

Enhanced fire detection in new dataset: a fine-grained and multiscale-optimized architecture

Zhengjie Wang^a and Zengmin Xu^{b,c,d,*}

^aTaiyuan University of Technology, College of Artificial Intelligence, JinZhong, China

^bGuilin University of Electronic Technology, School of Mathematics and Computing Science, Guangxi Colleges
and Universities Key Laboratory of Data Analysis and Computation, Guilin, China

^cCenter for Applied Mathematics of Guangxi (GUET), Guilin, China

^dAnview.ai, Guilin, China

ABSTRACT. Most existing deep learning methods are confined to single-scenario detection or coarse-grained binary classification tasks (fire/nonfire), often overlooking the critical value of fine-grained fire type classification in smart city development. However, when confronting flame targets with significant scale variations, existing models face severe challenges regarding feature extraction and multiscale adaptability, impeding precise detection. Addressing these issues, we propose EFDC-YOLO, an improved framework built upon YOLOv8. Specifically, we integrate a dynamic ghost convolution module with SS2D into the backbone network and embed a strip large kernel spatial module into the neck network. By effectively capturing discriminative spatial context information and enhancing multiscale feature representation capabilities, this architecture significantly elevates the model's recognition precision and robustness. In addition, we release a diverse dataset covering building, vehicle, and forest fires. Experimental results demonstrate that EFDC-YOLO outperforms the original YOLOv8, achieving mAP_{0.5} improvements of 3.2% and 1.7% on the public M⁴SFWD dataset and our self-constructed dataset, respectively. These results robustly demonstrate the immense potential of EFDC-YOLO for achieving high-precision, multitype fire monitoring in complex real-world scenarios. Our codes and dataset are publicly available at <https://github.com/Administor123/EFDC-YOLO/>

© 2026 SPIE and IS&T [DOI: [10.1117/1.JEI.35.4.041406](https://doi.org/10.1117/1.JEI.35.4.041406)]

Keywords: fire detection; dynamic conv; ghost conv; multiscale

Paper 251887SS received Dec. 30, 2025; revised Feb. 2, 2026; accepted Feb. 4, 2026; published Feb. 23, 2026.

1 Introduction

Fire is widely recognized as a significant disaster that disrupts daily life, causes extensive economic losses, jeopardizes safety, damages property and wildlife habitats, and can lead to environmental pollution.^{1,2} In 2023, adverse weather conditions in Canada contributed to widespread forest fires, releasing considerable combustion products into the atmosphere and resulting in environmental pollution.³ According to the U.S. Fire Administration (USFA), in 2020, the National Fire Protection Association reported 1.4 million fire incidents in the United States, resulting in 3500 deaths, 15,200 injuries, and approximately \$21.9 billion in economic losses.⁴ In China, data from the National Fire Rescue Administration indicate that from January to August 2024, there were 660,000 fire rescue reports, leading to 1324 deaths, 1760 injuries, and

*Address all correspondence to Zengmin Xu, xzm@guet.edu.cn

direct property losses totaling 4.92 billion yuan. To mitigate the losses caused by fires, early detection is crucial. Current fire detection methods predominantly rely on scalar and visual sensors. Traditional systems largely utilize scalar sensors, including flame, gas, and temperature detectors.^{5,6} Although these sensors are cost-effective and easy to install, they are primarily suitable for indoor environments and offer limited coverage. In addition, because these sensors are typically mounted on ceilings, the time required for flames and smoke to reach them can result in delays, reducing their efficacy as early warning systems. Recently, an increasing number of researchers have begun to apply computer vision technologies to fire detection.^{7,8} Compared with traditional physical sensors, vision-based fire detection solutions provide advantages such as wider coverage, lower deployment costs, quicker response times, and greater robustness. Despite significant advancements in deep-learning-based fire detection research, existing supervised learning algorithms remain constrained by several key bottlenecks hindering their practical deployment.

First, existing detection paradigms often oversimplify the task, lacking in-depth exploration of multitask joint representation. Many methods predominantly treat fire recognition as a coarse-grained classification problem (fire/nonfire)⁹ or limit detection to single scenarios, making it difficult to achieve joint modeling of “object localization + type discrimination.” In real-world emergency response, scenarios such as building fires, vehicle fires, and forest fires often exhibit significantly different background structures, combustion morphologies, and interference factors. The lack of reliable fire type information weakens the ability of management personnel to rapidly assess the nature of the event, thereby affecting decision-making regarding response strategies, cross-departmental collaboration, and resource allocation efficiency.

Second, balancing multiscale variations with computational efficiency remains a critical challenge affecting detection robustness. Although existing studies^{10–12} have attempted to address multiscale issues in fire detection by introducing complex mechanisms, these methods are typically accompanied by high computational costs or suffer from insufficient multiscale adaptability, often limiting their applicability to relatively simple scenarios. Therefore, designing a mechanism capable of adaptively adjusting the receptive field to match flames of varying scales, without compromising real-time performance, remains an unexplored gap in current research.

Furthermore, the scarcity of high-quality benchmark data limits the generalization capability and fair evaluation of algorithms. Yar et al. pointed out that annotated fire detection datasets are often difficult to access or remain unpublished;⁹ meanwhile, a survey by Han et al. indicates that existing datasets suffer from issues such as low resolution and limited scenario coverage.¹³ These factors hinder the fair evaluation of different methods under a unified benchmark and restrict the generalization validation and deployment of models in real-world complex environments.

Recent research in the field of computer vision offers new insights for addressing these problems. On the one hand, to more effectively capture the fine-grained spatial structure of objects, VMamba introduced the 2D selective scan (SS2D) mechanism,¹⁴ demonstrating superior capability in extracting global spatial context information. This is crucial for distinguishing between fire types that share similar textures but exhibit distinct spatial distributions. On the other hand, considering that flame morphologies often exhibit irregular and anisotropic characteristics, standard square convolution kernels struggle to align perfectly with their geometric shapes. The concept of strip pooling and strip convolution proposed by Hou et al.¹⁵ can effectively preserve long-range dependencies along specific directions while maintaining high computational efficiency. Furthermore, when combined with convolution designs using varying dilation rates, it becomes possible to flexibly adjust the effective receptive field to cover multiscale targets.

Motivated by these advanced concepts and aiming at the dynamic variability of fire morphologies and background complexity, this study proposes a general fire detection model based on an improved YOLOv8: Efficient Fire Detect and Classify model based on YOLO (EFDC-YOLO). This model not only accurately identifies fires and classifies them into specific types (e.g., building fire, vehicle fire, and forest fire) but also precisely localizes fire areas using bounding boxes. We do not aim to merely increase parameter counts to compete with large-scale server-side models in terms of precision; rather, we seek to break through the bottlenecks of existing methods in multitype and multiscale fire recognition while maintaining efficient inference.

The main contributions of this paper are summarized as follows:

- This is the first attempt to simultaneously address dynamic texture feature extraction and multitype fire classification based on the YOLOv8 framework. We integrate a self-designed dynamic ghost conv with the SS2D (DGSS) module into the backbone network. By modeling the fine-grained features of flames, this effectively resolves the inter-class confusion issues faced by binary classification models.
- To mitigate the limitations of fixed receptive fields in multiscale fire detection, we design the strip large kernel spatial module (SLSM). This module integrates strip depthwise separable convolutions (SDWConv) with a multiscale selection pooling mechanism to replace the SPPF module in the YOLOv8 neck. This design effectively balances local detail capture with global context perception.
- Beyond module innovations, we construct a multitype fire dataset to bridge the current domain gap, available for the research community to promote the development of smart cities.

2 Related Work

2.1 Deep Learning in Fire Detection

Deep learning techniques have extensively permeated various sub-tasks, including fire detection, flame segmentation, and image classification; this paper primarily focuses on the domain of object detection. Since 2017, a multitude of innovative paradigms based on deep learning have emerged in the academic community,¹⁶ aiming to continuously push the boundaries of performance.

In the pursuit of real-time performance and lightweight deployment, researchers have made substantial efforts. Jadon et al.¹⁷ developed FireNet, a lightweight network designed specifically for low-computing-power devices, achieving real-time monitoring. Li and Zhao¹⁸ enhanced detection capabilities while maintaining response speed by improving YOLOv3. Targeting embedded applications, Li et al.¹⁹ combined YOLOX with image segmentation to construct a model suitable for UAV edge computing; Huang et al.²⁰ proposed the ultra-lightweight YOLO-ULNet, successfully realizing real-time forest fire detection on resource-constrained sensing devices.

To further elevate detection precision in complex scenarios, various advanced feature fusion strategies have been introduced. Barmpoutis et al.²¹ innovatively combined Faster R-CNN with multidimensional texture analysis based on higher-order linear dynamic systems, significantly reducing false alarm rates. EFDNet, developed by Li et al.,²² utilizes multiscale features and channel attention to bolster the capability of capturing minute flames. Huang et al.²³ employed RCNN to extract deep features combined with ResNet to predict fire trends. Furthermore, the VFD network proposed by Wang et al.²⁴ achieved progress in balancing speed and accuracy in video stream detection.

2.2 Scenario-Specific Fire Detection

Adaptive optimization for specific application scenarios is also a research hotspot. Xu et al.²⁵ integrated YOLOv5 with EfficientDet to adapt to variable forest environments; h-EfficientDet, proposed by Li et al.,²⁶ demonstrated high accuracy in early forest fire detection. FCLGYOLO, designed by Ren et al.,²⁷ was specifically optimized for forest fires from UAV perspectives. In indoor scenarios, the model developed by Ahn et al.²⁸ has been proven effective in reducing damage caused by indoor fires.

However, most of the aforementioned methods suffer from significant limitations: they are typically optimized for single scenarios (exclusively for forest or indoor environments). This limitation is particularly pronounced within the comprehensive management framework of modern smart cities. In the smart city ecosystem, building, vehicle, and forest fires often fall under the jurisdiction of different functional departments (e.g., fire departments, traffic management bureaus, forestry bureaus). The lack of a unified capability to identify and distinguish between multiple fire types hinders precise cross-departmental warning distribution and emergency response. To overcome this bottleneck, this paper is dedicated to constructing a unified

detection framework that achieves fine-grained classification of multiscene fires by introducing advanced spatial feature extraction mechanisms.

2.3 Multiscale Flame Feature Modeling

Multiscale modeling is crucial for adapting to smoke and fire targets with vast scale spans in surveillance or UAV imagery. Classical strategies (e.g., FPN and PANet), which aggregate features by constructing fixed-resolution pyramids, have become standard practice. Recent studies^{29–32} have further explored optimizing pyramid structures or introducing complex attention mechanisms, achieving significant results in improving precision.

Nevertheless, these performance gains are often accompanied by increased consumption of computational resources. Excessive parameter counts and complex attention operations may constrain inference speed, posing challenges to real-time fire early warning applications. To seek a better trade-off between feature representation and computational efficiency, this paper proposes a lightweight strip large kernel spatial module (SLSM). This module eschews heavy attention stacking in favor of utilizing strip convolutions and multidilation rate designs to flexibly adjust the effective receptive field. This hybrid architecture aims to significantly bolster the model's representation capability for flames of varying scales with minimal additional computational burden.

3 Methodology

The YOLOv8 framework represents the state-of-the-art in single-stage object detection, featuring a distinct modular architecture composed of a backbone, a neck, and a head, as illustrated in Fig. 1. The backbone, an evolution of the CSPDarkNet53 design, integrates modules such as C2f and SPPF³³ to ensure efficient feature extraction. For multiscale feature aggregation, the neck employs a path aggregation network (PANet)³⁴ structure. Finally, a decoupled detection head is utilized to independently perform predictions for object classification and localization. Despite its exceptional performance, the high computational demand of YOLOv8 presents significant challenges for deployment on resource-constrained platforms. Considering the necessity of deploying models on edge devices, overly complex detection frameworks are prone to reduced efficiency due to computational redundancy. Therefore, this study selects YOLOv8n, the most lightweight variant in the series, as the baseline model for architectural optimization to strike a balance between inference speed and model complexity.

3.1 EFDC-YOLO Network Structure

Modifications to the baseline YOLOv8n are highlighted within red dashed boxes. Key innovations include: (1) integrating the dynamic ghost conv with the SS2D (DGSS) module into the backbone to capture discriminative spatial context and texture features; and (2) replacing the original spatial pyramid pooling-fast (SPPF) with the strip large kernel spatial module (SLSM) in the neck to bolster feature aggregation across varying scales. Detailed descriptions of these modules are provided in Secs. 3.2 and 3.3.

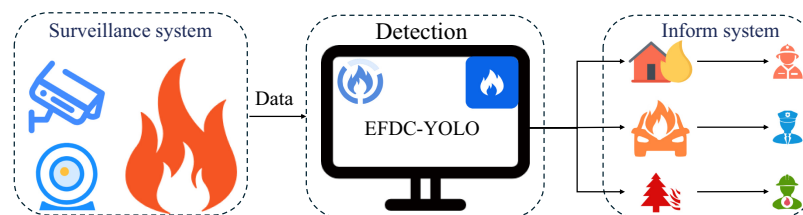


Fig. 1 EFDC-YOLO classifies different fire types and selectively delivers corresponding fire alerts to appropriate administrators to support efficient emergency response. To ensure system reliability, the inform system applies a decision logic that triggers an alarm only when the confidence score exceeds 0.6 for three consecutive frames, thereby reducing false alarms caused by transient noise or momentary disturbances.

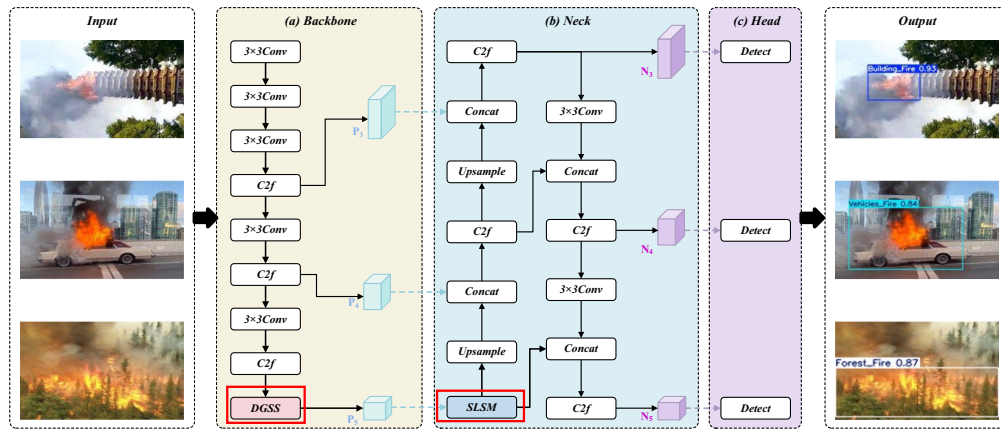


Fig. 2 Integrate DGSS into the backbone to enhance feature extraction and replace the original SPPF module with SLSM.

Building upon YOLOv8n, we developed a model called EFDC-YOLO designed for both fire detection and classification across fire types. The EFDC-YOLO architecture, shown in Fig. 2, incorporates a custom-designed DGSS module and replaces the original SPPF with the SLSM. To facilitate understanding, Secs. 3.2 and 3.3 provide detailed descriptions of these modules.

3.2 Fine-Grained Spatial Feature Enhancement Module DGSS

In fire detection tasks, different fire types exhibit unique characteristics. For instance, building fires are often accompanied by thick smoke and structural obstructions, vehicle fires commonly display localized explosions and irregular flame shapes, and forest fires are frequently obscured by surrounding vegetation. Relying on shallow features such as shape and color alone proves inadequate for accurately distinguishing these fire types. Our analysis indicates that fire types are closely linked to the spatial context in which they occur. Therefore, capturing a combination of flame shape, color, and spatial features during detection enhances the accuracy of fire detection and classification.

Convolutional neural networks (CNNs) are highly effective at extracting detailed flame features, whereas the SS2D module, as used in the VMamba¹⁴ model, can better capture both flame and surrounding spatial information while preserving a broad receptive field. To leverage these advantages, we combined convolution with SS2D in our design, using our proposed dynamic ghost conv (DGConv) to allow the module to capture rich, multilevel features. Meanwhile, the SS2D module aids in capturing the spatial context surrounding the flame. This combined structure forms our newly designed DGSS module. As illustrated in Fig. 3, image information is processed through four parallel branches, each capturing different aspects of the input, before being merged through a concatenation operation into a unified feature output.

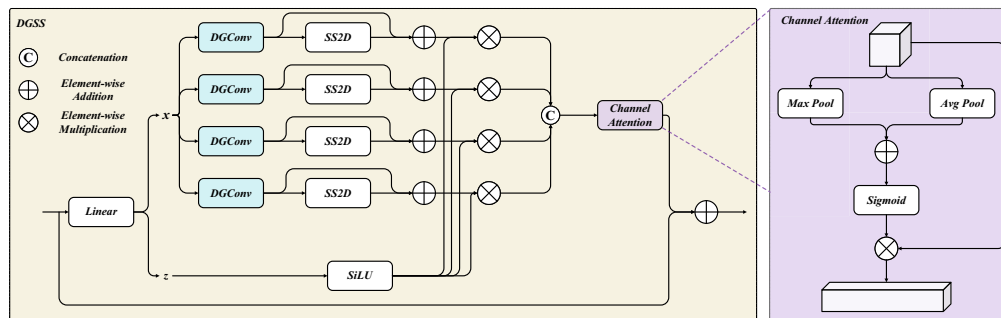


Fig. 3 Schematic of the DGSS architecture: the first half follows the VMamba structure to extract spatial information, with convolutions replaced by the proposed DGConv; the second half incorporates a channel attention module to capture channel-wise information.

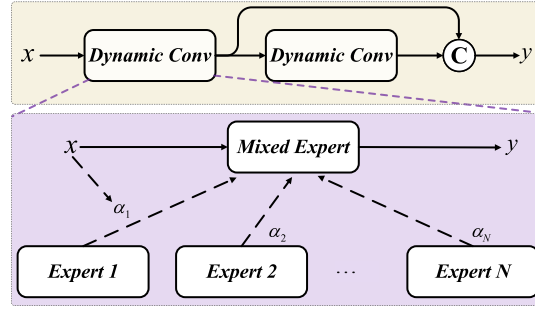


Fig. 4 Diagram of the DGConv architecture: leveraging the lightweight structure of GhostConv, it replaces the standard Conv with Dynamic Conv to achieve more flexible and powerful feature extraction capabilities. The illustration of Dynamic Conv is adapted with permission from Ref. 35.

Considering that variations in fire characteristics manifest not only in spatial distributions but also in distinct channel-specific feature information, we further integrated a channel attention mechanism within the DGSS module. This mechanism adaptively adjusts the weights across channels to emphasize critical feature areas within fire images. By doing so, the model concurrently attends to both the spatial location and the intricate details of the flames, thus enhancing the accuracy of fire type classification.

The core of DGSS is dynamic ghost convolution (DGConv), an innovative convolution mechanism specifically designed to capture key visual features in fire scenes. DGConv integrates the dynamic adjustment strategies of DynamicConv³⁵ with the efficient feature generation method of GhostConv.³⁶ By adaptively modifying the dimensions and structures of convolutional kernels and generating supplementary “ghost” feature maps, DGConv significantly improves the model’s ability to detect flames of diverse scales and shapes. DGConv employs a hybrid strategy, with its implementation strategy and design concept illustrated in Fig. 4.

Assume there are N dynamic experts. For a given input feature map $X \in \mathbb{R}^{C_{in} \times H \times W}$, the output feature map $Y \in \mathbb{R}^{C_{out} \times H' \times W'}$ can be obtained through the weight tensor $W_i \in \mathbb{R}^{C_{out} \times C_{in} \times H \times W}$ of the i ’th convolutional layer and the corresponding dynamic coefficient α_i , as shown in Eq. (1). In Eq. (1), $*$ denotes convolution.

$$Y = X * \sum_{i=1}^N \alpha_i W_i. \quad (1)$$

To produce the dynamic coefficient α_i , we employ a multilayer perceptron (MLP) based on the input. Initially, global average pooling is applied to transform X into a single vector. Then, a two-layer MLP module with softmax activation generates the dynamic coefficient α_i , as shown in Eq. (2).

The method for generating the dynamic coefficient α_i typically involves using an MLP module to generate coefficients based on the input. First, global average pooling is used to fuse the input information x into a vector. Then, a two-layer MLP module with softmax activation is used to dynamically generate the coefficients, as shown in Eq. (2).

$$\alpha_i = \text{softmax}(\text{MLP}(\text{Pool}(X))), \quad (2)$$

where $\alpha_i \in \mathbb{R}^M$. This approach slightly increases FLOPs compared with standard convolution but introduces additional parameters, enhancing the model’s ability to distinguish key features.

Subsequently, the generated feature map Y undergoes identity mapping and dynamic convolution operations, expanding the model’s receptive field to capture flame details more effectively. This combination not only strengthens the model’s feature representation capability but also optimizes computational efficiency, making it well suited for nuanced fire detection tasks.

3.3 Efficient Multiscale Flame Feature Modeling Module SLSM

In fire detection tasks, models frequently encounter difficulties in processing diverse fire scenes, managing occlusions, and distinguishing flames from similar background disturbances. The use

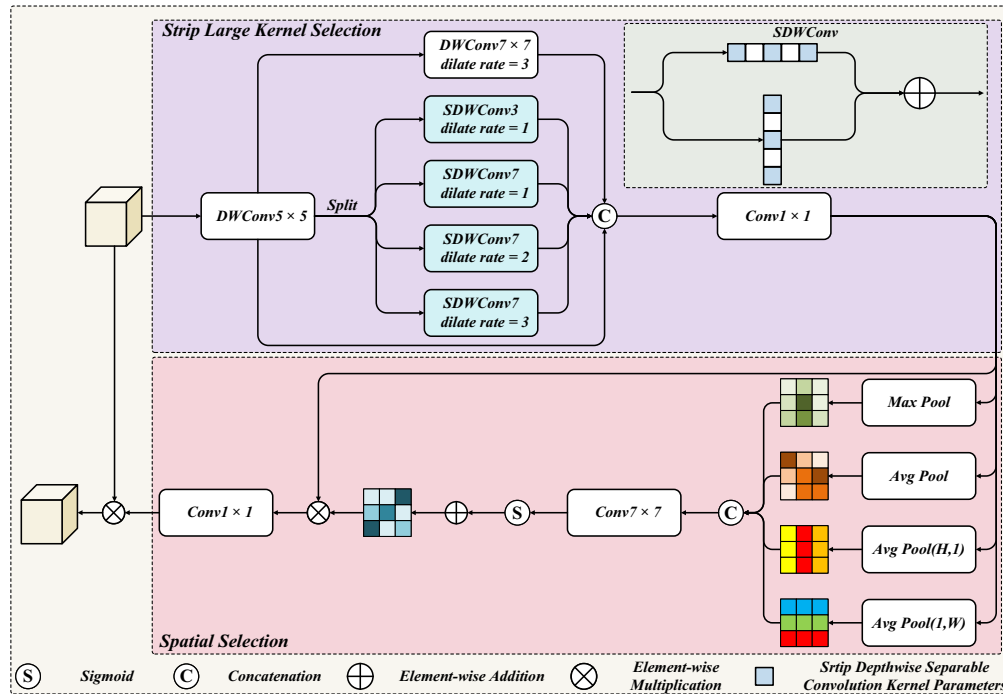


Fig. 5 SLSM architecture diagram: It leverages SDWConv with varying dilation rates to achieve efficient feature extraction and enhance detection capability for flames at multiple scales while improving computational efficiency through diverse pooling operations.

of a fixed receptive field limits the model's capability to recognize flames of varying scales. To address these challenges, we introduce the strip large kernel module (SLKM), inspired by the large kernel selection utilized in LSKNet,³⁷ based on strip depthwise separable convolutions (SDWConv), as illustrated in Fig. 5.

Through an analysis of flame texture features and the spatial distribution of fire occurrences, we implemented a strip large kernel selection within the SLKM module. This design acknowledges the significant relationship between the characteristics of different fire types and their spatial contexts. The SDWConv operates by applying convolution separately along the x and y directions of the input image, combining the resulting information via element-wise addition. This approach enhances the model's comprehension of spatial information while keeping the parameter count and complexity low.

To further bolster the module's capacity to interpret image data, the incoming image first undergoes a 5×5 depthwise separable convolution (DWConv). The processed information is then branched into two paths: one branch processes the data through a 7×7 DWConv, whereas the other is split and directed into SDWConv with varying kernel sizes. The outputs from these processes are subsequently fused using a Concat operation, followed by a 1×1 convolution for weighted feature integration. This architecture allows for dynamic adjustment of the optimal receptive field during training.

Recognizing the significant variability in flame size and shape during fire events, we developed a multiscale selection pooling mechanism following the application of strip large kernel selection. In this stage, the fire image data respectively undergo max pooling, average pooling, and a custom strip-shaped adaptive pooling tailored for flame texture features and spatial distribution characteristics. This diverse pooling strategy effectively extracts key features of flames across different scales.

The resulting multiscale feature maps are then processed using a 7×7 convolution kernel, followed by a nonlinear transformation via a Sigmoid activation function, which generates feature representations at various scales. These feature maps are fused through element-wise addition, culminating in the final multiscale feature map that enhances the model's understanding of scale variance in fire images.

4 Experiments

This section provides an overview of the experimental setup, hyper-parameter selection, dataset, evaluation metrics, and the validation results of the model.

4.1 Experimental Setup and Hyper-parameter Selection

Model training and testing are conducted on a server running Ubuntu 22.04 and equipped with an NVIDIA GeForce RTX 4090 GPU with 24 GB memory, using CUDA 11.8 and cuDNN 8.9.7. During training, all input images are resized to 640×640 pixels, a batch size of 16 is used, and the training process runs for 300 epochs. To ensure reproducibility and fair comparison, the random seed is fixed to 0 for all experiments, thereby reducing the impact of training randomness. The optimizer used for training is stochastic gradient descent (SGD) with a momentum of 0.937, an initial learning rate of 0.01, and a weight decay of 0.0005. To improve training efficiency, a cosine annealing strategy is applied to dynamically adjust the learning rate throughout the training process, enabling adaptive optimization based on model convergence behavior.

4.2 Dataset and Evaluation Metrics

Forest fires represent highly destructive natural disasters that significantly threaten ecosystems and human life. Therefore, there is a pressing need for technological interventions to prevent and control the spread of these fires. However, existing datasets related to forest fires primarily consist of ground-fixed angle images in a single modality, which fail to adequately capture the complexities of varied terrain, high humidity, low visibility weather conditions, and multiscale light flux density associated with forest fires. To address these limitations, Wang et al.³⁸ developed a multisenario, multiweather condition, and multi-illumination level synthetic dataset for forest fires (M⁴SFWD). This dataset includes remote sensing data on forest fires across diverse terrain types, weather conditions, light flux densities, and various quantities of wildfire objects, encompassing 3974 images labeled as either fire or smoke.

Although the M⁴SFWD dataset has made significant contributions to the study of forest fires, it mainly focuses on this specific type and does not encompass building fires or vehicle fires, which also significantly impact human life. To provide a more comprehensive dataset that includes various types of fires, we initiated the development of a new dataset that enables fire-fighters to manage different fire types more effectively. In our data collection efforts, we referred to the dataset proposed by Yar H⁹ and further categorized it into building fire, vehicle fire, and forest fire. In addition, to mitigate classification difficulties arising from class imbalance, we incorporated images of building, vehicle, and forest fires from the public DFS dataset.³⁹

To enhance the model's classification capabilities and generalization performance, we aimed to integrate fire images with their specific occurrence contexts during collection. In the building fire category, we collected images from a variety of building types, including American bungalows, modern office buildings, contemporary residential complexes, constructions, and historical structures. The vehicle fire category contained images depicting diverse vehicle fires, such as those involving motorcycles, buses, sedans, race cars, and new energy vehicles. These incidents had multiple causes, including vehicle collisions, fires occurring at racing events, and fires resulting from improper vehicle maintenance. The forest fire category comprised images of both forest and grassland fires, captured at different times of day, depicting flames ranging from large blazes to small sparks.

For building fires, flames are annotated jointly with buildings; for vehicle fires, flames are labeled together with vehicles; and for forest fires, flames are annotated in association with forest or grass regions. In addition, to address potential distractions caused by visually similar red objects during fire detection, we introduce an additional background category labeled "other." This category comprises hard negative samples that are prone to misidentified as fires, including toy lights resembling flames, table lamps, red flowers, sunlight, and red clothing. In practical deployment scenarios, such objects frequently trigger false alarms when incorrectly identified as fire events. The finalized dataset contains 3964 images, including 933 building fire samples, 1071 vehicle fire samples, 986 forest fire samples, and 974 images labeled as "others." Representative samples from the dataset are shown in Fig. 6.

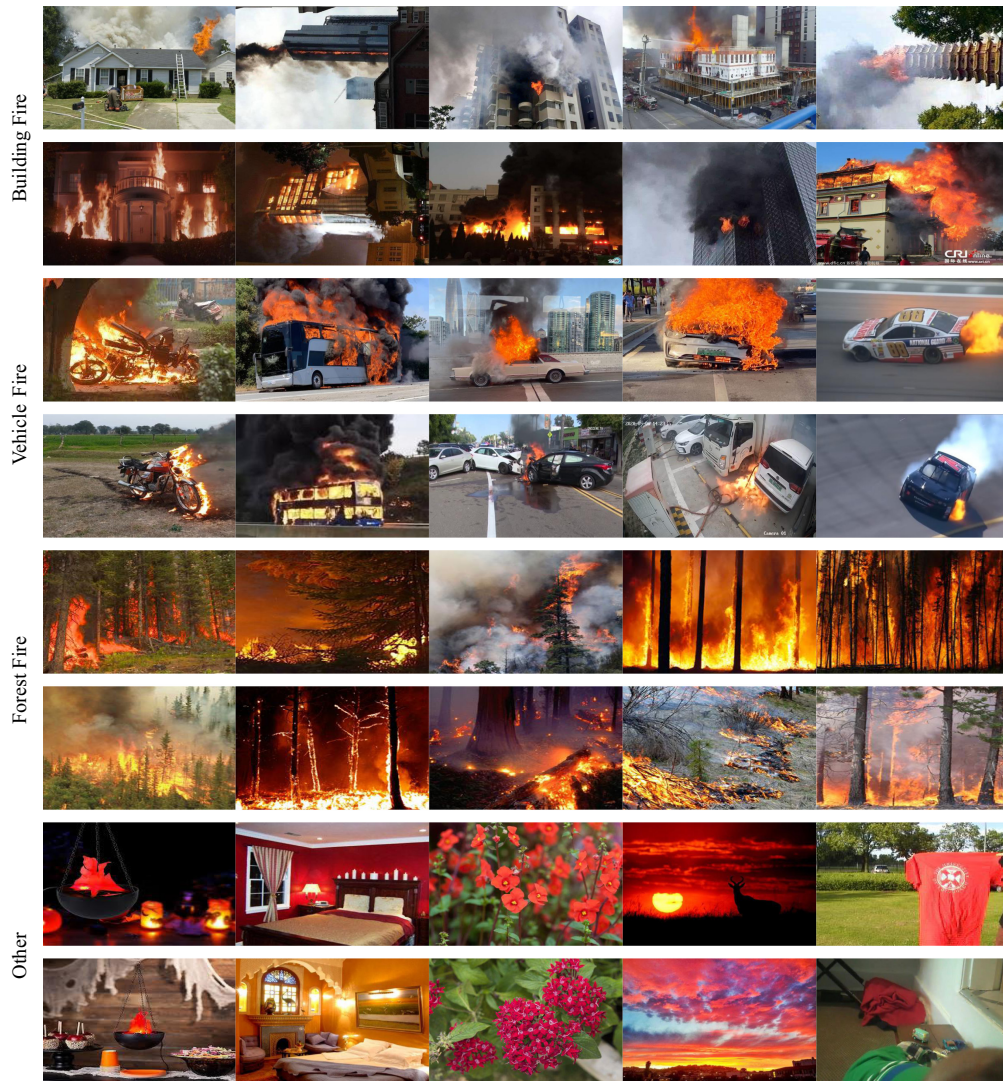


Fig. 6 Sample images of the self-created dataset.

We evaluated the model using precision, recall, and mean average precision (mAP). The calculation for precision is presented below:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (3)$$

Subsequently, recall is computed using the following equation:

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (4)$$

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively. The mean average precision (mAP) is derived by averaging the average precision (AP), which is calculated by determining the area under the precision–recall (P–R) curve, thus providing a more comprehensive assessment of the detection algorithm’s performance. Its calculation formula is shown in Eq. (5).

$$AP = \int_0^1 P(R) dR. \quad (5)$$

Once the AP for each category is obtained, the mean average precision (mAP) is computed according to Eq. (6), where n signifies the number of categories.

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^n \text{AP}_i. \quad (6)$$

When the intersection over union (IoU) threshold is set at 0.5, the resulting AP is denoted as $\text{AP}_{0.5}$. Thus, mAP at this threshold, denoted as $\text{mAP}_{0.5}$, is obtained by averaging $\text{AP}_{0.5}$ across all categories.

4.3 Performance Evaluation of the Proposed Model

In this section, we provide a brief overview of the performance comparison of our model against state-of-the-art methods, focusing on both quantitative and qualitative analyses.

4.3.1 Quantitative comparisons

To thoroughly evaluate the performance of the proposed model, we conduct a series of comprehensive comparative experiments on our self-created dataset. As the proposed method is an optimized version based on YOLOv11n, we compare it with mainstream state-of-the-art detection models. To ensure fairness under edge-computing constraints, we consistently select the lightest variant within each model series. Specifically, this includes seven YOLO-series models and the transformer-based detector RT-DETR-R18,⁴⁰ which serves as the most lightweight non-YOLO baseline. Furthermore, because forest fires represent one of the most catastrophic types of fire disasters, we also conduct experiments on the M⁴SFWD forest fire dataset. The detailed quantitative results are presented in Table 1.

As the experiments involve two datasets with different category definitions, we report the average detection results across three categories for the self-created dataset and across two categories for the M⁴SFWD dataset separately. Detailed detection results for each category are provided in Table 2 for further reference. Although the proposed model yields slightly lower precision or recall compared with certain variants, it achieves superior performance in terms of $\text{mAP}_{0.5}$, indicating stronger overall detection capability. Precision and recall are more susceptible to evaluation noise, whereas $\text{mAP}_{0.5}$ provides a more comprehensive assessment by jointly considering precision–recall performance across different IoU thresholds. These results demonstrate that the proposed method effectively enhances overall flame detection accuracy and improves robustness in recognizing flames at different scales. By better integrating fire characteristics with spatial information, the model exhibits strong adaptability and robustness in real-world fire detection scenarios.

Moreover, according to the no free lunch theorem, no single algorithm performs optimally across all tasks. Although YOLO-series detectors achieve excellent performance on general-

Table 1 Comparison of experimental results.

Methods	M ⁴ SFWD			Self-created dataset		
	Precision (%)	Recall (%)	$\text{mAP}_{0.5}$ (%)	Precision (%)	Recall (%)	$\text{mAP}_{0.5}$ (%)
YOLOv5n ⁴¹	84.8	80.9	87.0	79.8	68.8	74.7
YOLOv6n ⁴²	83.8	80.5	87.0	77.4	69.7	74.3
YOLOv7t ⁴³	83.4	84.4	87.1	64.8	67.9	67.5
YOLOv8n ⁴⁴	83.3	81.0	86.9	75.0	72.8	73.6
YOLOv9t ⁴⁵	84.4	81.5	87.1	78.1	67.3	73.4
YOLOv10n ⁴⁶	81.2	78.9	85.0	78.8	66.1	72.6
YOLOv11n ⁴⁷	84.0	82.3	87.6	77.9	72.1	75.5
RT-DETR-R18 ⁴⁰	85.0	78.0	85.9	69.7	71.7	72.9
Ours	84.5	81.0	87.7	79.4	72.4	76.8

Note: bold values indicate the best results in each column.

Table 2 Comparison of experimental results on each class.

Methods	M ⁴ SFWD			Self-created dataset			
	AP _{0.5} (%)		mAP _{0.5} (%)	AP _{0.5} (%)			mAP _{0.5} (%)
	Fire	Smoke		Building_Fire	Vehicles_Fire	Forest_Fire	
YOLOv5n ⁴¹	87.7	86.4	87.0	66.0	94.4	63.8	74.7
YOLOv6n ⁴²	87.7	86.2	86.9	66.4	92.4	64.1	74.3
YOLOv7 ⁴³	88.3	85.9	87.1	65.4	90.1	47.0	67.5
YOLOv8n ⁴⁴	87.9	85.9	86.9	68.1	93.4	59.4	73.6
YOLOv9t ⁴⁵	87.9	86.4	87.1	67.0	92.9	60.4	73.4
YOLOv10n ⁴⁶	87.0	83.1	85.0	65.6	92.2	59.9	72.6
YOLOv11n ⁴⁷	88.2	86.9	87.6	69.6	93.9	62.9	75.5
Ours	88.6	86.7	87.7	67.9	94.5	68.0	76.8

Note: bold values indicate the best results in each column.

purpose datasets such as COCO, the irregular shapes of flames, large-scale variations, and dataset-specific characteristics in fire detection limit the effectiveness of some alternative models in this domain.

4.3.2 Time complexity

We present a feasibility analysis of the proposed model by examining floating-point operations (GFLOPs), parameter count, GPU inference time, and model size. We compute these metrics for all compared models on both datasets, and the results are summarized in Table 3. The comparison covers multiple YOLO-series variants as well as the non-YOLO detector RT-DETR-R18.

The results in Table 3 reveal a clear trade-off for transformer-based detectors. Although RT-DETR demonstrates strong detection capability, it incurs significantly higher computational cost and slower inference speed compared with YOLO-series models. By contrast, the proposed method maintains a favorable balance between performance and efficiency. Although the introduction of the DGSS and SLSM modules slightly increases the parameter count and GFLOPs relative to the baseline YOLOv8n, the model remains computationally efficient and achieves competitive inference speed. When considered together with the superior accuracy reported in Table 1, the proposed method demonstrates strong overall capability and suitability for real-time fire detection scenarios where both accuracy and speed are critical.

To further highlight this advantage, we conduct additional comparisons with multiple YOLOv8 variants, and the corresponding analysis and results are provided in Table 4.

4.3.3 Qualitative analysis

According to the evaluation results shown in Table 1, we selected YOLOv5, YOLOv8, and YOLOv11 models from Ultralytics, alongside our proposed model, for comparison in the main text to assess their performance. This analysis involves identifying regions of interest and categorizing them based on specific labels, with experimental outcomes illustrated in Fig. 7.

As shown in Fig. 7, YOLOv5 and YOLOv8 frequently produce inaccurate bounding boxes in the “building fire” and “vehicle fire” categories, with issues such as overlapping or fragmented bounding boxes, and exhibit lower bounding box confidence in the “forest fire” category. Although YOLOv11 demonstrates higher bounding accuracy, its overall confidence is lower than that of our model, suggesting a higher likelihood of false positives during fire detection. This comparative experiment aligns well with the numerical results reported in Table 1.

To further validate the effectiveness of our model, we performed a specific visualization, as shown in Fig. 8. This figure demonstrates that the model accurately identifies various types of

Table 3 Model complexity analysis.

Methods	YOLOv5n ⁴¹	YOLOv6n ⁴²	YOLOv7t ⁴³	YOLOv8n ⁴⁴	YOLOv9t ⁴⁵	YOLOv10n ⁴⁶	YOLOv11n ⁴⁷	RT-DETR-R18 ⁴⁰	Ours
M ⁴ SFWD	GFLOPs	7.1	11.8	13.2	8.1	7.6	8.2	55.9	9.4
	Parameters (M)	2.50	4.23	6.02	3.00	1.97	2.69	18.8	6.38
	Inference time (ms)	1.8	1.8	4.3	2.0	1.6	1.5	4.5	2.1
	Model size (MB)	5.3	8.7	12.3	6.2	4.6	5.8	35.3	12.8
Self-created dataset	GFLOPs	7.1	11.8	13.2	8.1	7.9	8.2	55.9	8.7
	Parameters (M)	2.50	4.23	6.02	3.00	2.00	2.71	18.8	6.38
	Inference time (ms)	1.8	2.3	4.1	2.2	1.6	1.2	4.2	2.3
	Model size (MB)	5.2	8.5	11.9	6.1	4.5	5.6	35.1	12.7

Note: bold values indicate the best results in each column.

Table 4 Results of the comparison with the YOLOv8 series models.

	Methods	YOLOv8n	YOLOv8s	YOLOv8m	YOLOv8l	YOLOv8x	Ours
M ⁴ SFWD	mAP _{0.5}	86.9%	87.2%	86.7%	86.6%	86.4%	87.7%
	GFLOPs	8.1	28.4	78.7	164.8	257.4	9.4
	Parameters	3,006,038	11,126,358	25,840,918	43,608,150	68,125,494	6,386,319
	Inference time	2.0 ms	3.0 ms	4.5 ms	2.8 ms	2.9 ms	2.1 ms
	Model size (MB)	6.2	22.5	52	87.6	136.7	12.8
Self-created dataset	mAP _{0.5}	73.6%	77.1%	76.0%	75.3%	74.0%	76.8%
	GFLOPs	8.1	28.4	78.7	164.8	257.4	8.7
	Parameters	3,006,233	11,126,745	25,841,497	43,608,921	68,126,457	6,386,319
	Inference time	2.2 ms	2.5 ms	2.6 ms	2.7 ms	3.8 ms	2.3 ms
	Model size (MB)	6.1	22.5	52	87.7	136.7	12.7

**Fig. 7** Comparative analysis of the proposed model with YOLOv5, YOLOv8, and YOLOv11.

fires with consistently high confidence levels, highlighting its strong performance and reliability in practical detection scenarios.

4.3.4 Ablation studies

In this section, we use YOLOv8n as the baseline model to evaluate the effect of adding each module. To assess these modules' performance, we conducted three different experiments on two separate datasets. First, we integrated the DGSS module with YOLOv8n, then replaced the SPPF module with SLSM, and finally combined both with YOLOv8n for comparison. For clarity, Table 5 summarizes the results as category averages.

We first examine the model's performance on our custom dataset. As shown in Table 6, adding the DGSS module notably improves accuracy for the "Vehicles_Fire" and "Forest_Fire" categories, likely due to the strong spatial characteristics in these categories, aligning well with DGSS's focus on spatial information capture. After incorporating SLSM, the model shows a modest improvement for "Vehicles_Fire" and a substantial increase for "Forest_Fire," likely because both categories often feature fires of varying sizes, which SLSM captures through multi-scale feature extraction.

Next, we analyze the results from ablation experiments on the M⁴SFWD dataset for each category. The results indicate that integrating DGSS significantly enhances accuracy for "fire" and "smoke" likely due to the pronounced spatial characteristics that DGSS captures effectively. After adding SLSM, the model achieves a notable accuracy improvement for "fire" and a slight increase for "smoke" likely due to SLSM's ability to capture multiscale flame features common in these categories, where fires vary greatly in size and shape.

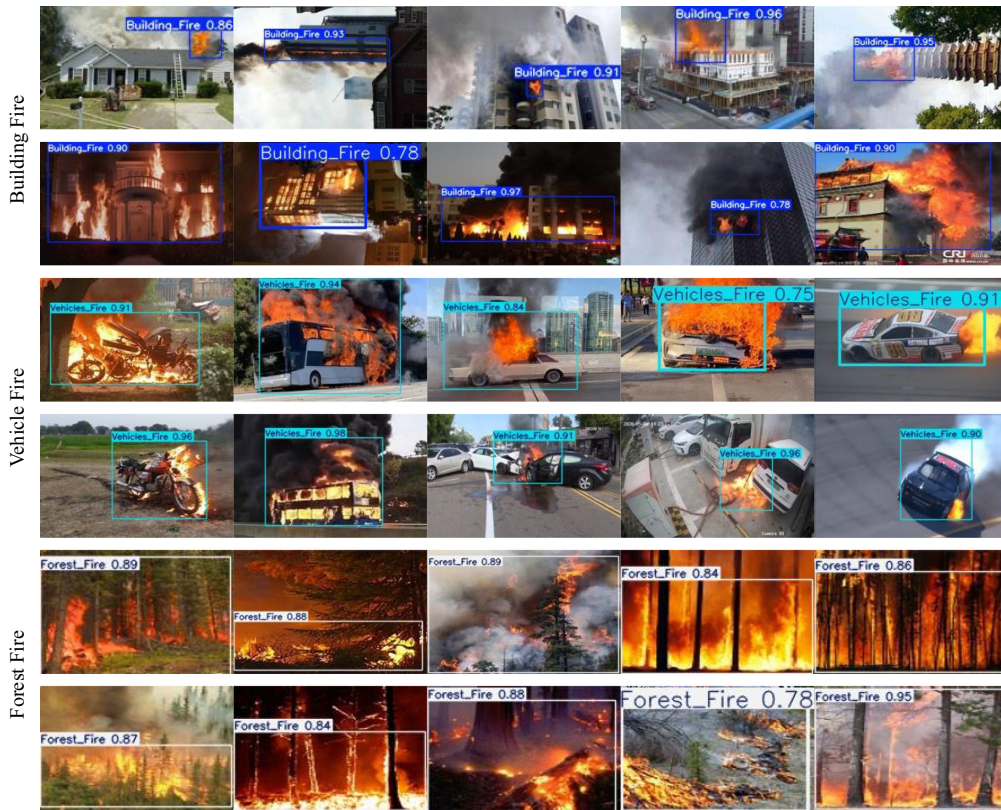


Fig. 8 Output results of our model using the self-created dataset.

Table 5 Ablation experiment results.

Methods	M ⁴ SFWD			Self-created dataset		
	Precision (%)	Recall (%)	mAP _{0.5} (%)	Precision (%)	Recall (%)	mAP _{0.5} (%)
YOLOv8n	82.3	80.0	86.0	75.0	72.8	73.6
YOLOv8n + DGSS	83.3	81.0	86.9	82.9	67.2	75.8
YOLOv8n + SLSM	84.3	79.8	86.8	77.3	71.5	74.3
Ours	84.5	81.0	87.7	79.4	72.4	76.8

Note: bold values indicate the best results in each column.

Table 6 Results of each module for each category in the ablation experiment.

Methods	M ⁴ SFWD		Self-created dataset		
	AP _{0.5} (%)		AP _{0.5} (%)		
	Fire	Smoke	Building_Fire	Vehicles_Fire	Forest_Fire
YOLOv8n	86.8	85.2	68.1	93.4	59.4
YOLOv8n + DGSS	87.7	86.2	65.6	94.3	67.7
YOLOv8n + SLSM	88.3	85.9	65.7	93.8	63.4
Ours	88.6	86.7	67.9	94.5	68.0

Note: bold values indicate the best results in each column.

5 Conclusion

With the growing emphasis on smart technology and “smart city” initiatives, there is an urgent demand for advanced fire detection technology. Fires pose severe risks to public safety, leading to significant economic losses and environmental damage. Consequently, the concept of “smart firefighting” has emerged as a crucial component of modern urban fire prevention systems. Although fire detection technology has advanced considerably, current research primarily targets the presence or absence of fires, lacking detailed fire type classification and efficient information transmission mechanisms—key challenges in fire emergency management.

To enhance the response efficiency and classification accuracy of smart firefighting systems, this paper presents an improved YOLOv8 model, referred to as EFDC-YOLO. EFDC-YOLO incorporates our newly developed DGSS and SLSM modules, which strengthen the capability to capture spatial flame characteristics and extract multiscale flame features, leading to significant improvements in fire type classification accuracy. Experimental results indicate that EFDC-YOLO outperforms the baseline model on both a custom dataset and the public M⁴SFWD dataset, achieving mAP_{0.5} scores of 76.8% and 87.7%, respectively, with inference times of 2.3 and 2.1 ms. These results demonstrate that EFDC-YOLO achieves a well-optimized balance between detection accuracy, computational speed, and model size, making it a highly efficient solution for smart firefighting applications.

In future research, although the proposed EFDC-YOLO model achieves a favorable balance between detection accuracy and speed, further optimization is necessary for practical applications, particularly for deployment on edge devices. This optimization will aim to enhance detection speed and reduce model parameters while maintaining high accuracy. In addition, the current dataset remains limited in image volume; expanding the dataset will be essential to improve the model’s generalization ability. The platform developed thus far serves only as a demonstration and has not yet been deployed in real-world applications. Future efforts will focus on incorporating knowledge distillation and pruning techniques to accelerate model inference, improve performance, and reduce both computational complexity and parameter count.

Disclosures

The authors declare that there are no financial interests, commercial affiliations, or other potential conflicts of interest that could have influenced the objectivity of this research or the writing of this paper.

Code and Data Availability

Our codes and dataset are publicly available at <https://github.com/Administor123/EFDC-YOLO/>

Acknowledgments

This research was supported in part by the Guangxi Natural Science Foundation (Grant No. 2024GXNSFAA010493), the National Natural Science Foundation of China (Grant No. 61862015), the Science and Technology Project of Guangxi (Grant No. Guike AD23023002), and the Guangxi Key Research and Development Program (Grant No. AB17195025).

References

1. S. Dogan et al., “Automated accurate fire detection system using ensemble pretrained residual network,” *Expert Syst. Appl.* **203**, 117407 (2022).
2. S. Khan et al., “DeepSmoke: deep learning model for smoke detection and segmentation in outdoor environments,” *Expert Syst. Appl.* **182**, 115125 (2021).
3. M. Filonchyk and M. P. Peterson, “Changes in aerosol properties at the El Arenosillo site in southern Europe as a result of the 2023 Canadian forest fires,” *Environ. Res.* **260**, 119629 (2024).
4. M. Ahrens and B. Evarts, “Fire loss in the United States during 2020,” National Fire Protection Association, 2022, <https://www.nfpa.org/education-and-research/research/nfpa-research/fire-statistical-reports/fire-loss-in-the-united-states> (accessed 15 February 2026).
5. H. Jan et al., “Raspberry pi assisted safety system for elderly people: an application of smart home,” in *First Int. Conf. of Smart Syst. and Emerg. Technol. (SMARTTECH)*, IEEE, pp. 155–160 (2020).

6. H. Yar et al., "Towards smart home automation using IoT-enabled edge-computing paradigm," *Sensors* **21**(14), 4932 (2021).
7. N. Yunusov et al., "Robust forest fire detection method for surveillance systems based on You Only Look Once version 8 and transfer learning approaches," *Processes* **12**(5), 1039 (2024).
8. D. Kong, Y. Li, and M. Duan, "Fire and smoke real-time detection algorithm for coal mines based on improved yolov8s," *PLoS One* **19**(4), e0300502 (2024).
9. H. Yar et al., "A modified Yolov5 architecture for efficient fire detection in smart cities," *Expert Syst. Appl.* **231**, 120465 (2023).
10. J. Huang et al., "A small-target forest fire smoke detection model based on deformable transformer for end-to-end object detection," *Forests* **14**(1), 162 (2023).
11. R. Li et al., "Smwe-GFPNet: a high-precision and robust method for forest fire smoke detection," *Knowl.-Based Syst.* **289**, 111528 (2024).
12. J. Lin, H. Lin, and F. Wang, "A semi-supervised method for real-time forest fire detection algorithm based on adaptively spatial feature fusion," *Forests* **14**(2), 361 (2023).
13. X. Han et al., "Benchmarking multi-scene fire and smoke detection," *Lect. Notes Comput. Sci.* **15041**, 203–218 (2024).
14. Y. Liu et al., "VMamba: visual state space model," in *Adv. in Neural Inform. Process. Syst.*, Vol. 37, pp. 103031–103063, Curran Associates, Inc. (2024).
15. Q. Hou et al., "Strip pooling: rethinking spatial pooling for scene parsing," in *Proc. IEEE/CVF Conf. on Comput. Vision and Pattern Recognit.*, pp. 4003–4012 (2020).
16. G. Cheng et al., "Visual fire detection using deep learning: a survey," *Neurocomputing* **596**, 127975 (2024).
17. A. Jadon et al., "FireNet: a specialized lightweight fire & smoke detection model for real-time IoT applications," 2019, <https://arxiv.org/abs/1905.11922> (accessed 15 February 2026).
18. P. Li and W. Zhao, "Image fire detection algorithms based on convolutional neural networks," *Case Stud. Thermal Eng.* **19**, 100625 (2020).
19. C. Li et al., "Fast forest fire detection and segmentation application for UAV-assisted mobile edge computing system," *IEEE Internet Things J.* **11**(16), 26690–26699 (2023).
20. L. Huang et al., "YOLO-ULNet: ultra-lightweight network for real-time detection of forest fire on embedded sensing devices," *IEEE Sens. J.* **24**(15), 25175–25185 (2024).
21. P. Barmpoutis et al., "Fire detection from images using faster R-CNN and multidimensional texture analysis," in *ICASSP 2019-2019 IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, IEEE, pp. 8301–8305 (2019).
22. S. Li, Q. Yan, and P. Liu, "An efficient fire detection method based on multiscale feature extraction, implicit deep supervision and channel attention mechanism," *IEEE Trans. Image Process.* **29**, 8467–8475 (2020).
23. P. Huang et al., "A combined real-time intelligent fire detection and forecasting approach through cameras based on computer vision method," *Process. Saf. Environ. Prot.* **164**, 629–638 (2022).
24. Y. Wang et al., "A fast video fire detection of irregular burning feature in fire-flame using in indoor fire sensing robots," *IEEE Trans. Instrum. Meas.* **71**, 7505414 (2022).
25. R. Xu et al., "A forest fire detection system based on ensemble learning," *Forests* **12**(2), 217 (2021).
26. M. Li et al., "Early forest fire detection based on deep learning," in *3rd Int. Conf. on Ind. Artif. Intell. (IAI)*, Shenyang, China, IEEE, pp. 1–5 (2021).
27. D. Ren et al., "FCLGYOLO: feature constraint and local guided global feature for fire detection in unmanned aerial vehicle imagery," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **17**, 5864–5875 (2024).
28. Y. Ahn, H. Choi, and B. S. Kim, "Development of early fire detection model for buildings using computer vision-based CCTV," *J. Build. Eng.* **65**, 105647 (2023).
29. T. Wang et al., "Improving YOLOx network for multi-scale fire detection," *Vis. Comput.* **40**(9), 6493–6505 (2024).
30. G. Chen et al., "Multi-scale forest fire recognition model based on improved YOLOv5s," *Forests* **14**(2), 315 (2023).
31. J. Fan et al., "Yolov8-WMD: a frequency domain multi-scale feature fusion algorithm for forest fire detection," *Signal Image Video Process.* **19**(10), 783 (2025).
32. L. Xiao et al., "EMG-YOLO: an efficient fire detection model for embedded devices," *Digit. Signal Process.* **156**, 104824 (2025).
33. K. He et al., "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(9), 1904–1916 (2015).
34. K. Wang et al., "PANet: few-shot image semantic segmentation with prototype alignment," in *Proc. IEEE/CVF Int. Conf. on Comput. Vision*, pp. 9197–9206 (2019).
35. K. Han et al., "ParameterNet: parameters are all you need for large-scale visual pretraining of mobile networks," in *Proc. IEEE/CVF Conf. on Comput. Vision and Pattern Recognit.*, pp. 15751–15761 (2024).
36. K. Han et al., "GhostNet: more features from cheap operations," in *Proc. IEEE/CVF Conf. on Comput. Vision and Pattern Recognit.*, pp. 1580–1589 (2020).

37. Y. Li et al., "Large selective kernel network for remote sensing object detection," in *Proc. IEEE/CVF Int. Conf. on Comput. Vision*, pp. 16794–16805 (2023).
38. G. Wang et al., "M4SFWD: a multi-faceted synthetic dataset for remote sensing forest wildfires detection," *Expert Syst. Appl.* **248**, 123489 (2024).
39. S. Wu et al., "A dataset for fire and smoke object detection," *Multimedia Tools Appl.* **82**(5), 6707–6726 (2023).
40. Y. Zhao et al., "DETRs beat YOLOs on real-time object detection," in *Proc. IEEE/CVF Conf. on Comput. Vision and Pattern Recognit.*, pp. 16965–16974 (2024).
41. G. Jocher, "Ultralytics yolov5," GitHub, 2020, <https://github.com/ultralytics/yolov5> (accessed 15 February 2026).
42. C. Li et al., "YOLOv6: a single-stage object detection framework for industrial applications," 2022, <https://arxiv.org/abs/2209.02976> (accessed 15 February 2026).
43. C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. on Comput. Vision and Pattern Recognit.*, pp. 7464–7475 (2023).
44. G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," GitHub, 2023, <https://github.com/ultralytics/ultralytics> (accessed 15 February 2026).
45. C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: learning what you want to learn using programmable gradient information," 2024, <https://arxiv.org/abs/2402.13616> (accessed 15 February 2026).
46. A. Wang et al., "YOLOv10: real-time end-to-end object detection," 2024, <https://arxiv.org/abs/2405.14458> (accessed 15 February 2026).
47. G. Jocher and J. Qiu, "Ultralytics yolo11," GitHub, 2024, <https://github.com/ultralytics/ultralytics> (accessed 15 February 2026).

Zhengjie Wang received his BS degree from Guilin University of Electronic Technology, Guilin, China, in 2023. He is currently pursuing his MS degree with the College of Artificial Intelligence, Taiyuan University of Technology, Taiyuan, China. His current research interests include computer vision and deep learning.

Zengmin Xu received his PhD from Wuhan University, Wuhan, China, in 2019. He is currently an associate professor with the School of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin, China. He has authored or coauthored over 30 refereed journal and conference papers. His research interests include computer vision and pattern recognition.