ICCK Transactions on Intelligent Systematics
http://dx.doi.org/10.62762/TIS.2026.406065

RESEARCH ARTICLE

Low-Light Fire and Smoke Detection in Electric Vehicles via a Lightweight Dual-Branch Architecture with Adaptive Reflectance Enhancement

Xiu Tang[1], Yaxin Man[1], Xiaoyan Wang[1], Rendong Ji[1,*], Xiaojun Zhang[1] and Yunlong Xu[1]

[1] Faculty of Electronic Information Engineering, Huai'an University, Huai'an 223003, China

Abstract

Fire and smoke detection in electric vehicle environments is essential for safety monitoring, yet existing methods often perform poorly under low-light conditions because of weak feature representation, flame overexposure, and limited nighttime data. In addition, many detection models cannot effectively balance accuracy and real-time efficiency, which restricts their deployment in practical applications. To address these issues, this study proposes a lightweight dual-branch detection framework for low-light fire and smoke recognition in electric vehicle environments. The proposed method integrates reflectance enhancement, adaptive feature weighting, and multi-scale feature learning within a unified architecture. Specifically, the auxiliary branch enhances weak fire-related features under low illumination through the Adaptive Reflectance Enhancement Module (AREM) and the Element-wise Adaptive Weight Convolution (EAConv), while the main branch employs a Self-Adaptive Feature Convolution (SAConv) module to strengthen multi-scale representation with limited computational overhead. Moreover, a self-constructed and publicly released dataset containing 3,512 images is introduced, including 1,910 daytime and 1,602 nighttime samples, to provide a more realistic benchmark for fire detection in electric vehicle environments under diverse lighting conditions. Experimental results show that the proposed method achieves 85.3% AP50 and 52.2% AP on the full dataset, and 52.9% AP50 and 26.5% AP on the nighttime subset. Compared with YOLOv11-n, the proposed model improves AP50 by 1.2% and AP by 1.7%, while demonstrating more stable performance in challenging low-light environments. These results indicate that the proposed framework provides a practical and reliable solution for real-time fire detection in electric vehicle environments.

Keywords: electric vehicle safety, electric vehicles, fire detection, low-light optimization.

Academic Editor:
Manolis Vavalis

Submitted: 20 April 2026
Revised: 03 August 2026
Accepted: 07 August 2026
Published: 01 September 2026

Vol. 3, **No.** 3, 2026.
10.62762/TIS.2026.406065

***Corresponding author:**
Rendong Ji
jrdgxy@163.com

Citation
Tang, X., Man, Y., Wang, X., Ji, R., Zhang, X., & Xu, Y. (2026). Low-Light Fire and Smoke Detection in Electric Vehicles via a Lightweight Dual-Branch Architecture with Adaptive Reflectance Enhancement. *ICCK Transactions on Intelligent Systematics*, 3(3), 145–161.

1 Introduction

Fire and smoke detection is essential for safety monitoring in electric vehicle (EV) environments, where early identification of fire hazards can effectively reduce accident severity and prevent secondary losses. Compared with general fire scenarios, fire incidents in EV charging stations often occur under low-light conditions and are characterized by small-scale flames, weak visual features, and complex background interference. These factors significantly increase the difficulty of accurate fire and smoke detection in real-world applications.

Early fire detection methods mainly relied on traditional computer vision techniques, such as Gaussian Mixture Models (GMM) and multi-color feature analysis, which were effective in constrained environments but often suffered from high false alarm rates and limited robustness in complex scenes [1]. With the rapid development of deep learning, convolutional neural network (CNN)-based methods have significantly improved fire detection accuracy in surveillance and industrial applications [2]. In addition, self-learning and feature-weighting strategies have been introduced to further refine feature representation and improve detection performance in diverse scenarios [3]. Recent research has increasingly focused on improving both detection accuracy and computational efficiency. Small-object fire detection models such as Fire-YOLO have shown promising performance in low-visibility conditions [4], while end-to-end structured networks have been designed to optimize feature extraction for fire and smoke detection [5]. However, most of these methods are primarily developed for general fire scenarios such as forest fires, and even lightweight or small-target-oriented approaches [6, 7] remain focused on outdoor aerial or woodland environments, leaving the detection of small-scale flames and subtle smoke in EV charging stations—particularly under nighttime illumination—largely unaddressed. Beyond the challenge of small target sizes, lighting variation further complicates the detection task. Low-light conditions remain a major challenge for fire detection. In nighttime environments, flame overexposure and low smoke contrast often degrade feature representation and reduce detection stability. To address such problems, high-performance detection frameworks have also been developed for domain-specific fire scenarios, such as industrial cable monitoring, where real-time reliability and robustness under complex background conditions are critical requirements [8]. Meanwhile, several studies have investigated fire detection under degraded visibility conditions such as haze, fog, and nighttime illumination by combining attention mechanisms with data augmentation and adaptive feature enhancement [9, 10]; however, these methods are evaluated on general surveillance or remote sensing datasets and have not been specifically validated for EV charging environments where overexposed flames and low-contrast smoke under artificial lighting introduce additional detection challenges. Meanwhile, lightweight CNN-based models have also been proposed to balance detection accuracy and real-time efficiency for practical applications [11]. Despite these advances, many existing fire detection models still require considerable computational resources, which limits their deployment in real-time EV monitoring systems [12, 13]. Although lightweight YOLO-based variants have been proposed to improve efficiency in fire detection [14], achieving a simultaneous balance between real-time performance and robustness under complex low-light conditions remains challenging. Other approaches, including low-complexity CNN models and hybrid methods such as BowFire, improve efficiency to some extent, but they often lack sufficient adaptability in complex and dynamic fire scenarios [15, 16].

Although considerable progress has been achieved in forest fire monitoring, industrial safety, and video surveillance, the applicability of existing methods to EV charging environments remains insufficiently explored [17]. Multi-sensor fusion and 3D convolution-based smoke detection methods have also been explored, but their high computational complexity limits their practicality in real-time EV fire monitoring systems [18, 19]. Based on the above analysis, several limitations remain in existing fire detection methods. First, currently available datasets lack sufficient coverage of EV fire scenarios, particularly under nighttime conditions. Second, existing models still struggle to effectively represent weak fire-related features under low illumination, resulting in unstable detection performance. Third, many high-accuracy methods rely on computationally intensive architectures, which limits their suitability for real-time deployment in resource-constrained environments. Therefore, there is still a need for a lightweight and robust detection framework specifically designed for low-light EV fire detection.

To address these challenges, this study proposes a lightweight dual-branch framework for fire and smoke recognition in EV environments. The proposed framework integrates AREM, EAConv, and SAConv to improve weak feature representation and multi-scale feature learning under low-light conditions. In addition, a dedicated EV-Fire dataset is constructed and publicly released, covering diverse fire scenarios under different illumination conditions, particularly nighttime environments. Experimental results demonstrate that the proposed framework achieves superior accuracy and robustness compared with existing real-time detectors.

The main contributions of this work are summarized as follows:

1. A lightweight dual-branch detection framework integrating reflectance enhancement, adaptive feature weighting, and multi-scale feature learning is proposed for low-light fire detection.
2. An auxiliary feature enhancement mechanism is designed to suppress background interference and enhance discriminative fire-related features.
3. A dedicated EV fire dataset is constructed with a focus on nighttime conditions, providing a more realistic benchmark for evaluation.
4. Extensive experiments demonstrate that the proposed method achieves superior performance, particularly in challenging low-light environments.

The remainder of this paper is organized as follows. Section 2 reviews related work on fire datasets and fire detection methods. Section 3 introduces the proposed lightweight dual-branch detection framework, including the AREM, EAConv, and SAConv modules. Section 4 presents the EV-Fire dataset, experimental settings, ablation studies, and comparison results. Finally, Section 5 concludes this paper and discusses future research directions.

2 Related Work

Publicly available datasets play an important role in the development of fire detection algorithms. Existing fire datasets, such as D-Fire and the KMU Fire & Smoke Database, have been widely used in early fire detection research and have contributed significantly to the advancement of fire recognition methods [20, 21]. However, most of these datasets mainly focus on indoor and outdoor scenes, wildfires, or conventional fire incidents, and contain only limited samples related to vehicle fires. This shortage of EV-related fire data restricts the training and evaluation of detection models in electric vehicle scenarios. Although the Fire Detection Dataset on Roboflow contains a certain number of vehicle fire images, it includes very few nighttime samples, and its relatively low image quality limits its applicability in low-light fire detection tasks [22]. Therefore, there is still a lack of specialized datasets that can effectively support fire detection research in EV environments, especially under nighttime conditions.

In terms of detection methods, recent studies have proposed various strategies to improve feature extraction and detection robustness in complex fire scenes. Chaoxia et al. [23] improved Faster R-CNN by introducing a color-guided anchoring strategy and a global information-guided mechanism, which enhanced detection accuracy and reduced false alarms. Barmpoutis et al. [24] combined deep learning with multidimensional texture analysis to improve practical fire detection performance. Pan et al. [25] proposed EAConv and EADown modules to improve detection efficiency and accuracy, especially for small-object fire detection and precise downsampling. These methods have shown promising results in complex environments, but they still face challenges in low-light scenes with strong background illumination or overexposed flame regions. To further improve detection performance in challenging scenarios, some studies have focused on enhancing feature extraction and model adaptability. Park and Ko [26] proposed a two-step real-time nighttime fire detection framework for urban environments, combining a lightweight ELASTIC-YOLOv3 with a temporal fire-tube mechanism to handle complex artificial lighting interference such as headlights and neon signs. Although this method demonstrates the importance of low-light-aware design for fire detection, it relies on temporal motion cues from consecutive video frames and does not address the weak feature representation caused by flame overexposure and low smoke contrast in EV charging station imagery, motivating the development of image-level reflectance enhancement strategies for static frame detection. Zheng et al. [27] proposed FTA-DETR, which combines a diffusion-based dataset enhancement framework with encoder optimization to address the issues of high computational cost and limited dataset quality in fire detection. Although these methods improve performance in specific fire detection tasks, they are mainly designed for forest fires or general fire scenarios and are not specifically optimized for EV fire detection under low-light conditions.

Overall, existing studies have made substantial progress in dataset construction, feature enhancement, and lightweight detection design. However, three limitations remain. First, currently available datasets still lack sufficient EV fire samples, especially under nighttime conditions. Second, existing methods have limited capability in representing weak fire-related features under low illumination, which leads to unstable detection performance. Third, many high-performance models rely on relatively complex architectures, making real-time deployment difficult in resource-constrained EV monitoring systems. These limitations motivate the development of a lightweight and robust detection framework specifically designed for low-light fire and smoke recognition in electric vehicle environments.

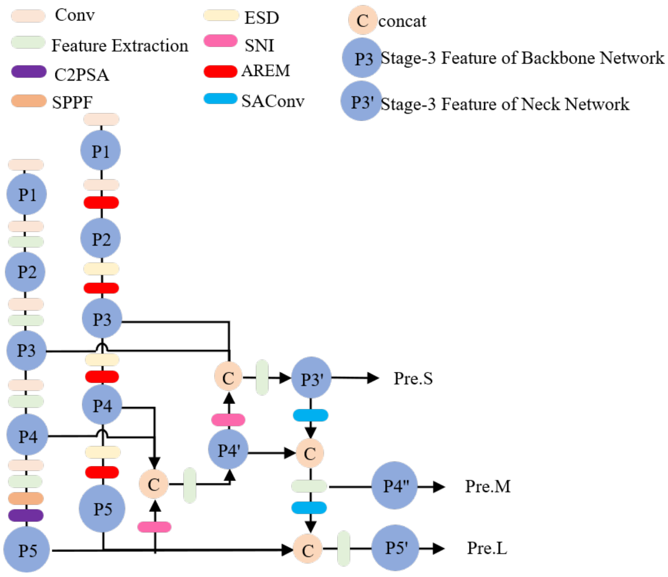


Figure 1. Overall architecture of model.

3 Materials and Methods

This section describes the proposed lightweight dual-branch framework for fire and smoke detection in electric vehicle environments. The overall design is motivated by the challenges of low-light fire detection, where flame overexposure, weak smoke contrast, and complex background interference often lead to unstable feature representation. To address these issues, the proposed framework introduces an auxiliary branch to enhance weak fire-related features and a multi-scale convolution module to strengthen feature extraction across different object scales. Specifically, the auxiliary branch incorporates the Adaptive Reflectance Enhancement Module (AREM) and the Element-wise Adaptive Weight Convolution (EAConv) to improve discriminative feature representation under low illumination, while the Self-Adaptive Feature Convolution (SAConv) module is used to enhance multi-scale modeling with limited computational overhead.

3.1 Framework Overview

To address the challenges of flame and smoke detection in electric vehicle charging stations, including low illumination, complex backgrounds, and large-scale variations, an enhanced fire detection framework based on YOLOv11 is proposed. As shown in Figure 1, the network consists of four main components: a backbone feature extraction branch, an auxiliary feature enhancement branch, a neck-based multi-scale fusion network, and a detection head.

The backbone branch extracts hierarchical semantic features from the input image, while the auxiliary branch enhances fire-related representations through the Adaptive Reflectance Enhancement Module (AREM) and the Enhanced Spatial Downsampling (ESD) module. The extracted features are then fused in the neck network, where the Soft Normalized Interpolation (SNI) module improves feature alignment during top-down fusion, and the Self-Adaptive Feature Convolution (SAConv) module strengthens cross-scale feature interaction during bottom-up aggregation. Finally, the fused features are fed into the detection head for classification and localization, enabling robust detection of flame and smoke targets under complex charging-station environments.

3.2 Auxiliary Structure

To improve fire and smoke detection in low-light environments, a dual-branch architecture is introduced. The main branch is responsible for extracting high-level semantic features and preserving the overall scene context, while the auxiliary branch is designed to enhance local fire-related features that are easily degraded under weak illumination. This complementary design allows the framework to simultaneously model global context and local discriminative cues, thereby improving robustness in challenging EV fire scenes.

As illustrated in Figure 2, the auxiliary branch integrates the Adaptive Reflectance Enhancement Module (AREM) to improve feature representation under low-light conditions. AREM enhances flame-related responses by suppressing overexposed regions and reducing background interference. In

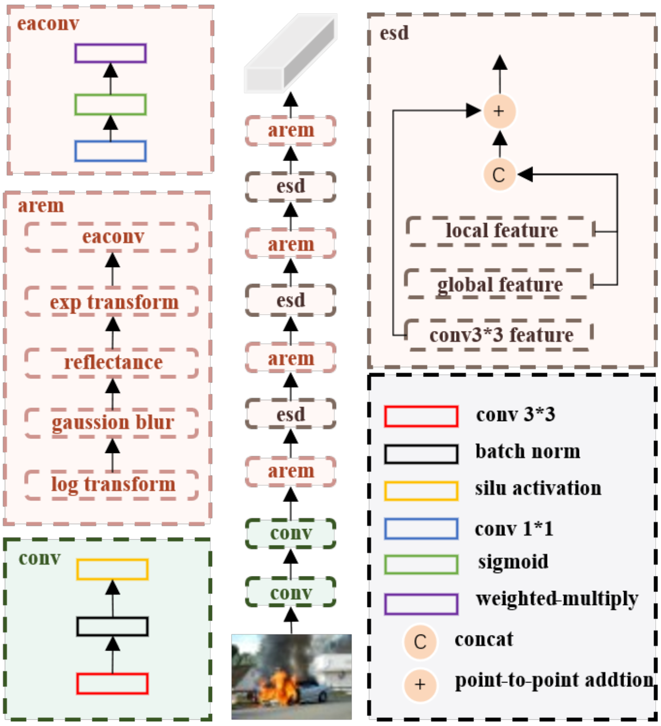


Figure 2. Illustration of the auxiliary structure.

addition, the auxiliary branch incorporates the Extended Spatial Window for Down-sampling (ESD) [28] module to improve the preservation of useful spatial information during down-sampling. This design strengthens the detection of small-scale flames and subtle smoke patterns while maintaining computational efficiency, making the proposed framework suitable for real-time fire detection in electric vehicle environments.

3.2.1 *Adaptive Reflectance Enhancement Module*

Detecting flames in nighttime or poorly illuminated environments is particularly challenging because flame regions are often affected by overexposure, while surrounding regions suffer from low contrast and unstable texture information. These factors weaken the discriminative representation of fire-related features and reduce detection reliability. To address this problem, the proposed framework introduces the Adaptive Reflectance Enhancement Module (AREM), which enhances flame feature representation through reflectance-based enhancement and adaptive feature modulation.

Unified Symbol Definition

1. Input fire image: $\boldsymbol{I} \in \mathbb{R}^{H\times W\times 3}$, where H = image height, W = image width, and the last dimension corresponds to three RGB color channels;
2. RGB channel index: $k \in \{R, G, B\}$, exclusively used for Retinex illumination-reflectance decomposition [29] in AREM preprocessing;
3. After reflectance enhancement, the enhanced image is fed into the EAConv adaptive convolution module. We define:
   - C_{in}: total number of input feature channels for EAConv;
   - C_{out}: total number of output feature channels for EAConv;
   - $i \in \{1, 2, 3, ..., C_{in}\}$: serial index of input feature channels;
   - $o \in \{1, 2, 3, ..., C_{out}\}$: serial index of output feature channels;
   - σ: only represents the standard deviation of Gaussian kernel;
   - $F_i(x, y)$: the enhanced feature map of the i-th input channel after reflectance enhancement, which serves as the input to the EAConv module.
4. Sigmoid activation function is independently defined as:

$$\mathrm{sigm}(z) = \frac{1}{1+\exp(-z)} \tag{1}$$

   to avoid symbol conflict with Gaussian standard deviation σ.

Let $I_k(x, y)$ denote the pixel intensity of the input image at spatial coordinate (x, y) for RGB channel $k \in \{R, G, B\}$. We define $\hat{I}_k(x, y)$ as the logarithmically transformed image, $L_k(x, y)$ as the illumination component, and $R_k(x, y)$ as the reflectance component for channel k, respectively. The complete processing pipeline of AREM is detailed as follows.

Photon saturation causes excessively high brightness in flame regions, leading to local detail loss. To mitigate this issue, we adopt logarithmic transformation to compress the dynamic range of pixel values, where an adjustable factor α controls the overall enhancement strength:

$$\hat{I}_k(x, y) = \log\left(1 + \alpha \cdot I_k(x, y)\right), \quad k \in \{R, G, B\} \tag{2}$$

where $I_k(x, y)$ is the raw pixel value at coordinate (x, y) of channel k, and $\hat{I}_k(x, y)$ is the corresponding transformed result. This transformation maps saturated high-intensity pixels into a compressed value range to preserve fine flame details.

To decouple flame textures from uneven ambient lighting, we estimate global illumination by applying

Gaussian smoothing with a $s \times s$ kernel and adjustable standard deviation σ. The illumination component is derived by convolving the log-transformed image with the Gaussian kernel:

$$\begin{aligned} L_k(x, y) &= (\hat{I}_k * G_\sigma)(x, y) \\ &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \hat{I}_k(x+i, y+j) \cdot G_\sigma(i, j) \end{aligned} \tag{3}$$

where $L_k(x, y)$ represents the illumination map of channel k, which smooths large-scale illumination fluctuations while retaining local structural information of flames; $G_\sigma(i, j)$ is the Gaussian function with standard deviation σ; and $*$ denotes the convolution operation. Since the Gaussian kernel is centrosymmetric, convolution and cross-correlation produce identical results in practical implementation, and we adopt this summation form for convenient programming deployment.

To suppress interference from uneven illumination variation, we separate the reflectance component via element-wise subtraction between the log-transformed image and illumination map:

$$R_k(x, y) = \hat{I}_k(x, y) - L_k(x, y) \tag{4}$$

where $R_k(x, y)$ is the reflectance image after illumination elimination for channel k.

Subsequently, exponential transformation is utilized to invert the logarithmic operation and further boost contrast of flame-related information, highlighting intrinsic spectral and textural features of flames:

$$\tilde{I}_k(x, y) = \exp\left(\beta \cdot R_k(x, y)\right) \tag{5}$$

where β is a tunable hyperparameter controlling enhancement intensity, and $\tilde{I}_k(x, y)$ is the reflectance map after exponential recovery. When photon saturation makes overexposed flame highlights appear as blank white areas lacking valid features, reflectance decomposition effectively recovers their original spectral attributes; subsequent contrast enhancement further sharpens flame textures and contours, improving detection robustness under dark nighttime scenes.

To dynamically adjust feature weights according to illumination distribution, we construct an illumination-adaptive mask via the sigmoid function with sensitivity hyperparameter γ:

$$M_k(x, y) = \mathrm{sigm}(\gamma \cdot L_k(x, y)) \tag{6}$$

where γ adjusts the response sensitivity of the mask to illumination variation, and $\mathrm{sigm}(\cdot)$ denotes the predefined sigmoid activation function. We multiply the reflectance-enhanced image with the illumination-aware mask pixel-wise to obtain the preliminary enhanced single-channel image for the k-th RGB channel:

$$\hat{I}_{enh,k}(x, y) = \tilde{I}_k(x, y) \cdot M_k(x, y) \tag{7}$$

Here, the subscript k denotes the index of the RGB channel, so $\hat{I}_{enh,k}(x, y)$ represents the preliminary enhanced image of the k-th single color channel calculated by Eq. (6). After processing all three RGB channels, we concatenate $\hat{I}_{enh,1}, \hat{I}_{enh,2}, \hat{I}_{enh,3}$ into a multi-channel feature tensor with C_{in} input channels, which is fed into the subsequent EAConv module.

The enhanced RGB image is subsequently forwarded to the backbone network for feature extraction. Flame features present obvious differences in spatial distribution and temporal evolution status; fixed feature extraction strategies cannot adaptively capture such dynamic variations. To solve this problem, we embed an EAConv adaptive convolution submodule after the reflectance enhancement branch inside AREM. EAConv first generates an adaptive weight map via 1×1 convolution, normalizes it through the sigmoid function to obtain a channel-spatial activation weight map, and recalibrates the enhanced flame features accordingly. This local modulation strategy adaptively revises feature representation according to real-time flame conditions at different positions, enabling effective modeling of spatiotemporal variation of flame characteristics.

For the multi-channel tensor, we use the general subscript $i \in \{1, 2, ..., C_{in}\}$ to denote arbitrary input channels, and the corresponding input feature of the i-th channel is denoted as $\hat{I}_{enh,i}(x, y)$. For EAConv calculation, we use the enhanced image $\hat{I}_{enh,i}(x, y)$ $(i \in \{1, 2, ..., C_{in}\})$ as input features. For the o-th output channel, the intermediate weight map $W_o(x, y)$ is calculated as:

$$W_o(x, y) = \sum_{i=1}^{C_{in}} w_{o,i} \cdot F_i(x, y) + b_o \tag{8}$$

where $w_{o,i}$ represents learnable kernel weights of the 1×1 convolution mapping the i-th input channel to the o-th output channel; b_o is the corresponding bias term; the summation traverses all C_{in} input channels.

We normalize the weight range via the sigmoid function to obtain normalized activation weights

$Z_o(x, y)$:

$$Z_o(x, y) = \text{sigm}(W_o(x, y)) \tag{9}$$

where $Z_o(x, y)$ is the adaptive modulation weight assigned to pixel (x, y) of the o-th output channel.

Before feature recalibration, we define $I_{\text{out},o}(x, y)$ as the unadjusted enhanced feature map of the o-th channel from the input multi-channel tensor, which serves as the original feature to be modulated by the weight map $Z_o(x, y)$. We multiply the normalized weight map with the enhanced feature map element-wise to implement feature recalibration and acquire the final adaptively adjusted output feature map:

$$I_{\text{out},o}(x, y) = Z_o(x, y) \cdot \hat{I}_{\text{enh},o}(x, y) \tag{10}$$

Substituting Eq. (9) into Eq. (10), and expanding $Z_o(x, y)$ using its explicit form derived from Eqs. (8) and (9), the complete adaptive feature redistribution process can be expressed as:

$$I_{\text{out},o}(x, y) = \text{sigm}\left(\sum_{i=1}^{C_{\text{in}}} w_{o,i} \cdot \hat{I}_{\text{enh},i}(x, y) + b_o\right) \cdot \hat{I}_{\text{enh},o}(x, y) \tag{11}$$

3.3 Self-Adaptive Feature Convolution Module

In practical EV fire scenarios, flame and smoke targets often exhibit significant scale variation. Small flames, thin smoke trails, and diffuse smoke regions may appear simultaneously in the same scene, which places higher demands on multi-scale feature representation. To address this issue, the proposed framework introduces the Self-Adaptive Feature Convolution module to improve the extraction of fire- and smoke-related patterns at different scales while maintaining lightweight computation.

As shown in Figure 3, SAConv adopts a multi-branch design to extract features with different receptive fields. After an initial channel projection, the input feature map is processed by several parallel branches, including an identity branch and three large-kernel convolution branches with kernel sizes of 9×9, 13×13, and 17×17. These branches enable the module to simultaneously capture local details and broader contextual information. Such a design is particularly suitable for fire and smoke detection, where targets may vary from compact flame edges to widely distributed smoke regions. The extracted multi-scale features are then concatenated and fused through a 1×1 convolution to generate the final output representation.

Mathematically, given an input tensor, $\mathcal{X} \in \mathbb{R}^{H\times W\times c_1}$, where H and W denote spatial dimensions and c_1 is the input channel dimensionality, the SAConv operation is formalized as a composition of an initial projection, multi-scale transformations, and a learnable fusion step:

$$\mathcal{Y} = \Phi_{\text{fuse}}\left(\bigoplus_{m=0}^{M-1} \mathcal{K}_m\left(\Pi_{\theta_0}^{c_1\to c'}(\mathcal{X})\right)\right), M = 4 \tag{12}$$

where $\mathcal{Y} \in \mathbb{R}^{H\times W\times c_2}$ is the output tensor with c_2 output channels. The components are defined as follows:

Initial Projection: The input $\mathcal{X}$ is first projected into an intermediate feature space with reduced channel dimensionality $c' = \lfloor c_2/4 \rfloor$:

$$\mathcal{X}_0 = \Pi_{\theta_0}^{c_1\to c'}(\mathcal{X}) = \sigma(\text{Conv}_{k,s,g,d}(\mathcal{X}; \theta_0)) \tag{13}$$

where, $\Pi_{\theta_0}^{c_1\to c'}(\mathcal{X}) : \mathbb{R}^{H\times W\times c_1} \to \mathbb{R}^{H\times W\times c'}$ is a parameterized projection operator with learnable weights θ_0, implemented as a convolution with kernel size k, stride s, groups g, dilation d and activation σ.

Multi-Scale Transformations: The intermediate features $\mathcal{X}_0$ are processed by $M = 4$ parallel branches, each applying a kernel-specific transformation $\mathcal{K}_m : \mathbb{R}^{H\times W\times c_1} \to \mathbb{R}^{H\times W\times c'}$:

$$\mathcal{K}_m(\mathcal{X}_0) = \begin{cases} \mathcal{X}_0, & m = 0 \\ \sigma(\text{Conv}_{k_m,1,c,d}(\mathcal{X}_0; \theta_m)), & m \in \{1, 2, 3\} \end{cases} \tag{14}$$

where $k_m \in \{9, 13, 17\}$ for $m = 1, 2, 3$ corresponding to kernel sizes of 9×9, 13×13, and 17×17, respectively, with learnable parameters θ_m.

Feature Fusion: The transformed features are concatenated along the channel dimension (Φ) and fused via a 1×1 convolution:

$$\mathcal{Y} = \Phi_{\text{fuse}}([\mathcal{K}_0(\mathcal{X}_0), \mathcal{K}_1(\mathcal{X}_0), \mathcal{K}_2(\mathcal{X}_0), \mathcal{K}_3(\mathcal{X}_0)]) \tag{15}$$

where $\Phi_{\text{fuse}} : \mathbb{R}^{H\times W\times Mc'} \to \mathbb{R}^{H\times W\times c_2}$ dynamically prioritizes fine-grained details and coarse-grained contextual information, ensuring a comprehensive representation of smoke and fire patterns.

Algorithm 1 outlines the Scale Adaptive Convolution (SAConv) module, designed to enhance the model's ability to detect subtle and dispersed fire and smoke features in challenging environments. By integrating larger receptive fields, SAConv improves feature extraction while preserving real-time inference speeds. This balance between detection accuracy

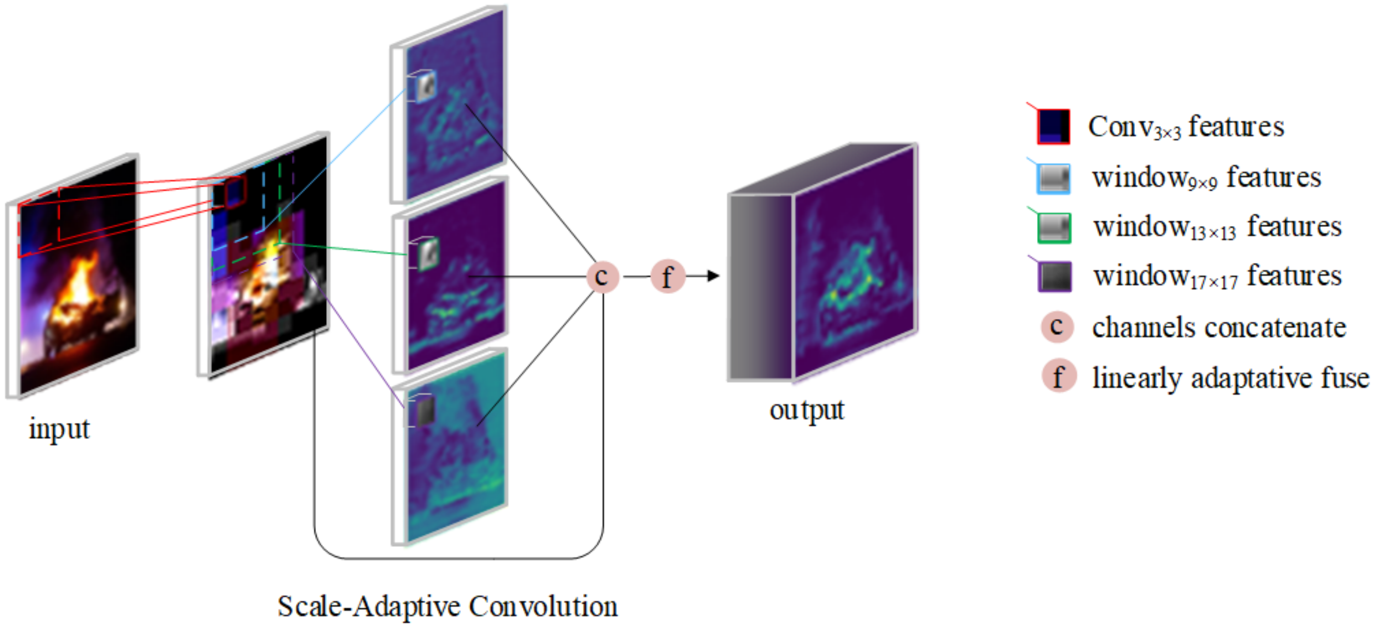


Figure 3. Illustrations of the SAConv.

and computational efficiency makes SAConv highly suitable for practical fire and smoke detection applications, as validated through experimental results.

Algorithm 1: Scale Adaptive Conv module

Initial Projection:

$$\mathcal{X}_0 = \Pi_{\theta_0}^{c_1 \to c'}(\mathcal{X}) = \sigma(\mathrm{Conv}_{k,s,g,d}(\mathcal{X};\theta_0))$$

Multi-Scale Transformation:

$$\mathcal{K}_m(\mathcal{X}_0) = \begin{cases} \mathcal{X}_0, & m = 0 \\ \sigma(\mathrm{Conv}_{k_m,1,c',d}(\mathcal{X}_0;\theta_m)), & m \in \{1,2,3\} \end{cases}$$

Feature Concatenation:

$$\mathcal{Z} = \bigoplus_{m=0}^{M-1} \mathcal{K}_m(\mathcal{X}_0) = [\mathcal{K}_0(\mathcal{X}_0), \mathcal{K}_1(\mathcal{X}_0), \mathcal{K}_2(\mathcal{X}_0), \mathcal{K}_3(\mathcal{X}_0)]$$

Feature Fusion:

$\mathcal{Y} = \Phi_{fuse}(\mathcal{Z})$, where $\Phi_{fuse} : \mathbb{R}^{H\times W\times Mc'} \to \mathbb{R}^{H\times W\times c_2}$

Output:

$\mathcal{Y} \in \mathbb{R}^{H\times W\times c_2}$

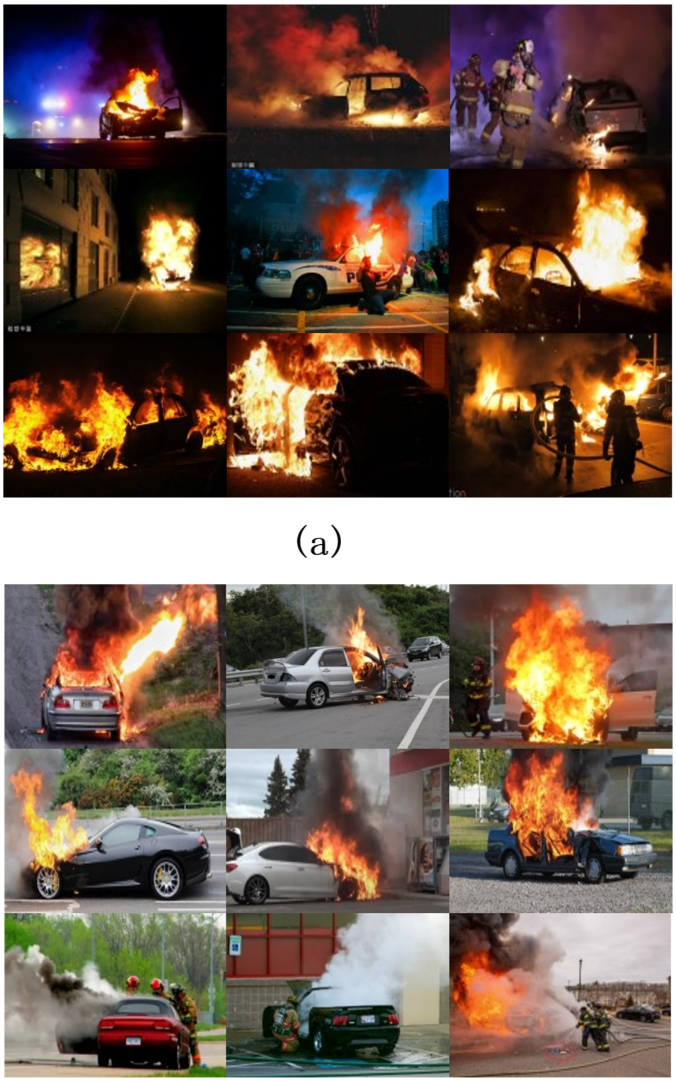
4 Experiments and Analysis

4.1 Data Sets

To evaluate the proposed method under realistic EV fire scenarios, a dedicated EV-Fire dataset was constructed for this study. The dataset contains 3,512 images collected from electric vehicle fire environments under diverse lighting conditions, including 1,910 daytime images and 1,602 nighttime images. Compared with general fire datasets, the EV-Fire dataset places greater emphasis on low-light scenes, which are more challenging because of poor visibility, flame overexposure, and complex background interference. This dataset therefore provides a more suitable benchmark for evaluating fire and smoke detection performance in EV environments. To avoid data leakage, the training and test sets were divided at the scene level, ensuring that highly similar images from the same scene did not appear in both subsets.

Figure 4 presents representative samples from the EV-Fire dataset. As shown in Figure 4 (a), nighttime images exhibit typical low-light characteristics, including weak background illumination, strong local brightness variation, and reduced target visibility. In contrast, the daytime samples shown in Figure 4 (b) provide clearer visual information and relatively more stable scene conditions. By covering both daytime and nighttime fire scenarios, the dataset supports a comprehensive evaluation of model performance under different lighting conditions.

Figure 5 shows the statistical distribution of the dataset. The training and testing sets maintain a relatively balanced distribution across daytime and nighttime scenes, which helps reduce evaluation bias caused by illumination differences. In addition, the annotation distribution of fire and smoke targets is also relatively balanced, allowing the dataset to support joint evaluation of flame and smoke detection performance.



(a)

(b)

Figure 4. Sample images from the fire dataset. (a) Nighttime vehicle fire images. (b) Daytime vehicle fire images.

Such a distribution is beneficial for improving model generalization across different fire-related patterns and scene conditions.

All images in the EV-Fire dataset were manually annotated using bounding boxes for two categories, namely fire and smoke. The annotations focus on spatial localization of fire-related targets and do not include additional metadata such as fire intensity or environmental attributes. This design is consistent with the objective of this study, which is to evaluate the effectiveness of visual detection methods for fire and smoke recognition in EV environments. Moreover, by specifically incorporating a substantial number of nighttime samples, the dataset provides an effective basis for assessing the robustness of detection models under low-light conditions.

4.2 Ablation Studies

All experiments were conducted on a Windows 11 workstation equipped with an Intel® Core™ i5-12600KF CPU, an NVIDIA GeForce RTX 4060 Ti GPU (16 GB VRAM), and 32 GB DDR4 system memory, using Python 3.10, PyTorch 1.13.0, and CUDA 11.7. The model was trained for a maximum of 100 epochs with a batch size of 32 and an input resolution of 640×640 pixels, using SGD as the optimizer with an initial learning rate of 0.01 and Mosaic augmentation disabled during the final 10 epochs to stabilize convergence. To ensure consistency and fairness across all comparative experiments, the confidence threshold was uniformly set to 0.25 and the IoU threshold to 0.7 during both validation and inference stages, with the number of data-loading workers set to 0 to guarantee stable training under the Windows environment. Detection performance was evaluated using standard COCO-style metrics, including mAP@0.5 and mAP@0.5:0.95, providing a comprehensive assessment of localization accuracy across both loose and strict IoU thresholds to ensure the credibility and reproducibility of the reported results.

To verify the effectiveness of the proposed components, ablation experiments were conducted on several lightweight YOLO baselines. The analysis focuses on three aspects: the contribution of AREM and EAConv, the effectiveness of the auxiliary structure, and the combined effect of the auxiliary structure and SAConv. The corresponding results are reported in Tables 1, 2 and 3.

Table 1 evaluates the effect of AREM and EAConv on six YOLO baselines. Overall, the proposed modules improve detection performance for most models, indicating that the feature enhancement strategy is effective in refining fire- and smoke-related representations. Taking YOLOv5-n [30] as an example, its baseline fire detection performance is 83.2% AP50 and 48.2%. When only the AREM module is embedded, the metrics rise to 84.6% AP50 and 48.7%. After integrating both AREM and EAConv modules, the fire performance is further boosted to 85.1% AP50 and 48.9%. Meanwhile, YOLOv11-n [31] improves from 83.8% AP50 and 55.0% AP to 84.8% AP50 and 56.0% AP with dual modules. Similar consistent performance gains on both smoke and fire tasks can also be observed in YOLOv6-n [32] and YOLOv7-t [33]. For YOLOv8-n [34], the improvement is more evident

Figure 5. Statistics of the fire dataset. (a) Distribution of daytime and nighttime fire dataset. (b) Label distribution of the fire dataset.

Table 1. Performance of AREM and EAConv at six baselines.

Method	AREM	EAConv	Smoke		Fire	
			AP50 (%)	AP (%)	AP50(%)	AP(%)
YOLOv5-n			79.2	38.8	83.2	48.2
	✓		79.1	38.7	84.6	48.7
	✓	✓	**79.1**	**38.9**	**85.1**	**48.9**
YOLOv6-n			80.0	47.3	81.8	47.9
	✓		80.8	47.2	81.9	47.5
	✓	✓	**81.0**	**47.1**	**83.1**	**48.1**
YOLOv7-t			74.5	33.4	82.4	49.5
	✓		74.1	33.5	83.4	50.1
	✓	✓	**74.0**	**34.0**	**83.5**	**50.8**
YOLOv8-n			78.9	45.0	82.7	55.1
	✓		82.1	46.1	84.1	54.5
	✓	✓	**82.3**	**46.7**	**83.3**	**55.1**
YOLOv10-n			81.2	44.6	81.1	50.7
	✓		81.3	45.5	80.9	50.6
	✓	✓	**81.6**	**45.8**	**80.5**	**50.8**
YOLOv11-n			84.4	45.9	83.8	55.0
	✓		84.3	45.4	84.6	55.7
	✓	✓	**84.2**	**46.1**	**84.8**	**56.0**

in smoke detection: smoke AP50 rises from 78.9% to 82.3% and smoke AP increases from 45.0% to 46.7% when both modules are applied. In contrast, the gain for fire detection is less consistent. The fire AP drops slightly to 54.5% when only AREM is used, and returns to the baseline value of 55.1% after EAConv is introduced. Meanwhile, fire AP50 first improves to 84.1% with AREM alone, then decreases to 83.3% when EAConv is further added, indicating that the effectiveness of EAConv may vary across different backbone architectures. By comparison, YOLOv10-n [35] shows a slight decline in fire AP50 as modules are added progressively, indicating that the effectiveness of the proposed enhancement strategy remains partly dependent on the underlying detector architecture.

Table 2 further examines the contribution of the auxiliary structure as a whole. The results show that the auxiliary branch consistently improves detection accuracy on most baselines while introducing only limited computational overhead. For example, YOLOv5-n-au achieves 85.1% AP50 and 48.9% AP, corresponding to gains of 1.9% and 0.7%, respectively,

Table 2. Ablation study on the auxiliary structure for the YOLO series.

Models		Parameters number (M)	AP50(%)-fire	AP(%)-fire	Latency b=32(ms)	FLOPs (G)
Baselines	YOLOv5-n	1.76	83.2	48.2	9.0	4.1
	YOLOv6-n	4.3	81.8	47.9	9.5	11.1
	YOLOv7-t	6.0	82.4	49.5	7.4	13.2
	YOLOv8-n	3.1	82.7	55.1	2.1	8.2
	YOLOv10-n	2.6	81.0	50.7	2.5	8.2
	YOLOv11-n	2.5	83.8	55.0	1.5	6.3
Baselines +auxiliary structure	YOLOv5-n-au	1.9	85.1(+1.9)	48.9(+0.7)	8.5	4.3
	YOLOv6-n-au	4.6	83.1(+1.9)	48.1(+0.2)	10.1	10.9
	YOLOv7-t-au	6.2	83.5(+1.1)	50.8(+1.3)	8.2	13.0
	YOLOv8-n-au	3.01	83.3(+0.6)	55.1(+0.0)	1.4	8.1
	YOLOv10-n-au	2.72	80.5(-0.5)	50.8(+0.1)	2.3	8.1
	YOLOv11-n-au	2.52	84.8(+1.0)	56.0(+1.0)	1.8	6.3

while YOLOv11-n-au achieves 84.8% AP50 and 56.0% AP, improving both AP50 and AP by 1.0%. At the same time, the changes in FLOPs and latency remain relatively small, suggesting that the auxiliary structure improves feature representation without substantially sacrificing efficiency. Nevertheless, the improvement remains limited for YOLOv8-n and slightly negative for YOLOv10-n in AP50, again showing that the proposed structure is more effective for some architectures than for others. Even so, the overall trend across the tested baselines supports the effectiveness of the auxiliary branch in enhancing fire-related feature extraction.

Table 3 analyzes the individual and combined contributions of SAConv and the auxiliary structure. When SAConv is introduced alone, most models achieve higher AP50 and AP, indicating that multi-scale feature extraction is beneficial for fire and smoke detection, especially for small flames and diffuse smoke. For example, YOLOv5-n-saconv improves to 83.7% AP50 and 47.0% AP, and YOLOv11-n-saconv reaches 85.0% AP50 and 51.2% AP. When only the auxiliary structure is applied, moderate but consistent gains are observed in most models. When both SAConv and the auxiliary structure are combined, the best overall performance is obtained on YOLOv11-n-saconv-au, which reaches 85.3% AP50 and 52.2% AP, corresponding to improvements of 1.2% and 1.7% over the baseline. This result indicates that the two components are complementary: the auxiliary structure mainly enhances weak fire-related features under complex illumination, while SAConv improves multi-scale representation capability. However, the combined strategy still does not improve all architectures equally, as YOLOv10-n-saconv-au shows a slight decline. Therefore, the ablation results suggest that the proposed design is generally effective, but its performance gain remains partly architecture-dependent.

Overall, the ablation experiments demonstrate that the proposed modules improve detection performance from different perspectives. AREM and EAConv mainly strengthen weak feature representation and suppress background interference, whereas SAConv enhances the modeling of fire and smoke targets at different scales. Their combination yields the best overall result on the strongest baseline, confirming that the proposed framework can improve detection accuracy while preserving lightweight characteristics.

4.3 Comparison Experiments

To further evaluate the effectiveness of the proposed method, comparison experiments were conducted under two settings: the full EV-Fire dataset containing both daytime and nighttime scenes, and the nighttime subset focusing on low-light conditions. This design allows the proposed method to be assessed not only in terms of overall detection performance, but also in terms of robustness under challenging illumination conditions. The corresponding results are reported in Tables 4 and 5. All latency results were measured under the same hardware and software environment using the same input resolution, inference setting, and batch size to ensure fair comparison.

Table 4 presents the comparison results on the full dataset. Among the lightweight YOLO baselines, YOLOv11-n achieves the best overall balance between accuracy and efficiency, reaching 84.1% AP50 and 50.5% AP with a latency of 1.5 ms. EFA-YOLO also shows strong efficiency, with only 1.9G FLOPs

Table 3. Ablation experiments of the SAConv and auxiliary structure.

Models		Param. (M)	FLOPs (G)	AP50(%)	AP(%)	Latency b=32(ms)
baseline	YOLOv5-n	1.7	4.1	81.2	43.6	9.0
	YOLOv6-n	4.6	11.3	81.1	47.6	4.2
	YOLOv7-t	6.3	13.2	78.4	41.4	7.4
	YOLOv8-n	3.0	8.2	80.8	50.1	2.1
	YOLOv10-n	2.6	8.2	81.1	47.7	2.5
	YOLOv11-n	2.5	6.3	84.1	50.5	1.5
Part I: SAConv (ours)	YOLOv5-n- saconv	1.6	4.0	83.7(+2.5)	47.0(+3.4)	10.4
	YOLOv6-n- saconv	4.5	11.0	82.1(+1.0)	48.0(+0.4)	4.5
	YOLOv7-t- saconv	6.3	13.5	80.6(+2.2)	42.1(+0.7)	8.6
	YOLOv8-n- saconv	2.9	8.0	83.8(+3.0)	50.9(+0.8)	4.1
	YOLOv10-n- saconv	2.7	8.1	79.6(-1.5)	48.0(+0.3)	2.6
	YOLOv11-n-saconv	2.5	6.2	85.0(+0.9)	51.2(+0.7)	2.0
Part II: auxiliary structure (ours)	YOLOv5-n-au	1.9	4.3	82.1(+0.9)	44.0(+0.4)	8.7
	YOLOv6-n-au	4.6	10.9	82.2(+1.1)	47.6(+0.0)	4.0
	YOLOv7-t-au	6.2	13.0	78.8(+0.4)	42.4(+1.0)	6.9
	YOLOv8-n-au	3.0	8.1	82.8(+2.0)	50.9(+0.8)	1.7
	YOLOv10-n-au	2.7	8.1	81.0(-0.1)	48.3(+0.6)	2.9
	YOLOv11-n-au	2.5	6.3	84.5(+0.4)	51.1(+0.6)	2.0
Part III: SAConv + auxiliary structure (ours)	YOLOv5-n- saconv-au	1.6	4.0	83.7(+2.5)	44.1(+0.5)	8.0
	YOLOv6-n- saconv-au	4.6	10.5	82.2(+1.1)	47.6(+0.0)	3.5
	YOLOv7-t- saconv-au	6.3	12.8	80.5(+2.1)	42.5(+1.1)	6.5
	YOLOv8-n- saconv-au	3.0	8.1	82.2(+1.4)	51.0(+0.9)	2.0
	YOLOv10-n-saconv-au	2.6	8.4	80.9(-0.2)	47.3(-0.4)	2.5
	YOLOv11-n-saconv-au	2.5	6.3	85.3(+1.2)	52.2(+1.7)	2.0

Table 4. Performance comparison of State-of-the-Art real-time detectors on the full "fire dataset".

Models	Year	Param. (M)	FLOPs (G)	AP50 (%)	AP (%)	Latency b=32(ms)
SSD512 [39]	2016	52.3	139.7	76.9	40.6	23.5
Faster R-CNN [38]	2017	41.0	180.1	80.4	49.7	41.3
DETR [37]	2020	41.3	86.4	81.6	46.4	55.6
YOLOv5-n [30]	2022	1.7	4.1	81.2	43.6	9.0
YOLOv6-n [32]	2022	4.6	11.3	81.1	47.6	4.2
YOLOv7-t [33]	2023	6.3	13.2	78.4	41.4	7.4
YOLOv8-n [34]	2024	3.0	8.2	80.8	50.1	2.1
YOLOv10-n [35]	2024	2.6	8.2	81.1	47.7	2.5
YOLOv11-n [31]	2024	2.5	6.3	84.1	50.5	1.5
EFA-YOLO [25]	2024	2.1	1.9	83.4	51.1	1.2
Enhanced YOLOX [36]	2024	3.7	11.3	82.8	43.1	7.4
Proposed Method	2024	2.5	6.3	85.3	52.2	1.8

and 1.2 ms latency, while achieving 83.4% AP50 and 51.1% AP. Meanwhile, Enhanced YOLOX, another 2024 model, achieves 82.8% AP50 and 43.1% AP, but with a higher computational cost (11.3G FLOPs) and latency of 7.4 ms, making it less ideal for speed-sensitive applications. In comparison, the proposed method achieves the best overall detection accuracy, reaching 85.3% AP50 and 52.2% AP with 6.3G FLOPs and 1.8 ms latency. Compared with YOLOv11-n, the proposed method improves AP50 by 1.2% and AP by 1.7%, while maintaining a lightweight computational cost. These results indicate that the proposed framework

Figure 6. Qualitative comparison of detection results among different models under daytime and nighttime scenarios.

improves feature representation without introducing a substantial efficiency penalty.

Although the improvement over the strongest baseline is moderate, it should be noted that YOLOv11-n already provides a relatively high-performance baseline. Under such conditions, further gains become more difficult to obtain. Therefore, the observed improvements suggest that the proposed framework is effective in refining discriminative fire-related features, rather than merely benefiting from a weaker baseline.

To complement the quantitative results in Table 4, Figure 6 provides a qualitative comparison under both daytime and nighttime scenarios. In daytime scenes, most compared methods are able to detect large and visually obvious flame regions. However, the proposed method produces tighter bounding boxes and more reliable confidence scores, indicating better localization quality. These qualitative observations are consistent with the quantitative gains reported on the full dataset and show that the proposed framework improves not only numerical performance but also prediction stability in representative EV fire scenes.

Table 5 further evaluates detection performance on the nighttime subset, where low illumination, background interference, and flame overexposure make the task more challenging. Under this setting, the performance gap between models becomes more evident. YOLOv7-t performs the worst, with only 14.7% AP50 and 3.83% AP, while YOLOv6-n and YOLOv10-n also show limited adaptability to nighttime fire detection. YOLOv11-n achieves 52.3% AP50 and 24.9% AP, making it the strongest YOLO baseline on the nighttime subset. The proposed method further improves the performance to 52.9% AP50 and 26.5% AP, with the same FLOPs as YOLOv11-n. Although the AP50 improvement is only 0.6%, the AP gain reaches 1.6%, indicating that the proposed framework provides more stable localization and ranking performance under low-light conditions.

Figure 7 provides the corresponding qualitative comparison on representative nighttime samples. Under low-light conditions, several baseline models show insufficient flame coverage, incomplete smoke detection, or unstable prediction results. In contrast, the proposed method is able to simultaneously detect flame and smoke targets more reliably. It shows better stability in both flame localization and smoke representation, especially in scenes with long viewing distance, low visibility, and strong background

Figure 7. Qualitative comparison of detection results on representative nighttime fire scenes.

Table 5. Performance comparison of State-of-the-Art real-time detectors on the "nighttime dataset".

Models	Year	Param. (M)	FLOPs (G)	AP50 (%)	AP (%)	Latency b=32(ms)
SSD300 [39]	2016	5.5	31.0	38.5	17.8	12.6
Faster R-CNN [38]	2017	41.0	180.1	54.2	26.1	85.2
DETR [37]	2020	41.3	86.1	46.5	21.2	68.0
YOLOv5-n [30]	2022	1.7	4.1	44.0	21.3	9.0
YOLOv6-n [32]	2022	4.6	11.3	32.9	13.4	5.1
YOLOv7-t [33]	2023	6.3	13.2	14.7	3.83	7.4
YOLOv8-n [34]	2024	3.0	8.2	46.4	21.3	2.1
YOLOv10-n [35]	2024	2.6	8.2	38.8	17.6	3.0
YOLOv11-n [31]	2024	2.5	6.3	52.3	24.9	2.3
Proposed Method	2024	2.5	6.3	52.9	26.5	3.0

interference. These observations further support the conclusion that the proposed framework is particularly effective in enhancing weak fire-related features under nighttime conditions.

Overall, the comparison experiments demonstrate that the proposed method achieves the best overall detection accuracy on the full dataset and maintains superior adaptability on the nighttime subset. Combined with the qualitative results shown in Figures 6 and 7, these findings show that the proposed framework improves both quantitative performance and practical robustness, making it more suitable for real-time fire and smoke detection in complex EV environments.

5 Conclusions

This study proposed a lightweight dual-branch framework for fire and smoke detection in electric vehicle environments, with a particular focus on low-light conditions. To improve feature representation under weak illumination and complex backgrounds, the proposed method integrates the Adaptive Reflectance Enhancement Module (AREM), the Element-wise Adaptive Weight Convolution (EAConv), and the Self-Adaptive Feature Convolution (SAConv). In addition, an EV-Fire dataset containing 3,512 images, including 1,910 daytime and 1,602 nighttime samples, was constructed and publicly released to support evaluation under diverse lighting conditions. Experimental results show that the proposed method achieves 85.3% AP50 and 52.2% AP on the full dataset, and 52.9% AP50 and 26.5% AP on the nighttime subset, demonstrating strong overall performance and good adaptability to low-light scenarios. The ablation results further verify that the proposed modules improve weak feature representation, suppress background interference, and enhance multi-scale feature extraction.

Overall, the proposed framework provides an effective and practical solution for real-time fire and smoke detection in electric vehicle environments. However, the performance gain of the proposed framework remains partly architecture-dependent, as slight performance degradation is observed on certain detector variants. Future work will focus on improving robustness under more complex conditions, such as dense smoke, severe occlusion, and strong artificial illumination, while further enhancing the generalization of the proposed design across different detector architectures.

Data Availability Statement

The dataset supporting the findings of this study is publicly available in the GitHub repository at https://github.com/Tx1101/EV-Fire.

Funding

This work was supported by the Natural Science Research Project of Jiangsu Higher Education Institutions of China under Grant 24KJA510002 and by the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant SJCX25_2194.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Foggia, P., Saggese, A., & Vento, M. (2015). Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion. *IEEE TRANSACTIONS on circuits and systems for video technology*, *25*(9), 1545-1556. [CrossRef]

[2] Reis, H. C., & Turk, V. (2023). Detection of forest fire using deep convolutional neural networks with transfer learning approach. *Applied Soft Computing*, *143*, 110362. [CrossRef]

[3] Jiang, M., Zhao, Y., Yu, F., Zhou, C., & Peng, T. (2022). A self-attention network for smoke detection. *Fire safety journal*, *129*, 103547. [CrossRef]

[4] Zhao, L., Zhi, L., Zhao, C., & Zheng, W. (2022). Fire-YOLO: A small target object detection method for fire inspection. *Sustainability*, *14*(9), 4930. [CrossRef]

[5] Li, Y., Zhang, W., Liu, Y., Jing, R., & Liu, C. (2022). An efficient fire and smoke detection algorithm based on an end-to-end structured network. *Engineering Applications of Artificial Intelligence*, *116*, 105492. [CrossRef]

[6] Wang, G., Li, H., Xiao, Q., Yu, P., Ding, Z., Wang, Z., & Xie, S. (2025). Fighting against forest fire: A lightweight real-time detection approach for forest fire based on synthetic images. *Expert Systems with Applications*, *262*, 125620. [CrossRef]

[7] Xue, Z., Lin, H., & Wang, F. (2022). A small target forest fire detection model based on YOLOv5 improvement. *Forests*, *13*(8), 1332. [CrossRef]

[8] Gao, D., Zhang, S., Ju, Y., & Yang, Q. (2023). A real-time fire and flame detection method for electric vehicle charging station based on machine vision. *Journal of Real-Time Image Processing*, *20*(2), 38. [CrossRef]

[9] Yar, H., Ullah, W., Khan, Z. A., & Baik, S. W. (2023). An effective attention-based CNN model for fire detection in adverse weather conditions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *206*, 335-346. [CrossRef]

[10] Huang, L., Liu, G., Wang, Y., Yuan, H., & Chen, T. (2022). Fire detection in video surveillances using convolutional neural networks and wavelet transform. *Engineering Applications of Artificial Intelligence*, *110*, 104737. [CrossRef]

[11] Rehman, A., Kim, D., & Paul, A. (2023). Convolutional neural network model for fire detection in real-time environment. *Computers, Materials and Continua*, *77*(2), 2289-2307. [CrossRef]

[12] Zhao, S., Liu, B., Chi, Z., Li, T., & Li, S. (2022). Characteristics based fire detection system under the effect of electric fields with improved Yolo-v4 and ViBe. *IEEE Access*, *10*, 81899-81909. [CrossRef]

[13] Zhao, H., Jin, J., Liu, Y., Guo, Y., & Shen, Y. (2024). FSDF: A high-performance fire detection framework. *Expert Systems with Applications*, *238*, 121665. [CrossRef]

[14] Xu, H., Li, B., & Zhong, F. (2022). Light-YOLOv5: A lightweight algorithm for improved YOLOv5 in

complex fire scenarios. *Applied Sciences*, *12*(23), 12312. [CrossRef]

[15] Jadon, A., Varshney, A., & Ansari, M. S. (2020). Low-complexity high-performance deep learning model for real-time low-cost embedded fire detection systems. *Procedia Computer Science*, *171*, 418-426. [CrossRef]

[16] Chino, D. Y., Avalhais, L. P., Rodrigues, J. F., & Traina, A. J. (2015, August). Bowfire: detection of fire in still images by integrating pixel color and texture analysis. In *2015 28th SIBGRAPI conference on graphics, patterns and images* (pp. 95-102). IEEE. [CrossRef]

[17] Gragnaniello, D., Greco, A., Sansone, C., & Vento, B. (2024). Fire and smoke detection from videos: A literature review under a novel taxonomy. *Expert Systems with Applications*, *255*, 124783. [CrossRef]

[18] Huo, Y., Zhang, Q., Zhang, Y., Zhu, J., & Wang, J. (2022). 3DVSD: An end-to-end 3D convolutional object detection network for video smoke detection. *Fire safety journal*, *134*, 103690. [CrossRef]

[19] Wang, S., Wu, M., Wei, X., Song, X., Wang, Q., Jiang, Y., ... & Lv, J. (2025). An advanced multi-source data fusion method utilizing deep learning techniques for fire detection. *Engineering Applications of Artificial Intelligence*, *142*, 109902. [CrossRef]

[20] de Venancio, P. V. A., Lisboa, A. C., & Barbosa, A. V. (2022). An automatic fire detection system based on deep convolutional neural networks for low-power, resource-constrained devices. *Neural Computing and Applications*, *34*(18), 15349-15368. [CrossRef]

[21] Ko, B., Cheong, K. H., & Nam, J. Y. (2010). Early fire detection algorithm based on irregular patterns of flames and hierarchical Bayesian Networks. *Fire safety journal*, *45*(4), 262-270. [CrossRef]

[22] T5 Capstone. (2024). *Fire Detection Dataset* [DS/OL]. Roboflow Universe. Retrieved from https://universe.roboflow.com/t5-capstone/fire-detection-9aome

[23] Chaoxia, C., Shang, W., & Zhang, F. (2020). Information-guided flame detection based on faster R-CNN. *IEEE Access*, *8*, 58923-58932. [CrossRef]

[24] Barmpoutis, P., Dimitropoulos, K., Kaza, K., & Grammalidis, N. (2019, May). Fire detection from images using faster R-CNN and multidimensional texture analysis. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8301-8305). IEEE. [CrossRef]

[25] Pan, W., Wang, X., & Huan, W. (2024). EFA-YOLO: An efficient feature attention model for fire and flame detection. *arXiv preprint arXiv:2409.12635*. [CrossRef]

[26] Park, M., & Ko, B. C. (2020). Two-step real-time night-time fire detection in an urban environment using static ELASTIC-YOLOv3 and temporal fire-tube. *Sensors*, *20*(8), 2202. [CrossRef]

[27] Zheng, H., Wang, G., Xiao, D., Liu, H., & Hu, X. (2024). FTA-DETR: An efficient and precise fire detection framework based on an end-to-end architecture applicable to embedded platforms. *Expert Systems with Applications*, *248*, 123394. [CrossRef]

[28] Li, H. (2024, September). Rethinking features-fused-pyramid-neck for object detection. In *European conference on computer vision* (pp. 74-90). Cham: Springer Nature Switzerland. [CrossRef]

[29] Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep Retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*. [CrossRef]

[30] Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Fang, J., ... & Thanh Minh, M. (2022). ultralytics/yolov5: v6. 1-tensorrt, tensorflow edge tpu and openvino export and inference. *Zenodo*. [CrossRef]

[31] Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*. [CrossRef]

[32] Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., ... & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*. [CrossRef]

[33] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023, June). 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). IEEE. [CrossRef]

[34] Varghese, R., & Sambath, M. (2024, April). Yolov8: A novel object detection algorithm with enhanced performance and robustness. In *2024 International conference on advances in data engineering and intelligent computing systems (ADICS)* (pp. 1-6). IEEE. [CrossRef]

[35] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. *Advances in neural information processing systems*, *37*, 107984-108011. [CrossRef]

[36] Wang, T., Wang, J., Wang, C., Lei, Y., Cao, R., & Wang, L. (2024). Improving YOLOX network for multi-scale fire detection. *The Visual Computer*, *40*(9), 6493-6505. [CrossRef]

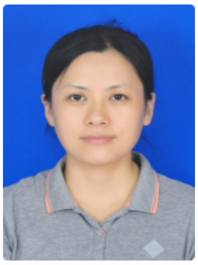
[37] Carion, N., Massa, F., Synnaert, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. In *European conference on computer vision (ECCV)* (pp. 213-229). Springer, Cham. [CrossRef]

[38] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, *39*(6), 1137-1149. [CrossRef]

[39] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In *European conference on computer vision (ECCV)* (pp. 21-37). Springer, Cham. [CrossRef]

Xiu Tang is an M.Eng. candidate. She has published one EI-indexed paper and holds three published invention patents. She has received multiple honors and awards, including the Third Prize in the 16th Lanqiao Cup EDA Provincial Competition, the Second Prize in the 15th Lanqiao Cup Embedded Systems Provincial Competition, and has presided over one provincial-level scientific research project. She has also received the titles of Merit Student and Outstanding Graduate Student, as well as the National Scholarship, one First-Class Scholarship, and two Second-Class Scholarships. (Email: Tangxiu_1110@163.com)

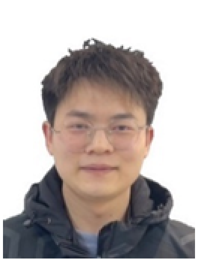
Yaxin Man, M.Eng., holds a Master's degree in Engineering. His academic training focuses on transportation, traffic information and control, and related fields, with research interests in deep learning-based image processing and intelligent transportation systems. During his postgraduate studies, he actively participated in scientific research projects and academic activities, demonstrating solid professional knowledge and practical skills. (Email: 15866676685@163.com)



Xiaoyan Wang, Ph.D. in Engineering, serves as a Jiangsu Provincial Science and Technology Associate, an Excellent Graduation Design Team Instructor of Jiangsu Province, and a leading talent of the "533 Talent Program" in Huai'an. Her research has long focused on photoelectric detection, particularly in spectral detection and analysis as well as signal and information processing. She has presided over several research projects, including those funded by the National Natural Science Foundation of China, the Ministry of Science and Technology's Spark Program, Jiangsu Province Industry–University–Research Cooperation, and major projects of Jiangsu universities, and has participated in more than 10 provincial or ministerial-level projects. She has published over 20 research papers, been granted 2 invention patents and 8 software copyrights, and contributed to one textbook and one monograph. (Email: wxygxy@163.com)

Rendong Ji, Ph.D. in Engineering from Nanjing University of Aeronautics and Astronautics, is a Professor at Huaiyin Institute of Technology. He concurrently serves as a member of the Optical Testing Professional Committee of the Chinese Optical Society and holds other academic positions. He has been recognized as a leading talent in Huai'an City's "533 Talent Program."He has presided over more than 10 national and provincial-level research projects and over 20 industry-funded projects. He has published more than 20 SCI/EI-indexed papers and has been granted 6 invention patents. He has also led one teaching reform project funded by the Ministry of Education and published one monograph and one textbook. (Email: jrdgxy@163.com)

Xiaojun Zhang is currently pursuing the M.S. degree in Transportation Engineering at Huaiyin Institute of Technology, Huai'an, China. His research interests include deep learning and road crack image detection. (Email: zxxxxj8800@163.com)



Yunlong Xu received the M.S. degree in transportation from Huaiyin Institute of Technology. He is currently engaged in research in the field of deep learning. (Email: yunlongxv293@gmail.com)