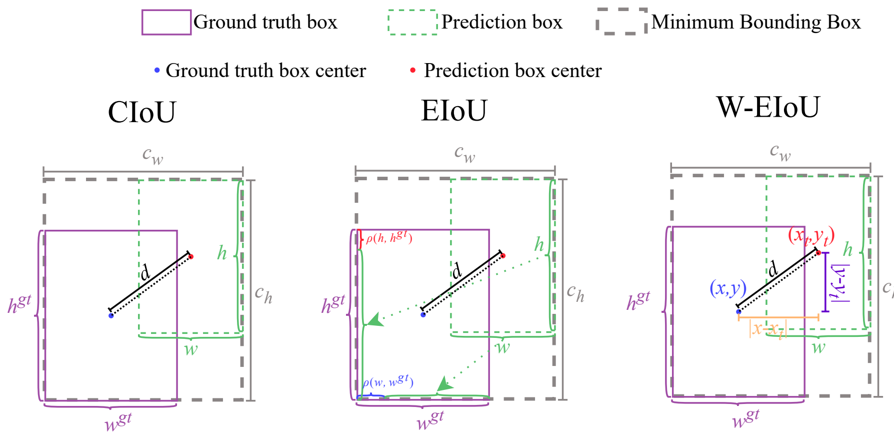


Fig. 8. Schematic diagram of loss function. d is the distance between two center points ($d = \rho(b, b^{gt})$).

The schematic diagrams of CIoU, EIoU, and W-EIoU are shown in Fig. 8. Compared with EIoU, W-EIoU does not penalize the width and height errors directly but instead adjusts them indirectly via the center point position error. Directly penalizing width and height errors could diminish the influence of ϕ, as these errors are localized and fail to reflect the overall difficulty of a sample. W-EIoU further enhances the penalty mechanism for positional deviations, with a specific focus on the relative center point offset with respect to the minimum enclosing rectangle. Using a non-monotonic focusing mechanism, gradient gains are dynamically allocated based on the degree of anchor box outlierness. The center point error acts as a unified control over the entire loss term, allowing ϕ to weight the overall loss adaptively and thus better handle difficult samples. W-EIoU enhances the model's learning capacity and generalization performance across samples of varying quality, which is particularly critical for detecting irregularly shaped forest fire targets in complex environments.

3. Experiments

3.1. Dataset

The forest fire dataset employed in this study was compiled from publicly available sources, video frame extraction, and curated web-scraped images. Public datasets include Fire Luminosity Airborne-based Machine learning Evaluation (FLAME) (Shamsoshoara et al., 2021) from Northern Arizona University and The Wildfire Dataset (El-Madafri et al., 2023) from Kaggle. After aggregation and filtering, a total of 2,558 raw images capturing diverse forest fire scenarios and perspectives were obtained and annotated using LabelImg software. Representative samples are illustrated in Fig. 9. To enhance model generalization and expand data diversity, multiple augmentation techniques were applied, including mixup, shifting, flipping, rotation, scaling, blurring, padding, cropping, translation, and affine transformation. The augmented dataset consists of 7,000 images. For performance evaluation, overfitting prevention, and optimization guidance, the dataset was partitioned into training, validation, and test sets at an 8:1:1 ratio.

Fig. 9. Example images of forest fire dataset.

Table 1
Training Environment Configuration.

Component	Specification
CPU	Intel(R) Xeon(R) Platinum 8255C @ 2.50GHz
GPU	NVIDIA Tesla T4 (16GB VRAM)
Operating System	Ubuntu 20.04.6
Framework	Pytorch2.0
Python Version	Python3.11
CUDA Version	CUDA11.8

3.2. Training configuration & Evaluation metrics

To ensure experimental reliability and isolate environmental variability, all experiments were conducted under identical hardware and software conditions. The training environment specifications are detailed in Table 1. Hyperparameters (summarized in Table 2) were initially configured based on YOLOv8's default settings and adjusted iteratively according to validation metrics. Model convergence analysis revealed optimal convergence at 350 epochs, with no significant improvement observed beyond this threshold.

The model's performance was comprehensively evaluated using six key metrics: Precision (P), Recall (R), Average Precision (AP), mean Average Precision (mAP), Parameters (Params), and FLOPs. Precision quantifies the model's ability to minimize false positive predictions, with higher values indicating fewer incorrect positive classifications. Recall measures the model's effectiveness in identifying all true positive instances, thus reflecting the likelihood of correctly detecting a positive sample. AP, calculated as the area under the P-R curve, provides a balanced assessment of

Table 2
Hyperparameter Settings.

Hyperparameters	Value
Image Size	640×640
Epochs	350
Batch-size	32
Initial Learning Rate	0.01
Optimizer	SGD
Optimizer Momentum	0.937
Weight Decay	0.0005

the trade-off between precision and recall across varying confidence thresholds. mAP, derived by averaging AP values across all classes, offers a holistic measure of multi-class detection accuracy. The calculation formulas are shown in Eqs. (11)–(15).

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$AP = \int_0^1 P(R)\, dR \tag{13}$$

$$mAP = \frac{1}{n} \sum_{i=1}^{n} AP_i \tag{14}$$

$$mAP50{-}95 = \frac{1}{N} \sum_{t=1}^{N} mAP(t) \tag{15}$$

where TP represents the number of positive samples that are actually positive and correctly predicted as positive; FP represents the number of negative samples incorrectly predicted as positive; FN represents the number of positive samples incorrectly predicted as negative; n denotes the number of categories; AP_i denotes the average precision of category i; N denotes the number of IoU thresholds, usually 10, ranging from 0.5 to 0.95 with a step size of 0.05; and $mAP(t)$ denotes the mAP at IoU threshold t.

In object detection tasks, mAP with varying IoU thresholds is commonly used. For example, mAP50 refers to the mAP obtained when evaluating the model at an IoU threshold of 0.5, whereas mAP50-95 represents the mAP calculated across multiple IoU thresholds. This metric provides a more comprehensive assessment of the model's performance under varying localization accuracy levels.

3.3. Ablation experiments

To evaluate the performance of the proposed modules, we designed a series of ablation experiments and assessed the results using the previously mentioned metrics. The results of the ablation experiments are presented in Table 3. From the results, it is evident that GHGNet significantly reduces the number of parameters while improving P, R, mAP50, and mAP50-95 to varying degrees. Although GERB shows a slight decline in R, it further reduces both the parameter count and computational complexity of the model. Meanwhile, FPSConv demonstrates substantial improvements over the baseline model across most metrics. Compared to the baseline model, the improved model achieves increases of 2.11%, 1.42%, and 1.15% in R, mAP50, and mAP50-95, respectively. Additionally, the parameter count and computational complexity are reduced by 31% and 26%, respectively, compared to the baseline model.

In addition to validating the effectiveness of each module, this study also conducts ablation experiments by replacing the backbone network in the baseline model and substituting different loss functions in the improved model. The results of these experiments are presented in Table 4 and Table 5, respectively.

To enable a more intuitive comparison of performance across different backbone networks, mAP50, mAP50-95, Params, and FLOPs are visualized in Fig. 10. In this plot, small purple bubbles closer to the top-left corner indicate that the corresponding backbone networks exhibit higher accuracy and superior lightweight design characteristics.

Table 3
Experimental results of module ablation.

	GHGNet/A	GERB/B	FPSConv/C	P	R	mAP50	mAP50-95	Params/M	FLOPs/G	Latency/ms
YOLOv8n	-	-	-	0.8693	0.7751	0.8528	0.5360	3.01	8.09	4.75
+A	✓	-	-	0.8723	0.7839	0.8610	0.5442	2.31	6.82	5.66
+B	-	✓	-	0.8734	0.7520	0.8478	0.5203	2.20	6.05	4.47
+C	-	-	✓	0.8761	0.7832	0.8584	0.5417	3.15	8.09	4.66
+A+B	✓	✓	-	0.8709	0.7841	0.8547	0.5323	1.92	6.02	5.48
+A+C	✓	-	✓	0.8790	0.7786	0.8634	0.5463	2.46	6.82	5.48
+B+C	-	✓	✓	0.8645	0.7835	0.8594	0.5422	2.76	7.29	4.82
+A+B+C	✓	✓	✓	0.8737	0.7962	0.8670	0.5475	2.07	6.02	5.53

Note: Latency = preprocess + inference + postprocess, and the latency was measured on the NVIDIA Tesla T4 GPU with an input image size of 640*640.

Table 4
Experimental results of backbone network ablation.

Different backbone networks	P	R	mAP50	mAP50-95	Params/M	FLOPs/G	Latency/ms
HGNetV2 (Zhao et al., 2024)	0.8663	0.7818	0.8571	0.5460	2.35	6.90	5.78
StarNet (Ma et al., 2024)	0.8654	0.7701	0.8516	0.5278	2.21	6.47	5.64
ShuffleNetv2 (Ma et al., 2018)	0.8727	0.7851	0.8569	0.5412	2.79	7.42	8.39
MobileNetv4 (Qin et al., 2025)	0.8478	0.7488	0.8323	0.5019	5.70	22.54	6.15
EfficientViT (Liu et al., 2023)	0.8574	0.7818	0.8569	0.5369	4.01	9.42	26.14
GHGNet	0.8723	0.7839	0.8610	0.5442	2.31	6.82	5.66

Table 5
Ablation experiment results of different loss functions in the improved model

Different loss functions	P	R	mAP50	mAP50-95	Params/M	FLOPs/G	Latency/ms
DIoU (Zheng et al., 2020)	0.8828	0.7774	0.8570	0.5303	2.07	6.02	5.72
GIoU (Rezatofighi et al., 2019)	0.8825	0.7794	0.8647	0.5470	2.07	6.02	5.56
SIoU (Gevorgyan, 2022)	0.8699	0.7926	0.8657	0.5403	2.07	6.02	5.68
WIoUv3 (Tong et al., 2023)	0.8678	0.7908	0.8670	0.5408	2.07	6.02	5.68
W-EIoU	0.8757	0.8027	0.8680	0.5478	2.07	6.02	5.48

Overall, GHGNet demonstrates strong performance across all evaluated aspects. Although GHGNet is slightly lower in R than ShuffleNetV2, it achieves better performance than ShuffleNetV2 in other metrics and is also significantly better than other lightweight backbone networks. In Table 5, although the proposed loss function yields lower P values than DIoU loss and GIoU loss, it outperforms DIoU loss and GIoU loss in other metrics by optimizing the quality of bounding box regression, and also surpasses other loss functions. These results validate the balanced design of GHGNet in feature extraction efficiency and discriminative ability, as well as the optimization potential of W-EIoU for target localization accuracy in complex scenes.

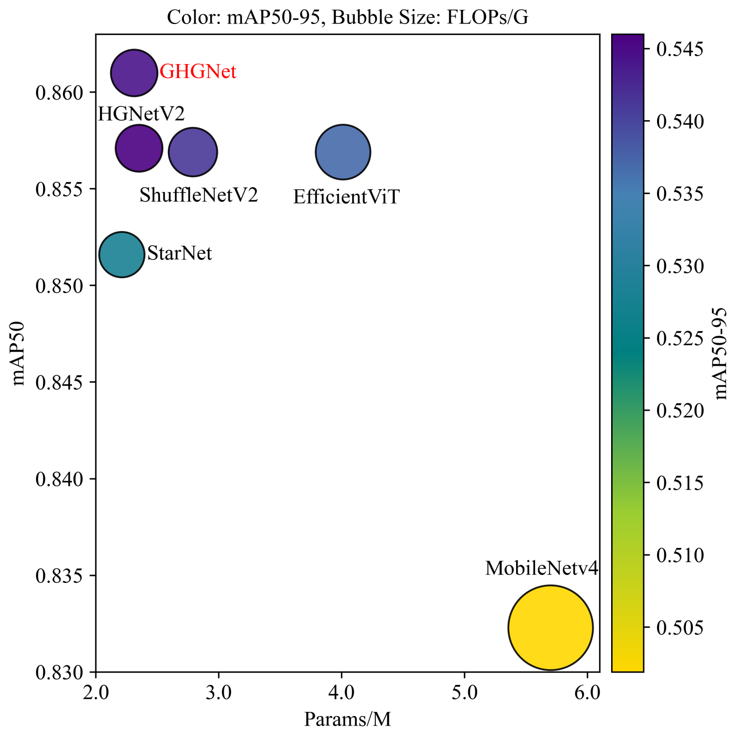


Fig. 10. Bubble chart of ablation test results of different backbone networks.

3.4. Comparative analysis with other models

By comparing the forest fire target detection results of the proposed YOLO-MP model with those of the unmodified YOLOv8 model, we observed that YOLO-MP achieves significant improvements in detection performance. To gain a more comprehensive understanding of the YOLO-MP model's capabilities, we designed a series of comparative experiments. The results of these comparative experiments are presented in Table 6.

Based on the data in Table 6, it is evident that the YOLO-MP model achieves higher detection accuracy while maintaining a relatively small parameter count and computational complexity. Specifically, the mAP50 of YOLO-MP is 1.52% higher than that of YOLOv8n, and its mAP50-95 is 1.18% higher. Compared to the model with the second-lowest parameter count, the proposed model reduces parameters by 8.4%, and by 31.23% relative to the baseline model.

To better demonstrate the effectiveness of the proposed model, we selected several images from the test set for visual comparison and heatmap analysis. The corresponding heatmaps illustrate the likelihood of object detection through variations in color intensity, providing an intuitive explanation of which regions contribute most to the model's predictions. Specifically, warmer colors indicate higher model attention and importance toward fire-related targets, whereas cooler colors indicate lower attention and importance. The comparison results are shown in Figure Fig. 11, with the left images presenting the detection results of the baseline model and the right images showing the results of YOLO-MP. Even in the presence of objects visually similar to forest fire targets, YOLO-MP accurately identifies the actual fire-related instances. These heatmaps not only visually confirm the improvement in detection accuracy of YOLO-MP but also highlight the most informative regions of the model, providing a transparent explanation of its decision-making process.

The comparison of detection performance for occluded small targets is illustrated in Fig. 12. The YOLOv8 model fails to recognize the occluded small forest fire targets, whereas the YOLO-MP model successfully identifies them. As

Table 6
Comparison experiment results.

Different models	P	R	mAP50	mAP50-95	Params/M	FLOPs/G	Latency/ms
Faster R-CNN (Ren et al., 2017)	0.6530	0.4347	0.4929	0.2353	41.75	91.48	92.00
SSD (Liu et al., 2016)	0.7344	0.7011	0.7363	0.3988	23.75	136.81	45.54
YOLOv5n (Jocher et al., 2022)	0.8764	0.7619	0.8452	0.5221	2.50	7.06	5.32
YOLOv8n	0.8729	0.7751	0.8528	0.5360	3.01	8.09	4.75
YOLOv9t (Wang et al., 2025)	0.8544	0.7535	0.8361	0.4942	2.62	10.74	17.98
YOLOv10n (Wang et al., 2024)	0.8703	0.7616	0.8526	0.5432	2.26	6.52	6.25
YOLOv11n (Jocher et al., 2023b)	0.8660	0.7832	0.8459	0.5283	2.58	6.31	6.22
Hyper-YOLO (Feng et al., 2025)	0.8642	0.7832	0.8552	0.5477	3.94	10.76	8.63
YOLOv12n (Tian et al., 2025)	0.8234	0.7066	0.7626	0.3791	2.51	5.82	12.39
YOLOv13n (Lei et al., 2025)	0.7815	0.6865	0.7219	0.3434	2.45	6.20	18.49
RT-DETR (Zhao et al., 2024)	0.8641	0.7950	0.8714	0.5482	18.83	29.68	28.80
YOLO-MP (This paper)	0.8757	0.8027	0.8680	0.5478	2.07	6.02	5.48

Note: All the above experiments were conducted under the same environment

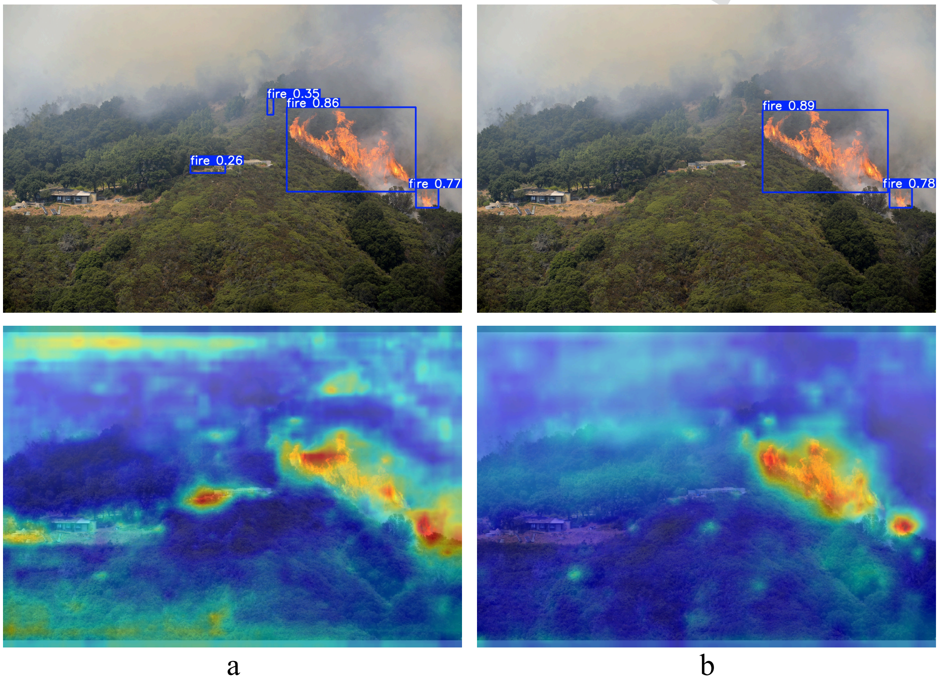


Fig. 11. Anti-interference comparison. (The figure below shows the corresponding heatmap for the figure above.) (a) YOLOv8 model error detection. (b) YOLO-MP model normal recognition without error detection.

shown in Fig. 13, the YOLOv8 model produces overlapping detection boxes during prediction, whereas the YOLO-MP model improves localization accuracy and minimizes redundant detection boxes. As shown in Fig. 14, in the forest fire detection task, the YOLOv8 model struggles to effectively extract global contextual information of the fire targets. Its detection boxes only mark localized burning areas, resulting in inaccurate localization of the fire's extent. In contrast, the YOLO-MP model achieves precise detection of large-scale forest fire targets and provides more comprehensive regional localization through multi-scale global information integration.

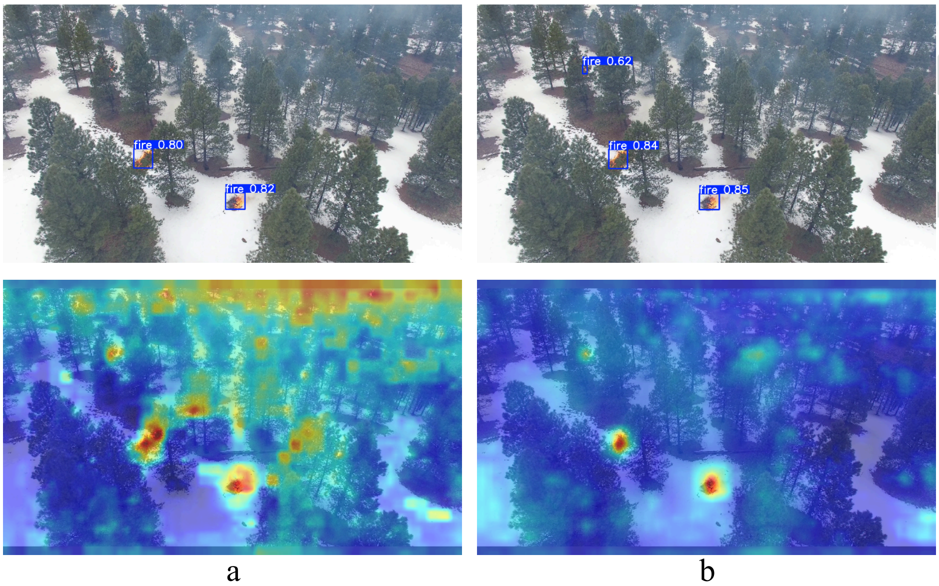


Fig. 12. Comparison of recognition of extremely small occluded targets. (The figure below shows the corresponding heatmap for the figure above.) (a) YOLOv8 model: missed detection. (b) YOLO-MP model: normal recognition without missed detection.

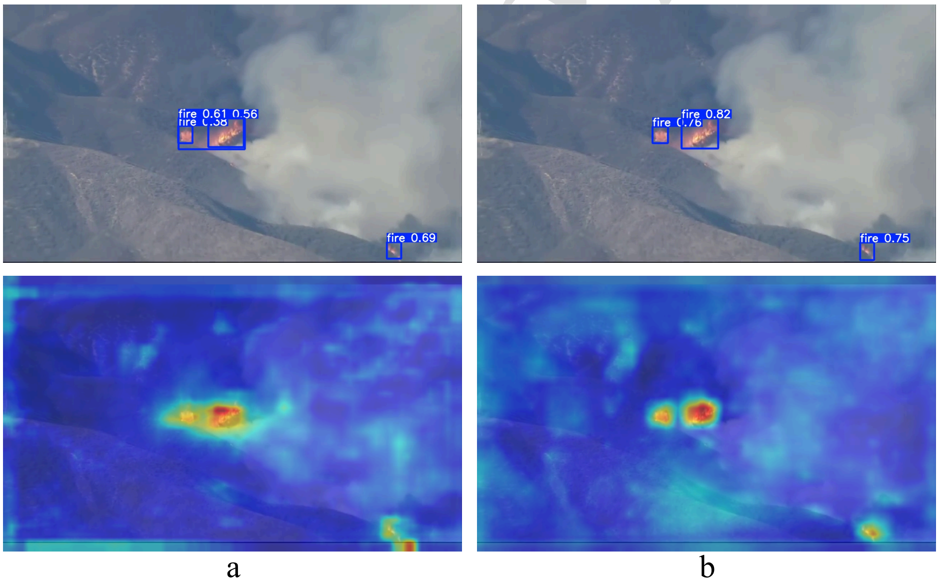


Fig. 13. Comparison of detection box redundancy. (The figure below shows the corresponding heatmap for the figure above.) (a) Overlapping detection boxes in YOLOv8. (b) Non-overlapping detection boxes in the YOLO-MP model.

3.5. Statistical validation

To comprehensively evaluate the performance stability of the proposed model, we adopted a statistical analysis approach based on multiple experiments. Specifically, under identical experimental settings (including hardware, software, and dataset configurations), the model was trained and tested in five independent runs, and performance indicators such as mAP50 and mAP50-95 were recorded. The detailed results are presented in Table 7.

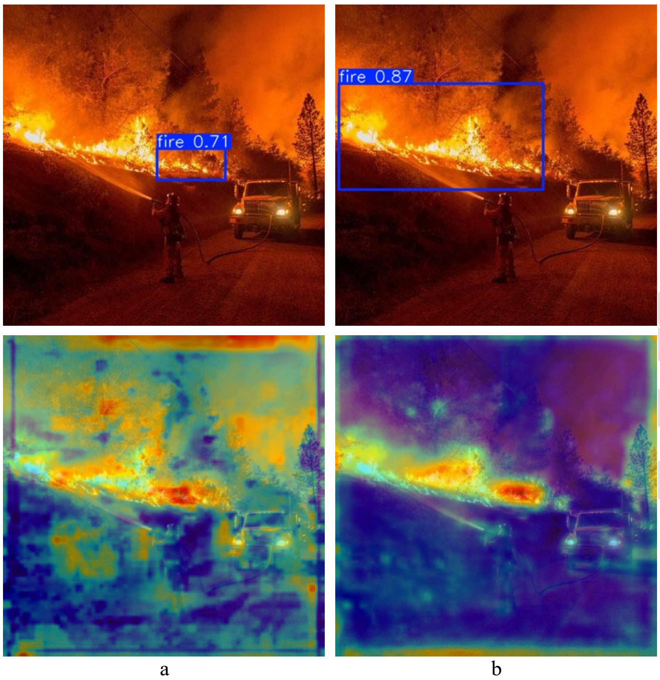


Fig. 14. Comparison of global information extraction capabilities. (The figure below shows the corresponding heatmap for the figure above.) (a) YOLOv8 only extracts locally. (b) YOLO-MP model extraction is more comprehensive.

Table 7
Summary of Model Performance Metrics

Performance Metric	Mean	Standard Deviation	95% Confidence Interval (95% CI)
Precision	0.8746	0.0028	(0.8711, 0.8780)
Recall	0.7981	0.0080	(0.7882, 0.8080)
mAP50	0.8682	0.0024	(0.8652, 0.8712)
mAP50-95	0.5448	0.0045	(0.5392, 0.5504)

The statistical analysis results indicate that the proposed model exhibits strong stability in key performance metrics: the mean value of mAP50 was 0.8682, with a standard deviation of 0.0024, and the 95% confidence interval (calculated using the t-distribution) was (0.8652, 0.8712); the mean value of mAP50-95 was 0.5448, with a standard deviation of 0.0045, and the 95% confidence interval was (0.5392, 0.5504). The narrow fluctuation ranges confirm that the model's performance is not accidental, highlighting its high consistency and reproducibility across multiple experiments. Therefore, the series of experiments provides strong evidence that the proposed model not only achieves superior detection accuracy but also demonstrates excellent performance stability.

3.6. Generalization experiment

To further evaluate the generalization ability of our model, we conducted experiments on the Fire & Smoke Dataset. The dataset encompasses urban, forest, industrial, and indoor scenarios. It contains more than 17,000 images with high-quality annotations, derived from real incidents, surveillance footage, and other sources, comprehensively covering the various appearances of fires under different conditions. During the experiments, we maintained the same hyperparameters and training procedures as described in Section 3.2 to ensure fairness. The results are presented in Table 8.

In the generalization experiments, YOLO-MP demonstrated an excellent balance across key metrics. Although its precision is slightly lower than that of YOLOv8n, its recall is significantly higher. The mAP50 is marginally

Table 8
Generalization experiment results.

Different models	P	R	mAP50	mAP50-95	Params/M	FLOPs/G	Latency/ms
YOLOv8n	0.5271	0.6256	0.5808	0.3103	3.01	8.09	4.74
YOLOv10n	0.5347	0.6162	0.5775	0.3102	2.26	6.52	6.25
YOLOv11n	0.5271	0.6392	0.5792	0.3111	2.58	6.31	6.22
YOLOv12n	0.4837	0.5287	0.4816	0.2448	2.51	5.82	12.39
YOLOv13n	0.4866	0.5281	0.4856	0.2494	2.45	6.20	18.49
RT-DETR	0.4814	0.5614	0.5035	0.2711	18.83	29.68	28.80
YOLO-MP	0.5128	0.6556	0.5816	0.3106	2.07	6.02	5.48

better than that of other models, and mAP50-95 shows a slight improvement, indicating the model's stability and generalization capability across different IoU thresholds. More importantly, YOLO-MP requires far fewer parameters and computational resources than other detection models, highlighting its outstanding lightweight characteristics. Overall, these experimental results indicate that YOLO-MP can achieve high detection precision while improving recall and maintaining computational efficiency, making it a well-balanced and efficient model for generalized scenarios.

4. Discussion

The frequent occurrence of forest fires worldwide has become a critical environmental issue, with resulting economic and societal losses continually increasing. The reduction in forest area and the destruction of valuable natural resources are deeply concerning (Xu et al., 2021; Fan et al., 2024). During forest fire detection, the morphological and dynamic characteristics of flames exhibit greater complexity and variability compared to other objects. Additionally, forest fire detection further exacerbates the challenge of fire detection (Emmy Prema et al., 2018; Hossain et al., 2020). Nevertheless, accurate and efficient forest fire detection is of significant practical value for reducing socio-economic losses, protecting forest areas, and preserving valuable natural resources. Therefore, it has become an urgent problem that requires prompt and effective solutions.

In this study, the proposed model achieves notable improvements in both lightweight design and forest fire detection accuracy through several innovative modifications. These include redesigning the model architecture, introducing novel feature extraction techniques, and optimizing the loss function. These improvements represent not only technical advancements but also a meaningful integration of cutting-edge algorithms with the goal of environmental protection. By enhancing detection performance, the YOLO-MP algorithm enables earlier identification of potential forest fire hazards, thus contributing to reducing disaster-related losses and preserving ecological balance. Compared to traditional detection methods, this approach—driven by big data and advanced algorithms—offers greater efficiency, reduced operational costs, and more comprehensive and objective monitoring, thereby helping to minimize human-induced errors in fire management (Carta et al., 2023). YOLO-MP is a lightweight, automated forest fire detection model that can be deployed on various edge devices to operate in complex field environments without human intervention. This capability overcomes challenges associated with operating under harsh weather conditions (Nazir and Kaleem, 2024). The model maintains robust detection performance in scenarios involving visually similar targets, complex backgrounds, occlusions, and overlapping objects. This advancement not only advances detection technologies but also provides new momentum for forest ecosystem conservation efforts. It fosters broader interdisciplinary collaboration and exploration.

In addition, the model can be integrated with related algorithms to establish a full-cycle forest fire monitoring and management platform (Chandra et al., 2022), enabling closed-loop management from prevention to post-fire recovery. During the fire prevention stage, the system can perform fire susceptibility analysis across forest regions by incorporating terrain, climate, vegetation, and other multi-dimensional data (Kantarcioglu et al., 2023; Pham et al., 2024; Sivrikaya et al., 2024). This allows for precise identification of high-risk areas. The YOLO-MP model can then be applied to these zones for focused monitoring. When a fire occurs, the YOLO-MP model facilitates timely early warning and uploads visual and other relevant fire-related information to the platform. This provides decision-making support for regulatory authorities and, in combination with fire spread prediction algorithms (Zhou et al., 2025), aids firefighting teams in formulating scientifically informed suppression strategies, thereby enhancing emergency response efficiency. In terms of decision support, the management platform can convert information from YOLO-MP, such as confidence levels, into a three-tier risk assessment to guide decision-making. For instance, when the model detects a potential fire with a confidence score <0.5, it is classified as low risk, triggering an automatic alert for manual verification to ensure no potential threats are overlooked. If the model detects a confidence score between 0.5 and 0.85 and the fire scale index is moderate, it is classified as medium risk. In this case, the system recommends that nearby firefighting resources remain on standby while continuously monitoring the situation through real-time data updates. When a confidence score >0.85 is detected along with high fire intensity levels and a rapidly increasing fire scale index, it is classified as high risk. The system then automatically initiates emergency responses, including mobilizing resources, planning evacuation for communities near the forest, and coordinating rescue operations with regional fire departments. After the fire is extinguished, relevant algorithms can be employed to assess the burned area (Afira and Wijayanto, 2022), quantify vegetation loss, and evaluate ecological damage (Novo et al., 2024). These outputs provide data support for post-fire impact assessment. Furthermore, the platform can be linked with ecological restoration models to generate targeted vegetation reconstruction and soil remediation plans based on loss assessment results (Jiang et al., 2024; Zahura et al., 2024; Fernández-Guisuraga et al., 2024). This offers comprehensive technical support for post-disaster ecological recovery. By enabling intelligent and refined forest fire management through the integration of multiple technologies, this full-cycle model enhances proactive fire prevention and improves the scientific basis for post-fire restoration.

5. Limitations and Future Work

Although the proposed model achieves higher detection accuracy and fewer parameters in forest fire detection, certain limitations remain. For example, in complex scenarios involving dense smoke or multilayer vegetation occlusion, false positives and false negatives may still occur. In addition, the dataset lacks sufficient samples, particularly images of forest fires under extreme weather conditions or complex terrains. This may reduce the model's generalization ability in such scenarios. Training deep learning-based models requires a long time, and a small dataset may lead to overfitting. Moreover, the model relies on RGB image inputs, which are influenced by weather, environmental factors, and the image acquisition devices.

In future work, we will further optimize the network design to enhance the robustness of YOLO-MP and deploy it in real-world environments to evaluate its performance. We also intend to expand the dataset by collecting multimodal data, such as images captured by infrared cameras, which are less affected by lighting and weather conditions. For complex terrain scenarios, drones equipped with depth cameras will be deployed to acquire three-dimensional information of fire sources, providing the model with richer feature representations. Furthermore, we will consider redesigning the network to handle data loss, noise interference (Yang et al., 2025a), and the characteristics of multimodal datasets. These efforts aim further to improve the performance of the forest fire detection model, expand its application scenarios, and enhance its practical utility, thereby contributing more intelligence and capability to the effective prevention and control of global forest fires.

6. Conclusions

This study proposes a novel lightweight forest fire detection model, YOLO-MP, based on an enhanced YOLOv8 architecture. The model incorporates a new lightweight backbone network named GHGNet, which simultaneously reduces the parameter count and computational cost while improving feature extraction capability. Additionally, we apply FPSConv to enlarge the receptive field and strengthen multi-scale feature aggregation. Furthermore, GERB is introduced to optimize gradient flow. Finally, a new loss function, W-EIoU, is designed to improve localization performance.

The YOLO-MP model accurately identifies flame targets in forest fires, effectively reducing both false positive and false negative rates. Its lightweight design significantly reduces the number of parameters and the computational complexity. Consequently, the model is particularly well-suited for deployment and operation on resource-constrained devices, such as embedded systems and mobile terminals, thereby expanding its application scenarios and practical utility while providing robust technical support for real-world forest fire monitoring. By enabling more accurate and efficient forest fire detection, YOLO-MP provides a practical technical tool for ecological monitoring and forest ecosystem protection. This highlights the importance of combining algorithmic innovation with ecological applications and contributes to interdisciplinary efforts addressing global environmental challenges.

CRediT authorship contribution statement

Hongwei Zhu: Writing – original draft, Visualization, Software, Conceptualization. **Weiwei Ling:** Formal analysis, Conceptualization. **Huabiao Yan:** Visualization, Validation, Conceptualization. **Xinghai Zhong:** Writing – review & editing, Validation, Supervision, Software. **Feng Liao:** Writing – review & editing, Visualization, Validation.

Declaration of competing interest

The authors declare no conflicts of interest.

Data availability

The code, data and model used to obtain the presented results can be found here: https://github.com/Braised-Peppa-Pig/YOLO-MP

Acknowledgments

This work was supported in part by the Special Project for High-Skilled and High-Level Talents of Jiangxi College of Applied Technology (Grant No. JXYY-G2023001).

References

Abramov, N., Emelyanova, Y., Fralenko, V., Khachumov, V., Khachumov, M., Shustova, M., Talalaev, A., 2024. Intelligent methods for forest fire detection using unmanned aerial vehicles. Fire 7, 89. https://doi.org/10.3390/fire7030089.

Afira, N., Wijayanto, A.W., 2022. Mono-temporal and multi-temporal approaches for burnt area detection using sentinel-2 satellite imagery (a case study of rokan hilir regency, indonesia). Ecological Informatics 69, 101677. https://doi.org/10.1016/j.ecoinf.2022.101677.

Ahajjam, A., Allgaier, M., Chance, R., Chukwuemeka, E., Putkonen, J., Pasch, T., 2025. Enhancing prediction of wildfire occurrence and behavior in alaska using spatio-temporal clustering and ensemble machine learning. Ecological Informatics 85, 102963. https://doi.org/10.1016/j.ecoinf.2024.102963.

Aslan, Y.E., Korpeoglu, I., Özgür Ulusoy, 2012. A framework for use of wireless sensor networks in forest fire detection and monitoring. Computers, Environment and Urban Systems 36, 614–625. https://doi.org/10.1016/j.compenvurbsys.2012.03.002.

Baek, J., Alhindi, T.J., Jeong, Y.S., Jeong, M.K., Seo, S., Kang, J., Shim, W., Heo, Y., 2023. A wavelet-based real-time fire detection algorithm with multi-modeling framework. Expert Systems with Applications 233, 120940. https://doi.org/10.1016/j.eswa.2023.120940.

Bhadoria, R.S., Pandey, M.K., Kundu, P., 2021. Rvfr: Random vector forest regression model for integrated & enhanced approach in forest fires predictions. Ecological Informatics 66, 101471. https://doi.org/10.1016/j.ecoinf.2021.101471.

Borges, P.V.K., Izquierdo, E., 2010. A probabilistic approach for vision-based fire detection in videos. IEEE Transactions on Circuits and Systems for Video Technology 20, 721–731. https://doi.org/10.1109/TCSVT.2010.2045813.

Carta, F., Zidda, C., Putzu, M., Loru, D., Anedda, M., Giusto, D., 2023. Advancements in forest fire prevention: A comprehensive survey. Sensors 23, 6635. https://www.mdpi.com/1424-8220/23/14/6635.

Çelik, T., Demirel, H., 2009. Fire detection in video sequences using a generic color model. Fire Safety Journal 44, 147–158. https://doi.org/10.1016/j.firesaf.2008.05.005.

Chandra, R., Agarwal, S., Singh, N., 2022. Semantic sensor network ontology based decision support system for forest fire management. Ecological Informatics 72, 101821. https://doi.org/10.1016/j.ecoinf.2022.101821.

Chen, J., Kao, S.h., He, H., Zhuo, W., Wen, S., Lee, C.H., Chan, S.H.G., 2023. Run, don't walk: Chasing higher flops for faster neural networks, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12021–12031. https://doi.org/10.1109/CVPR52729.2023.01157.

Coen, J.L., Schroeder, W., 2013. Use of spatially refined satellite remote sensing fire detection data to initialize and evaluate coupled weather-wildfire growth model simulations. Geophysical Research Letters 40, 5536–5541. https://doi.org/10.1002/2013GL057868.

Ding, X., Zhang, X., Han, J., Ding, G., 2022. Scaling up your kernels to 31×31: Revisiting large kernel design in cnns, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11953–11965. https://doi.org/10.1109/CVPR52688.2022.01166.

Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., Sun, J., 2021. Repvgg: Making vgg-style convnets great again, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13728–13737. https://doi.org/10.1109/CVPR46437.2021.01352.

El-Madafri, I., Peña, M., Olmedo-Torre, N., 2023. The wildfire dataset: Enhancing deep learning-based forest fire detection with a diverse evolving open-source dataset focused on data representativeness and a novel multi-task learning approach. Forests 14, 1697. https://doi.org/10.3390/f14091697.

Emmy Prema, C., Vinsley, S.S., Suresh, S., 2018. Efficient flame detection based on static and dynamic texture analysis in forest fire detection. Fire Technology 54, 255–288. https://doi.org/10.1007/s10694-017-0683-x.

Fan, D., Wang, M., Liang, T., He, H., Zeng, Y., Fu, B., 2024. Estimation and trend analysis of carbon emissions from forest fires in mainland china from 2011 to 2021. Ecological Informatics 81, 102572. https://doi.org/10.1016/j.ecoinf.2024.102572.

Feng, Y., Huang, J., Du, S., Ying, S., Yong, J.H., Li, Y., Ding, G., Ji, R., Gao, Y., 2025. Hyper-yolo: When visual object detection meets hypergraph computation. IEEE Transactions on Pattern Analysis and Machine Intelligence 47, 2388–2401. https://doi.org/10.1109/TPAMI.2024.3524377.

Fernández-Guisuraga, J.M., Fernández-Manso, A., Quintano, C., Fernández-García, V., Cerrillo, A., Marqués, G., Cascallana, G., Calvo, L., 2024. Firemap: Cloud-based software to automate the estimation of wildfire-induced ecological impacts and recovery processes using remote sensing techniques. Ecological Informatics 81, 102591. https://doi.org/10.1016/j.ecoinf.2024.102591.

Gevorgyan, Z., 2022. Siou loss: More powerful learning for bounding box regression. arXiv preprint. arXiv:2205.12740.

Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C., 2020. Ghostnet: More features from cheap operations, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1577–1586. https://doi.org/10.1109/CVPR42600.2020.00165.

Hossain, F.A., Zhang, Y.M., 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. Journal of Unmanned Vehicle Systems 8, 285–309. https://doi.org/10.1139/juvs-2020-0009.

Jiang, B., Chen, W., Zou, Y., Wu, C., Wu, Z., Kang, X., Xiao, H., Sakai, T., 2024. Post-fire vegetation dynamic patterns and drivers in greater hinggan mountains: Insights from long-term remote sensing data analysis. Ecological Informatics 83, 102850. https://doi.org/10.1016/j.ecoinf.2024.102850.

Jocher, G., Chaurasia, A., Qiu, J., 2023a. YOLO by Ultralytics. https://github.com/ultralytics/ultralytics.

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., Michael, K., TaoXie, Fang, J., imyhxy, Lorna, Zeng, Y., Wong, C., V, A., Montes, D., Wang, Z., Fati, C., Nadar, J., Laughing, UnglvKitDe, Sonck, V., tkianai, yxNONG, Skalski, P., Hogan, A., Nair, D., Strobel, M., Jain, M., 2022. ultralytics/yolov5: v7.0 - yolov5 sota realtime instance segmentation. https://doi.org/10.5281/zenodo.7347926.

Jocher, G., Qiu, J., Chaurasia, A., 2023b. Ultralytics YOLO. https://github.com/ultralytics/ultralytics.

Kantarcioglu, O., Kocaman, S., Schindler, K., 2023. Artificial neural networks for assessing forest fire susceptibility in türkiye. Ecological Informatics 75, 102034. https://doi.org/10.1016/j.ecoinf.2023.102034.

Ko, B.C., Cheong, K.H., Nam, J.Y., 2009. Fire detection based on vision sensor and support vector machines. Fire Safety Journal 44, 322–329. https://doi.org/10.1016/j.firesaf.2008.07.006.

Lei, M., Li, S., Wu, Y., Hu, H., Zhou, Y., Zheng, X., Ding, G., Du, S., Wu, Z., Gao, Y., 2025. Yolov13: Real-time object detection with hypergraph-enhanced adaptive visual perception. arXiv preprint. arXiv:2506.17733.

Li, J., Zhou, G., Chen, A., Wang, Y., Jiang, J., Hu, Y., Lu, C., 2022. Adaptive linear feature-reuse network for rapid forest fire smoke detection model. Ecological Informatics 68, 101584. https://doi.org/10.1016/j.ecoinf.2022.101584.

Lin, J., Lin, H., Wang, F., 2023. A semi-supervised method for real-time forest fire detection algorithm based on adaptively spatial feature fusion. Forests 14, 361. https://doi.org/10.3390/f14020361.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. Ssd: Single shot multibox detector, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham. pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.

Liu, X., Peng, H., Zheng, N., Yang, Y., Hu, H., Yuan, Y., 2023. Efficientvit: Memory efficient vision transformer with cascaded group attention, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14420–14430. https://doi.org/10.1109/CVPR52729.2023.01386.

Ma, N., Zhang, X., Zheng, H.T., Sun, J., 2018. Shufflenet v2: Practical guidelines for efficient cnn architecture design, in: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer Vision – ECCV 2018, Springer International Publishing, Cham. pp. 122–138. https://doi.org/10.1007/978-3-030-01264-9_8.

Ma, X., Dai, X., Bai, Y., Wang, Y., Fu, Y., 2024. Rewrite the stars, in: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5694–5703. https://doi.org/10.1109/CVPR52733.2024.00544.

Nazir, S., Kaleem, M., 2024. Object classification and visualization with edge artificial intelligence for a customized camera trap platform. Ecological Informatics 79, 102453. https://doi.org/10.1016/j.ecoinf.2023.102453.

Novo, A., Fernández, C., Míguez, C., Suárez-Vidal, E., 2024. Analysing the capacity of multispectral indices to map the spatial distribution of potential post-fire soil losses based on soil burn severity. Ecological Informatics 83, 102793. https://doi.org/10.1016/j.ecoinf.2024.102793.

Pham, V.T., Do, T.A.T., Tran, H.D., Do, A.N.T., 2024. Classifying forest cover and mapping forest fire susceptibility in dak nong province, vietnam utilizing remote sensing and machine learning. Ecological Informatics 79, 102392. https://doi.org/10.1016/j.ecoinf.2023.102392.

Qin, D., Leichner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., Wang, W., Banbury, C., Ye, C., Akin, B., Aggarwal, V., Zhu, T., Moro, D., Howard, A., 2025. Mobilenetv4: Universal models for the mobile ecosystem, in: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (Eds.), Computer Vision – ECCV 2024, Springer Nature Switzerland, Cham. pp. 78–96. https://doi.org/10.1007/978-3-031-73661-2_5.

Redmon, J., Farhadi, A., 2018. Yolov3: An incremental improvement. arXiv preprint. arXiv:1804.02767.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S., 2019. Generalized intersection over union: A metric and a loss for bounding box regression, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 658–666. https://doi.org/10.1109/CVPR.2019.00075.

Rodriguez-Jimenez, F., Lorenzo, H., Acuña-Alonso, C., Alvarez, X., 2023. Pls-pm analysis of forest fires using remote sensing tools. the case of xurés in the transboundary biosphere reserve. Ecological Informatics 75, 102010. https://doi.org/10.1016/j.ecoinf.2023.102010.

Sarkar, M.S., Majhi, B.K., Pathak, B., Biswas, T., Mahapatra, S., Kumar, D., Bhatt, I.D., Kuniyal, J.C., Nautiyal, S., 2024. Ensembling machine learning models to identify forest fire-susceptible zones in northeast india. Ecological Informatics 81, 102598. https://doi.org/10.1016/j.ecoinf.2024.102598.

Sathishkumar, V.E., Cho, J., Subramanian, M., Naren, O.S., 2023. Forest fire and smoke detection using deep learning-based learning without forgetting. Fire Ecology 19, 9. https://doi.org/10.1186/s42408-022-00165-0.

Shamsoshoara, A., Afghah, F., Razi, A., Zheng, L., Fulé, P.Z., Blasch, E., 2021. Aerial imagery pile burn detection using deep learning: The flame dataset. Computer Networks 193, 108001. https://doi.org/10.1016/j.comnet.2021.108001.

Sivrikaya, F., Günlü, A., Ömer Küçük, Ürker, O., 2024. Forest fire risk mapping with landsat 8 oli images: Evaluation of the potential use of vegetation indices. Ecological Informatics 79, 102461. https://doi.org/10.1016/j.ecoinf.2024.102461.

Strubell, E., Verga, P., Belanger, D., McCallum, A., 2017. Fast and accurate entity recognition with iterated dilated convolutions, in: Palmer, M., Hwa, R., Riedel, S. (Eds.), Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Copenhagen, Denmark. pp. 2670–2680. https://doi.org/10.18653/v1/D17-1283.

Tian, Y., Ye, Q., Doermann, D., 2025. Yolov12: Attention-centric real-time object detectors. arXiv preprint. arXiv:2502.12524.

Tong, Z., Chen, Y., Xu, Z., Yu, R., 2023. Wise-iou: Bounding box regression loss with dynamic focusing mechanism. arXiv preprint. arXiv:2301.10051.

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., Ding, G., 2024. Yolov10: Real-time end-to-end object detection, in: Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., Zhang, C. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc.. pp. 107984–108011.

Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M., 2023. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721.

Wang, C.Y., Mark Liao, H.Y., Wu, Y.H., Chen, P.Y., Hsieh, J.W., Yeh, I.H., 2020. Cspnet: A new backbone that can enhance learning capability of cnn, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1571–1580. https://doi.org/10.1109/CVPRW50498.2020.00203.

Wang, C.Y., Yeh, I.H., Mark Liao, H.Y., 2025. Yolov9: Learning what you want to learn using programmable gradient information, in: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (Eds.), Computer Vision – ECCV 2024, Springer Nature Switzerland, Cham. pp. 1–21. https://doi.org/10.1007/978-3-031-72751-1_1.

Xu, X., Li, F., Lin, Z., Song, X., 2021. Holocene fire history in china: Responses to climate change and human activities. Science of The Total Environment 753, 142019. https://doi.org/10.1016/j.scitotenv.2020.142019.

Yang, X., Hua, Z., Zhang, L., Fan, X., Zhang, F., Ye, Q., Fu, L., 2023. Preferred vector machine for forest fire detection. Pattern Recognition 143, 109722. https://doi.org/10.1016/j.patcog.2023.109722.

Yang, Y., Li, S., Wu, P., Moshayedi, A.J., Feng, W., 2025a. Design and analysis of a novel zeroing neural network for matrix decomposition under data loss and noise interference. Applied Mathematical Modelling 147, 116201. https://doi.org/10.1016/j.apm.2025.116201.

Yang, Y., Wu, P., Katsikis, V.N., Li, S., Feng, W., 2025b. A novel real-time noise-resilient zeroing neural network and its applications to matrix problem solving. Mathematics and Computers in Simulation https://doi.org/10.1016/j.matcom.2025.01.006.

Yang, Z., Shao, Y., Wei, Y., Li, J., 2024. Precision-boosted forest fire target detection via enhanced yolov8 model. Applied Sciences 14, 2413. https://doi.org/10.3390/app14062413.

Zahura, F.T., Bisht, G., Li, Z., McKnight, S., Chen, X., 2024. Impact of topography and climate on post-fire vegetation recovery across different burn severity and land cover types through random forest. Ecological Informatics 82, 102757. https://doi.org/10.1016/j.ecoinf.2024.102757.

Zhang, L., Wang, M., Fu, Y., Ding, Y., 2022a. A forest fire recognition method using uav images based on transfer learning. Forests 13, 975. https://doi.org/10.3390/f13070975.

Zhang, Y.F., Ren, W., Zhang, Z., Jia, Z., Wang, L., Tan, T., 2022b. Focal and efficient iou loss for accurate bounding box regression. Neurocomputing 506, 146–157. https://doi.org/10.1016/j.neucom.2022.07.042.

Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., Chen, J., 2024. Detrs beat yolos on real-time object detection, in: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16965–16974. https://doi.org/10.1109/CVPR52733.2024.01605.

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D., 2020. Distance-iou loss: Faster and better learning for bounding box regression. Proceedings of the AAAI Conference on Artificial Intelligence 34, 12993–13000. https://doi.org/10.1609/aaai.v34i07.6999.

Zhou, Z., Xu, Q., Yi, J., Li, Y., Zhang, S., Li, W., 2025. Modeling the spread of forest fires through cellular automata by leveraging deep learning to derive transition rules. Ecological Informatics 88, 103150. https://doi.org/10.1016/j.ecoinf.2025.103150.

Highlights

- Employ a lightweight backbone network GHGNet to enhance feature extraction capability while reducing computational complexity and model parameters.
- Utilize FPSConv to achieve multi-scale feature extraction and expand the receptive field.
- Introduce the GERB module in the neck to enhance gradient flow and feature fusion.
- Improve target localization accuracy through the use of W-EIoU loss.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: