ARTICLE

An Improved YOLOv5s-Based Smoke Detection System for Outdoor Parking Lots

Ruobing Zuo[1], Xiaohan Huang[1], Xuguo Jiao[2,3] and Zhenyong Zhang[1,4,5,*]

[1]State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang, 550025, China

[2]School of Information and Control Engineering, Qingdao University of Technology, Qingdao, 266520, China

[3]State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China

[4]Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Qilu University of Technology (Shandong Academy of Sciences), Jinan, 250353, China

[5]Text Computing & Cognitive Intelligence Engineering Research Center of National Education Ministry, Guizhou University, Guiyang, 550025, China

*Corresponding Author: Zhenyong Zhang. Email: zyzhangnew@gmail.com

ABSTRACT

In the rapidly evolving urban landscape, outdoor parking lots have become an indispensable part of the city's transportation system. The growth of parking lots has raised the likelihood of spontaneous vehicle combustion, a significant safety hazard, making smoke detection an essential preventative step. However, the complex environment of outdoor parking lots presents additional challenges for smoke detection, which necessitates the development of more advanced and reliable smoke detection technologies. This paper addresses this concern and presents a novel smoke detection technique designed for the demanding environment of outdoor parking lots. First, we develop a novel dataset to fill the gap, as there is a lack of publicly available data. This dataset encompasses a wide range of smoke and fire scenarios, enhanced with data augmentation to ensure robustness against diverse outdoor conditions. Second, we utilize an optimized YOLOv5s model, integrated with the Squeeze-and-Excitation Network (SENet) attention mechanism, to significantly improve detection accuracy while maintaining real-time processing capabilities. Third, this paper implements an outdoor smoke detection system that is capable of accurately localizing and alerting in real time, enhancing the effectiveness and reliability of emergency response. Experiments show that the system has a high accuracy in terms of detecting smoke incidents in outdoor scenarios.

KEYWORDS

YOLOv5s; smoke detection; deep learning; SENet

1 Introduction

With the accelerated pace of urbanization, automobiles have become an indispensable mode of transportation in daily life [1,2]. However, the increase in vehicle usage has led to a corresponding

rise in incidents of spontaneous vehicle combustion, posing significant threats to public safety and challenging the security of communities [3]. The consequences of such incidents can be particularly severe in outdoor parking lots, characterized by their exposed environment and high vehicle density.

Spontaneous vehicle combustion, like other forms of public space fires, are distinguished by their sudden onset and potential to trigger cascading effects. In the initial stage of a fire, smoke often serves as one of the most important indicators, making timely detection an urgent issue [4]. Effective smoke detection not only provides early warnings within the critical "golden minutes" of a fire outbreak but also secures valuable time for emergency response, thereby minimizing casualties and property damage.

Conventional smoke detection methods typically rely on physical sensors [5–9]. Smoke sensors perform well in enclosed or semi-enclosed environments, but may not be effective in outdoor environments [10]. In outdoor environments like spacious outdoor parking lots, it is easy to be affected by environmental factors like natural light changes, dust, wind speed, humidity, or other meteorological conditions, frequently leading to high false alarm rates or failure to send out timely alerts at critical times.

Compared to sensors, image-based smoke detection is less affected by changes in the outdoor environment [11]. With the development of computer vision and image processing technology, image-based smoke detection methods began to gain traction.

Most of the early image-based smoke detection methods use conventional image processing techniques to manually extract smoke features, such as optical flow method [12], wavelet transform [13], segmentation methods like frame differential method [14,15] and Gaussian mixture model [16]. These methods rely on handcrafted feature extraction and then use support vector machines (SVM) [17], Ada Boost [18] or neural networks [19] for classification. However, these early methods have their limitations. Smoke features such as texture, color and motion are easily affected by environmental changes, which leads to low robustness of these methods in smoke detection under different environments, and high false positive rate and false negative rate may exist in some scenarios.

In recent years, the rapid development of deep learning has made deep learning algorithms powerful tool to address these challenges [20–22]. Deep learning has a great capacity in feature learning [23], which has been fully applied in the field of object detection. Compared to methods rely on handcrafted feature extraction, deep learning methods like Convolutional Neural Networks (CNNs) [24] can adapt more flexibly to the complexities and variabilities of outdoor environments, thereby significantly enhancing the accuracy and reliability of detection [25]. Furthermore, these technologies enable real-time monitoring and dynamic tracking of smoke, greatly improving response times to initial fire outbreaks.

Despite the immense potential of deep learning-based methods in outdoor smoke detection, several challenges must be overcome for successful real-world application, such as enhancing model robustness under varying lighting and weather conditions, and optimizing algorithms to reduce the computational complexity. Most deep learning methods require extensive and high-quality datasets, and network architectures need to be designed appropriately for different tasks. However, there is currently a lack of public data sets on images of smoke and fire in parking lots.

To tackle these issues, this paper proposes an improved outdoor parking lot smoke detection system based on YOLOv5s. Our work focuses not only on enhancing the accuracy and efficiency of detection technology but also on adapting the system to the change of outdoor environments, thus

providing more reliable technological support for public safety. The main contributions of this paper can be summarized as follows:

1. We create a dataset of fire incidence in outdoor parking lots. A diverse collection of smoke and fire data is gathered through various means and subsequently expanded using data augmentation techniques. Using this approach, the robustness and generality of smoke detection model are greatly improved, allowing for more effective detection performance in a variety of outdoor conditions.
2. We employ the lightweight YOLOv5s as the principal detection model, innovatively integrated with the Squeeze-and-Excitation Network (SENet) [26] for optimization. This approach not only improves the model's accuracy in recognizing smoke but also ensures high efficiency in real-time detection.
3. We implement a smoke detection and early warning system for outdoor parking lots. This system is capable of real-time detection and precise localization of smoke and fires within parking areas, alerting users in a timely manner.

The remainder of this paper is organized as follows: It summarizes related work in Section 2. The details of the improved YOLOv5s and the underlying methods are described in Section 3. In Section 4, the new dataset is introduced. Section 5 analyzes the experimental environment and results. In Section 6, we describe the implementation of the system. Finally, conclusions are drawn in Section 7.

2 Related Work

In the realm of image-based smoke detection research, main methods can be classified as: (1) traditional image-based methods, which rely on handcrafted feature extraction, (2) deep learning-based methods, which using deep learning algorithm to automatically extract features.

2.1 Traditional Image-Based Smoke Detection

Traditional image-based smoke detection methods rely on image processing technologies to manually extract smoke features such as color, texture, motion. Chen et al. [27] used a RGB model and disorder measurement to extract smoke pixels. Qin et al. [14] used color moments, LBP and HOG to represent smoke features, and use SVM as a classifier. Islam et al. [16] used the Gaussian's Mixture Model (GMM) and HSV (hue-saturation-value) color segmentation to perform smoke segmentation in the pretreatment stage, and dynamic smoke growth analysis was performed. Ajith et al. [28] made use of motion information from fire video frames. Gubbi et al. [13] proposed a method for smoke characterization by using wavelets and SVM. Dimitropoulos et al. [29] proposed a high-order decomposition method of multidimensional image data to analyze the dynamic texture of smoke for smoke detection.

While these methods have achieved certain success, they still face challenges in accurately recognizing smoke and resisting interference in complex and variable outdoor environments.

2.2 Deep Learning-Based Smoke Detection

Deep learning-based object detection can be divided into two-stage algorithms represented by RCNN [30] and one-stage algorithms represented by YOLO [31]. Compared to two-stage algorithms, one-stage algorithms have better real-time performance and are more suitable for real-time smoke detection. Luo et al. [32] proposed a smoke detection algorithm based on the convolutional neural networks (CNN). Mahmoud et al. [33] used CNN and transfer learning developed a time-efficient

fire detection system. Yin et al. [34] proposed a deep normalization and deep convolutional neural networks (DCNN) for smoke detection. Saponara et al. [4] presented a smoke detection method which use YOLOv2 CNN in real-time surveillance systems. Lin et al. [35] developed a smoke detection framework based on a combination of 3D CNN and RCNN. Zeng et al. [36] used Faster R-CNN, SSD, R-FCN to detect smoke. Zhang et al. [37] used Faster R-CNN to detect wildland forest fire smoke.

While existing deep learning-based methods theoretically offer better adaptability and accuracy, they still confront practical challenges in outdoor applications, particularly under varying lighting and weather conditions in outdoor parking lots. Given these limitations of existing methods, this paper aims to propose a more accurate, responsive, and low false alarm rate solution for outdoor parking lot smoke detection, offering a new perspective and technical pathway to address this public safety challenge.

3 Methodology

In this section, we describe the methodology employed to smoke and fire detection in outdoor parking lots. The structure of YOLOv5s model is explained in Section 3.1. Section 3.2 then goes into detail about how the model is optimized by incorporating SE attention mechanisms.

3.1 YOLOv5s Network Structure

YOLOv5 is the fifth iteration of YOLO algorithm. It is a one-stage algorithm based on regression and has been widely used in the domain of object detection. Structured with a network of convolutional layers, pooling layers, and fully connected layers, it deeply extracts image features to predict the location and classification of targets with precision. To realize smoke detection in outdoor parking lots, the real-time performance is highly required. Compared to two-stage algorithms such as R-CNN and Fast R-CNN which first detect the target region and then classify the object, YOLOv5 achieve end-to-end detection and use a single network to complete the target region detection and classification, providing faster inference speed and better adaptability to real-time requirements. This makes it particularly suitable for implementing real-time smoke detection in the outdoor parking lot environments. The YOLOv5 framework introduces multiple model variants, namely YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5n, and YOLOv5x. The YOLOv5m, YOLOv5l and YOLOv5x models deepen the network based on YOLOv5s, which improves accuracy but also incurs increasing computational overhead. Conversely, the YOLOv5n model achieves relatively lower accuracy. In comparison to the other models, the compact and computationally efficient model YOLOv5s demonstrates faster detection speed while maintaining satisfactory accuracy levels.

Considering specific constraints of the algorithm's application environment, this paper primarily utilizes the lightweight YOLOv5s model for model training and further implementations. Fig. 1 presents the network architecture of YOLOv5s, systematically segmented into four components: The input end, Backbone, Neck, and the prediction module.

Input End The input end of YOLOv5s employs several techniques for processing data images, such as Mosaic data augmentation for enhanced diversity, adaptive image scaling for consistent sizing, and auto-learning bounding box anchors to adapt to image features, thereby improving detection accuracy. Moreover, we use data augmentation techniques to enrich the original dataset used for training to improve training effectiveness.
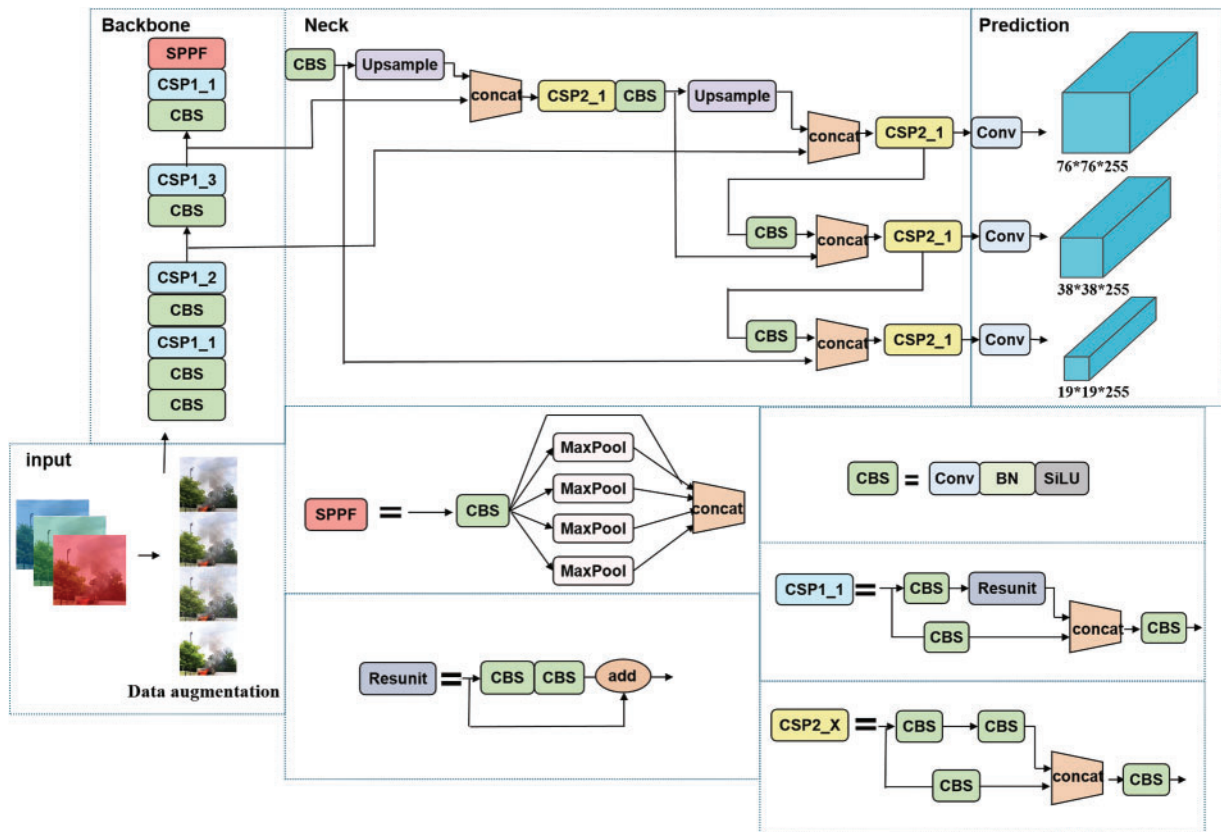
Figure 1: Network architecture of YOLOv5s

Backbone The Backbone of YOLOv5s is a streamlined convolutional neural network that extracts image features. It consists of convolutional layers for varied feature extraction, pooling layers for reducing spatial dimensions, and activation layers for adding non-linearity.

Neck The Neck connects the Backbone and prediction module. It downsizes feature maps for easier processing, employing convolutional layers, pooling layers, and residual connections to efficiently transition from feature extraction to prediction.

Prediction Module The prediction module of YOLOv5s, using convolutional layers, transforms the feature maps into object detection outputs. It predicts object categories, locations, and bounding boxes, while simultaneously adjusting feature map sizes to match the original image.

3.2 Integrate into SE Attention Mechanism

To enhance the capability of model in recognition and classification tasks, we introduce SENet [26] for model optimization. SENet, drawing inspiration from human attention mechanisms, allows the model to focus more intently on significant information within the input feature map, thus boosting performance.

The network utilizes a Squeeze-and-Excitation block (SE block) that first globally averages pools each channel of the feature map, compressing spatial information into a scalar. Subsequently, using the excitation step, feature responses are intensified. A fully connected layer coupled with a sigmoid

function maps this scalar into the [0,1] interval, after which the calculated weights are applied back to the original feature map, achieving dynamic feature recalibration. The structure of the SE block is depicted in Fig. 2.
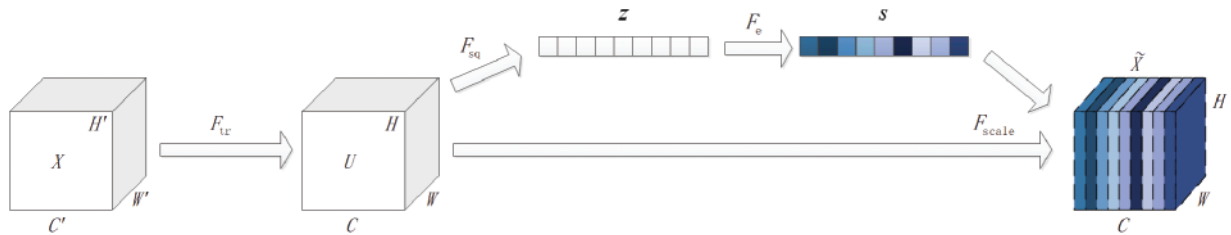

Figure 2: The Squeeze-and-Excitation block

As shown in Fig. 2, the transformation function F_{tr} reshapes the input feature map $X \in \Re^{H' \times W' \times C'}$ into feature maps $U \in \Re^{H \times W \times C}$, where each feature map U is represented by a set of channels $U = [u_1, u_2, \ldots, u_C]$, with $u_c \in \Re^{H \times W}$ and $H \times W$ representing its spatial dimensions.

Then the SE block operates through a sequence of operations, starting with the squeeze operation, as shown in Eq. (1). This global average pooling function, denoted as F_{sq}, shrinks the spatial dimensions of the feature map to produce a channel descriptor z_c for each channel.

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} [u_c]_{i,j}. \tag{1}$$

Building upon this, the excitation operation, detailed in Eq. (2), employs a simple gating mechanism with a sigmoid activation function δ, which generates weights for each channel. This gating mechanism applies a set of learned parameters W_1 and W_2 to the squeezed channel descriptor z to yield the scalar s, encapsulating the excitation output.

$$s = F_{ex}(z, W) = \sigma(W_2 \cdot \delta(W_1 \cdot z)). \tag{2}$$

Finally, the output of the SE block is manifested in Eq. (3), where the output feature map $\tilde{x}_c$ is obtained by scaling the original feature map u_c with the learned activations s_c. This scaling function, represented by F_{scale}, serves as a dynamic recalibration, thereby tailoring the network's focus to salient features while processing the input through subsequent transformations.

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c u_c. \tag{3}$$

In summary, the primary advantage of SENet lies in its robust representational capacity, which significantly improves outcomes across various image processing tasks. Furthermore, the SE block can be seamlessly integrated into other network structures, enhancing the model's ability to recognize key features. In this paper, we integrate the SE block into the convolutional neural network's backbone, inserting SE blocks into each CSP structure within the backbone of YOLOv5, to precisely adjust the feature responses of each channel. This integration elevates the model's sensitivity to essential features and overall performance. Its structure is shown in Fig. 3.
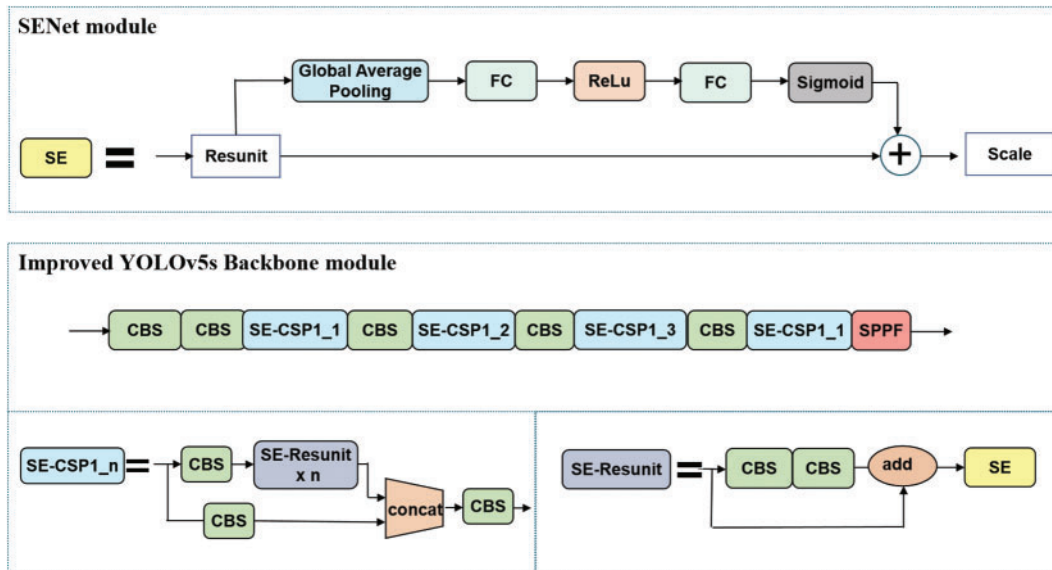
Figure 3: The structure of YOLOv5s and YOLOv5s_SE backbone modules

4 Dataset Preparation

In this section, we present the foundation of our research—the dataset. Section 4.1 details the building of a unique dataset, which was necessary because there were no publicly available datasets specifically designed for smoke detection in outdoor parking lots. Section 4.2 discusses the augmentation strategies implemented to enrich this dataset, enhancing the robustness and generalizability of the detection models.

4.1 Data Collection

Smoke detection in outdoor parking lots is particularly challenging due to the variability introduced by diverse backgrounds and meteorological conditions, such as fog, clouds, rain, and extraneous elements like vehicles, pedestrians, and lighting conditions. Considering the lack of publicly available datasets for parking fire incidents, we aggregated a dataset comprising 4,722 images of smoke and flame across a spectrum of backgrounds from public repositories and internet resources, as shown in Table 1.

Table 1: Primary sources of the original dataset

Source	Category	Link
University of Science and Technology of China	Smoke images	http://smoke.ustc.edu.cn/datasets.htm
Keimyung University	Smoke videos	http://cvpr.kmu.ac.kr/

(Continued)

Table 1 (continued)

Source	Category	Link
Network video	Spontaneous vehicle combustion videos	https://www.bilibili.com/video/BV17k4y1y7Ut/?spm_id_from=333.999.0.0&vd_source=65f0271ddc2feccbf8b4642f35ac83c3
Network video	Spontaneous vehicle combustion videos	https://www.bilibili.com/video/BV1pt4y117mr/?spm_id_from=333.788.recommend_more_video.-1&vd_source=65f0271ddc2feccbf8b4642f35ac83c3
Network video	Spontaneous vehicle combustion videos	https://www.bilibili.com/video/BV1CW411f7pi/?spm_id_from=333.788.recommend_more_video.4&vd_source=65f0271ddc2feccbf8b4642f35ac83c3

Moreover, to refine detection performance for the specialized scenario of vehicular fire incidents in parking lots, we collected 29 videos depicting the spontaneous combustion of vehicles from the network platform.

Subsequent to frame extraction, filtering, dimensional normalization, and targeted cropping, we augment the dataset with an additional 5,963 images, rendering it more representative of outdoor parking contexts. The resultant dataset, encompassing 10,685 distinct smoke and fire image features, constitutes the foundational dataset for this study, as shown in Table 2.

Table 2: Final composition of the original dataset

Category	Quantity	Description
Smoke and fire in complex backgrounds	4,722	Different levels of lighting, weather, occasions, and background complexity.
Frames from spontaneous vehicle combustion videos	5,963	The specific scenario of vehicle spontaneous combustion.
Total	10,685	–

4.2 Data Augmentation

To enhance the precision of image recognition and the generalization capability of the model, we utilized data augmentation techniques to expand the original dataset of fires and smoke. Given the limited scale of public datasets for fire and smoke detection, we initially processed frames from vehicle combustion videos, including extraction, selection, size normalization, and appropriate cropping to enrich the data samples. We also employed fundamental image processing methods such as adjusting the brightness, contrast, saturation of images, and adding Gaussian noise to some images. Smoke detection in outdoor parking lots is susceptible to factors such as lighting and weather conditions. By adjusting the brightness, contrast and saturation of the image, we simulated the change of ambient light

conditions. By adding Gaussian noise to the image, we simulated the inference of weather conditions such as rain and fog, as well as vehicles and pedestrians on smoke detection. These steps were employed to generate additional image data by simulating a greater variety of environmental and interference conditions, effectively enlarged the dataset size, which resulted in a more diverse and enriched dataset, thereby improving the model's adaptability to diverse scenarios. The effect of data augmentation is shown in Fig. 4.

Figure 4: Data augmentation effect

Moreover, to ensure the quality of the dataset, we conducted precise manual annotations on the augmented smoke and fire images with bounding boxes, using LabelImg to aid in and guarantee the accuracy of the annotation process. These annotated datasets were further divided into training, validation, and test sets in an approximate ratio of 8:1:1, supporting the needs for model training and evaluation. The partition of the augmented datasets is detailed in Table 3.

Table 3: Partition of the augmented dataset

Dataset	Total	Training set	Validation set	Test set
Once-augmented dataset	21763	17129	2139	2495
Twice-augmented dataset	38071	30663	3708	3700

After two rounds of data augmentation, the dataset expanded from 10,685 to 38,071 images, enhancing the robustness of the model and providing ample training samples for fire and smoke detection.

5 Experiments

In this section, we describe the experiments we conducted to validate the effectiveness of our proposed model. Section 5.1 details the experimental settings and performance metrics. Section 5.2 presents and analyzes the results of the outcomes of our experiments.

5.1 Experimental Settings

In this experiment, we configured the training parameters carefully to ensure an optimal balance between efficiency and performance. The learning rate was set at 0.01, with a weight decay of 0.0005, and momentum adjusted to 0.937. The batch size was established at 96, and the model underwent training over 800 epochs. Training platform parameters are detailed in Table 4.

Table 4: Hardware and software platforms

Device	Version
CPU	Intel Core Processor (Broadwell, IBRS)
GPU	Tesla V100
Operating system	Ubuntu 18.04
Deep learning framework	PyTorch 1.11.0

To verify the performance of the model, a comprehensive set of metrics including Precision (P), Recall (R), Average Precision (AP), and the mean Average Precision (mAP). Precision and Recall are calculated using Eqs. (4) and (5), as follows:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

In the formula, we mark the result as: TP if the model detects objects in positive videos; FP if the model detects smoke objects in negative videos; TN if the model does not detect objects in negative videos; FN if the model does not detect objects in positive videos [4].

As shown in Eq. (6), mean average precision (mAP) is then calculated by summing the average precision of all categories and then dividing by the total number of categories. Regarding average precision (AP), a key metric for target detection accuracy, is defined as the area under the curve formed by precision and recall rates at various class confidence levels.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{6}$$

5.2 Experimental Results and Analysis

To evaluate the effectiveness of the data augmentation, we conduct with the same algorithm experiments on two validation datasets, respectively, namely once-augmented dataset and twice-augmented dataset. The results of training the model on the once-augmented dataset and the twice-augmented dataset are presented as follows: Fig. 5 illustrates the comparative curves of the average mAP, precision, recall, and classification loss for the two datasets; Fig. 6 presents the comparison of Precision-Recall (PR) curves obtained from training on both datasets.

As shown in Fig. 5, with an increasing number of training epochs, the models trained on both the once-augmented and twice-augmented datasets exhibit an increase in mAP, precision, and recall, gradually stabilizing at precision values exceeding 0.8 and recall values exceeding 0.7. The classification loss decreases gradually and stabilizes at relatively low values. After 300 epochs, the model trained on the twice-augmented dataset outperforms the once-augmented dataset in terms of mAP, precision, and recall. These results confirm the high stability and predictive accuracy of the model, and indicate that data augmentation has improved the model's predictive precision. Moreover, it can be observed that the once-augmented dataset has fully converged around 400 epochs. Therefore, we set the training epochs for the dataset before the second round of data augmentation to 400 for training.
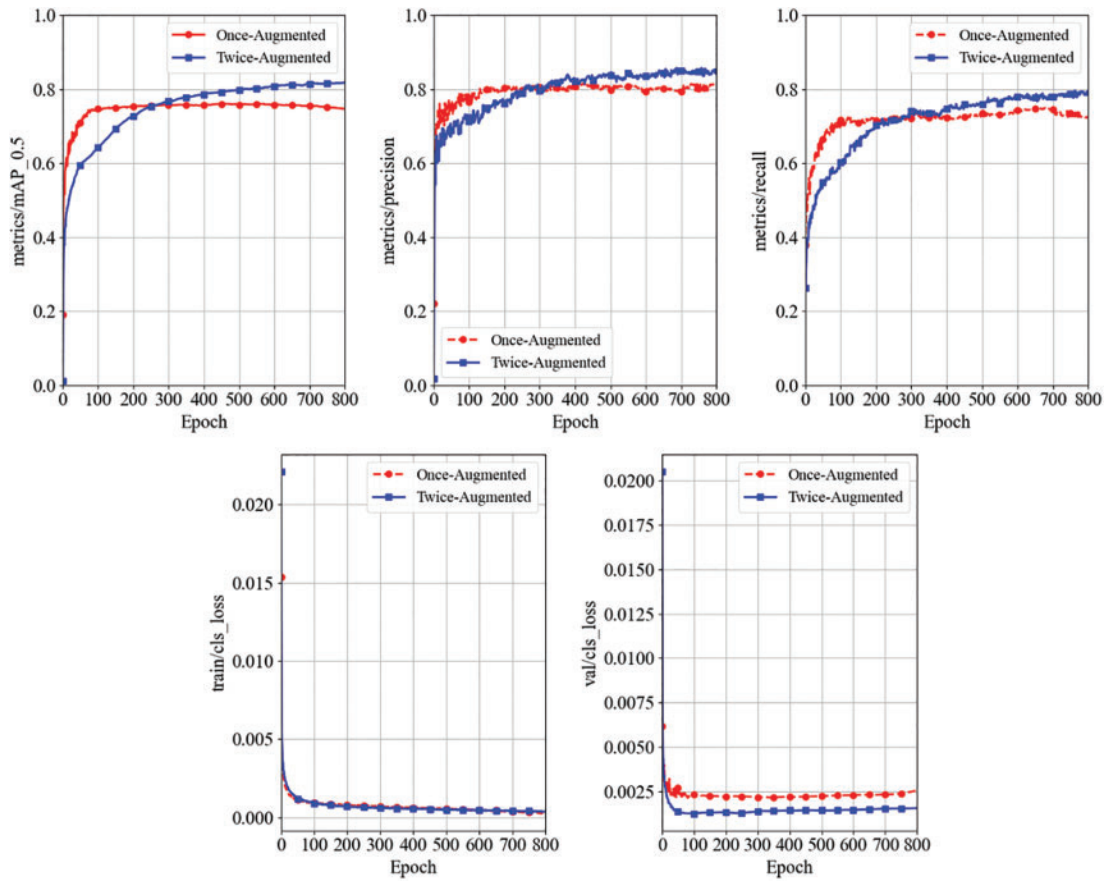
Figure 5: The results on once-augmented and twice-augmented dataset
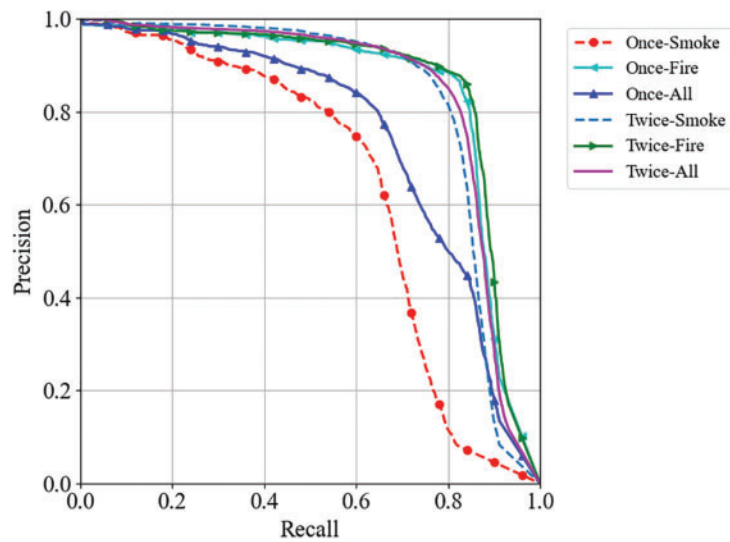

Figure 6: The P-R curves of once-augmented and twice-augmented dataset

In Fig. 6, Once-All represents the average precision-recall curve of smoke and fire detection after the first round of data augmentation, while Twice-All indicates the average precision-recall curve of smoke and fire detection after the second round of data augmentation. As shown in Fig. 6, the smoke and fire detection accuracy improved on twice-augmented dataset compared to once-augmented dataset, and it can be observed that the data augmentation process has consistently enhanced the average precision of smoke and fire detection. This enhancement reaffirms the effectiveness of data augmentation techniques in improving the training results of the model.

Based on the training results from the validation set, we fine-tuned the model and ultimately tested it on the testing set. The test set comprised 2,495 unique vehicle combustion images that were not pre-labeled with actual bounding boxes, including 495 images that did not contain smoke or fire and served as distractors. The model, enhanced with data augmentation and the attention mechanism, demonstrated a relatively high accuracy in detecting smoke and fire targets within images. As illustrated in Fig. 7, the detection results of smoke and fire are depicted in various scenarios. The model demonstrates good detection performance across various environments, showing minimal susceptibility to environmental factors such as complex backgrounds, weather conditions, and changes in lighting, demonstrating strong robustness.

Figure 7: The detection results of smoke and fire

The results from the testing dataset indicate that the rates of missed and false detections for smoke and fire targets are at a low level. To compare the training performance of different algorithms, the datasets before and after data augmentation were used. Training was conducted under the condition of video resolution of 1920×1080 with different algorithms, and the results are presented accordingly.

As shown in Table 5, compared to original dataset, the models trained on datasets augmented after each iteration exhibited significant improvements in average mAP for smoke and fire detection. When utilizing the YOLOv5s algorithm, the first round of data augmentation improved the mean average precision of smoke and fire detection by 0.09. Building upon the once-augmented dataset, the second round of data augmentation further increased the average precision of smoke and fire by 0.069. The integration of YOLOv5 with the SE attention mechanism also effectively enhanced the average precision of smoke and fire detection. To illustrate, in the twice-augmented dataset, compared to solely using the YOLOv5s method, incorporating SENet increases the smoke detection mAP from 0.814 to 0.828 and the fire detection mAP from 0.830 to 0.859.

At the same time, it can be observed that with the addition of the SE module, the model experiences an increase in computation load, leading to a slight decrease in the detection FPS. Nonetheless, it still meets the real-time requirements. Consequently, the algorithm's performance aligns with the application scenario, enabling timely detection of smoke in parking areas and thus reducing the risk of property loss to the public.

Table 5: The performance of the proposed models

Dataset	Method	AP		mAP@ 0.5	FPS
		Smoke	Fire		
Original dataset	YOLOv5s	0.550	0.776	0.663	90.92
	YOLOv5s + nSE	0.574	0.808	0.691	76.56
Once-augmented dataset	YOLOv5s	0.646	0.859	0.753	89.95
	YOLOv5s + nSE	0.654	0.872	0.763	76.31
Twice-augmented dataset	YOLOv5s	0.814	0.830	0.822	90.47
	YOLOv5s + nSE	0.828	0.859	0.843	76.50

6 System Implementation

We leverage the enhanced YOLOv5 algorithm and OpenCV computer vision technology to implement a smoke detection system. This system is developed using the Visual Studio 2019 integrated development environment and programmed in C/C++. Additionally, we develop the system interface with Qt5, and the user interface (UI) is enhanced with Qt Designer, resulting in an advanced smoke detection and early warning system for outdoor parking spaces. To further reduce the rate of undetected spontaneous vehicle combustion and to refine system functionality, our system is also equipped with fire detection capabilities. An overview of the algorithm and structure of our system is described in Fig. 8.
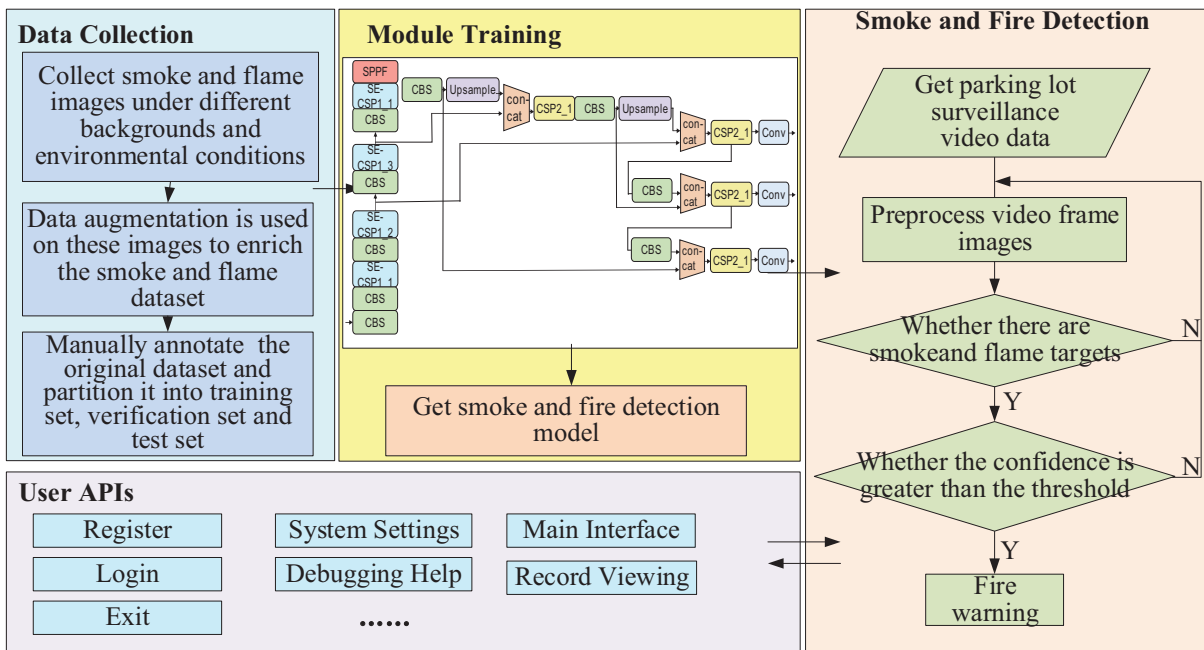
Figure 8: System module flow chart

As shown in Fig. 8, the Data Collection module illustrates the process of initial data collection and processing, and the Module Training process demonstrates how we generate the parking lot smoke and fire detection model leveraging the improved YOLOv5 network structure. It is worth noting that the smoke and fire detection model obtained here uses the YOLOv5-nSE network structure mentioned above, demonstrating strong performance in detecting smoke and fire targets. Some of the functional APIs available to users of the system are presented in the User APIs module. And finally, the Smoke and Fire Detection module describes how the system works and generates fire warnings.

Specifically, the process of smoke and fire detection algorithm is as follows:

Step 1: The system acquires video data either directly from surveillance cameras or by importing from local storage.

Step 2: The video data is preprocessed to ensure it aligns with the input specifications of our detection model, and is then displayed on a user interface for monitoring.

Step 3: The implemented algorithm analyzes video frames to identify the presence of smoke or fire targets.

Step 4: The system triggers an alert and marks the detection on the interface when identifying a target with a confidence level. These frames are also saved for later review and evidence preservation.

Step 5: The system proceeds to process and analyze subsequent video frames to maintain continuous surveillance if smoke or fire target is not detected.

Moreover, the interface of the system is shown in Fig. 9.

Figure 9: System interface

The functions that users can achieve through the user interface are described as follows:

Main Interface It controls the playback of locally imported or camera-fed videos, handling start, pause, and progress bar operations. Upon smoke or fire detection, it displays bounding boxes on the screen and emits an alert signal.

System Settings It includes basic settings for importing local video paths and user settings for direct camera connection by entering camera addresses and names, facilitating real-time monitoring of multiple areas.

Record Viewing For ease of investigating the cause and timing of fire incidents, the system provides a detection record viewing function, allowing examination of the time and visuals from when a fire event occurred.

Debugging Help This module provides instructions for using the system, enabling users to quickly become proficient.

User Exit This module ensures a user can conveniently exit the system.

7 Conclusion

In this paper, we develop an advanced smoke detection system for outdoor parking lots by integrating an improved YOLOv5s model with the SE attention mechanism. Our approach addresses the challenges posed by complex outdoor environments, utilizing a newly constructed and enriched dataset for training and validation. While the system exhibits impressive precision and swift detection responses, it still encounters challenges with very small smoke instances. This is attributed to the feature extraction layers of convolutional neural networks may lose critical information during downsampling operations. Additionally, real-world factors also pose significant challenges, such as inadequate camera resolution and extreme weather. In the future, we aim to refine our methodologies, potentially by incorporating more sophisticated attention mechanisms, to enhance robustness of the system and decrease the likelihood of false positives. Future work will also focus on optimizing the model for even faster deployment and greater reliability, ensuring that the system can effectively contribute to public safety by providing timely alerts in the crucial early stages of vehicle combustion incidents.

Author Contributions: The authors confirm contribution to the paper as follows: Research conception and design: Zhenyong Zhang; data collection: Ruobing Zuo, Xuguo Jiao; analysis and interpretation of results: Ruobing Zuo, Xiaohan Huang; draft manuscript preparation: Ruobing Zuo, Xiaohan Huang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] G. S. Kushwaha and N. K. Sharma, "Green initiatives: A step towards sustainable development and firm's performance in the automobile industry," *J. Clean. Prod.*, vol. 121, no. 3, pp. 116–129, 2016. doi: 10.1016/j.jclepro.2015.07.072.

[2] M. Tauqeer *et al.*, "Driver's emotion and behavior classification system based on Internet of Things and deep learning for Advanced Driver Assistance System (ADAS)," *Comput. Commun.*, vol. 194, no. 3, pp. 258–267, 2022. doi: 10.1016/j.comcom.2022.07.031.

[3] H. R. Boehmer, M. S. Klassen, and S. M. Olenick, "Fire hazard analysis of modern vehicles in parking facilities," *Fire Technol.*, vol. 57, no. 5, pp. 2097–2127, 2021. doi: 10.1007/s10694-021-01113-1.

[4] S. Saponara, A. Elhanashi, and A. Gagliardi, "Real-time video fire/smoke detection based on CNN in antifire surveillance systems," *J. Real Time Image Process.*, vol. 18, no. 3, pp. 889–900, 2020. doi: 10.1007/s11554-020-01044-0.

[5] A. Gaur *et al.*, "Fire sensing technologies: A review," *IEEE Sens. J.*, vol. 19, no. 9, pp. 3191–3202, 2019. doi: 10.1109/JSEN.2019.2894665.

[6] J. Fonollosa, A. Solórzano, and S. Marco, "Chemical sensor systems and associated algorithms for fire detection: A review," *Sens.*, vol. 18, no. 2, pp. 553, 2018. doi: 10.3390/s18020553.

[7] L. A. Cestari, C. L. Worrell, and J. A. Milke, "Advanced fire detection algorithms using data from the home smoke detector project," *Fire Saf. J.*, vol. 40, no. 1, pp. 1–28, 2005. doi: 10.1016/j.firesaf.2004.07.004.

[8] D. Gutmacher, U. Hoefer, and J. Wöllenstein, "Gas sensor technologies for fire detection," *Sens. Actuators B: Chem.*, vol. 175, pp. 40–45, 2012. doi: 10.1016/j.snb.2011.11.053.

[9] S. J. Chen, D. C. Hovde, K. D. Peterson, and A. W. Marshall, "Fire detection using smoke and gas sensors," *Fire Saf. J.*, vol. 42, no. 8, pp. 507–515, 2007. doi: 10.1016/j.firesaf.2007.01.006.

[10] S. B. Kukuk and Z. H. Kilimci, "Comprehensive analysis of forest fire detection using deep learning models and conventional machine learning algorithms," *Int. J. Comput. Exp. Sci. Eng.*, vol. 7, no. 2, pp. 84–94, 2021. doi: 10.22399/ijcesen.950045.

[11] D. Sheng, L. Deng, and J. Wang, "Automatic smoke detection based on SLIC-DBSCAN enhanced convolutional neural network," *IEEE Access*, vol. 9, pp. 63933–63942, 2021. doi: 10.1109/ACCESS.2021.3075731.

[12] X. Jiang, C. Hu, and Z. Fan, "Research on flame detection method by fusion feature and sparse representation classification," *Int. J. Comput. Commun. Eng.*, vol. 5, no. 4, pp. 238–245, 2016. doi: 10.17706/IJCCE.2016.5.4.238-245.

[13] J. Gubbi, S. Marusic, and M. Palaniswami, "Smoke detection in video using wavelets and support vector machines," *Fire Saf. J.*, vol. 44, no. 8, pp. 1110–1115, 2009. doi: 10.1016/j.firesaf.2009.08.003.

[14] R. Kaabi, S. Frizzi, M. Bouchouicha, F. Fnaiech, and E. Moreau, "Video smoke detection review: State of the art of smoke detection in visible and IR range," in *Proc. SM2C*, Sfax, Tunisia, 2017, pp. 81–86.

[15] L. Qin, X. Wu, Y. Cao, and X. Lu, "An effective method for forest fire smoke detection," *J. Phys.: Conf. Ser.*, vol. 1187, no. 5, pp. 052045, 2019. doi: 10.1088/1742-6596/1187/5/052045.

[16] M. R. Islam, M. Amiruzzaman, S. Nasim, and J. Shin, "Smoke object segmentation and the dynamic growth feature model for video-based smoke detection systems," *Symmetry*, vol. 12, no. 7, pp. 1075, 2020. doi: 10.3390/sym12071075.

[17] D. Xiong and L. Yan, "Early smoke detection of forest fires based on SVM image segmentation," *J. For. Sci.*, vol. 65, no. 4, pp. 150–159, 2019. doi: 10.17221/82/2018-JFS.

[18] X. Xu and J. Xu, "Automatic fire smoke detection based on image visual features," in *Proc. CISW*, Harbin, China, 2007, pp. 316–319.

[19] Y. C. Hu and X. B. Lu, "Real-time video fire smoke detection by utilizing spatial-temporal ConvNet features," *Multimed. Tools Appl.*, vol. 77, no. 22, pp. 29283–29301, 2018. doi: 10.1007/s11042-018-5978-5.

[20] Z. Habib, M. A. Mughal, M. A. Khan, and M. Shabaz, "WiFOG: Integrating deep learning and hybrid feature selection for accurate freezing of gait detection," *Alex. Eng. J.*, vol. 86, no. 6, pp. 481–493, 2024. doi: 10.1016/j.aej.2023.11.075.

[21] Z. Zhang *et al.*, "Vulnerability of machine learning approaches applied in IoT-based smart grid," *IEEE Internet Things J.*, 2024. doi: 10.1109/JIOT.2024.3349381.

[22] Z. Zhang, Z. Yang, D. K. Y. Yau, Y. Tian, and J. Ma, "Data security of machine learning applied in low-carbon smart grid: A formal model for the physics-constrained robustness," *Appl. Energy*, vol. 347, no. 1, pp. 121405, 2023. doi: 10.1016/j.apenergy.2023.121405.

[23] S. Wang and J. Xiang, "A minimum entropy deconvolution-enhanced convolutional neural networks for fault diagnosis of axial piston pumps," *Soft Comput.*, vol. 24, no. 4, pp. 2983–2997, 2020. doi: 10.1007/s00500-019-04076-2.

[24] R. Yamashita, M. Nishio, R. K. Gian, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights Imag.*, vol. 9, no. 4, pp. 611–629, 2018. doi: 10.1007/s13244-018-0639-9.

[25] N. Hussain *et al.*, "Intelligent deep learning and improved whale optimization algorithm based framework for object recognition," *Hum. Centric Comput. Inf. Sci.*, vol. 11, pp. 34–50, 2021.

[26] J. Hu, S. Li, and S. Gang, "Squeeze-and-excitation networks," in *Proc. CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141.

[27] T. H. Chen, P. H. Wu, and Y. C. Chiou, "An early fire-detection method based on image processing," in *Proc. ICIP*, Singapore, 2004, pp. 1707–1710.

[28] M. Ajith and M. Martinez-Ramon, "Unsupervised segmentation of fire and smoke from infra-red videos," *IEEE Access*, vol. 7, pp. 182381–182394, 2019. doi: 10.1109/ACCESS.2019.2960209.

[29] K. Dimitropoulos, P. Barmpoutis, and N. Grammalidis, "Higher order linear dynamical systems for smoke detection in video surveillance applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 5, pp. 1143–1154, 2016. doi: 10.1109/TCSVT.2016.2527340.

[30] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. CVPR*, Columbus, OH, USA, 2014, pp. 580–587.

[31] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. CVPR*, Las Vegas, NV, USA, 2016, pp. 9–788.

[32] Y. M. Luo, L. Zhao, P. Z. Liu, and D. T. Huang, "Fire smoke detection algorithm based on motion characteristic and convolutional neural networks," *Multimed. Tools Appl.*, vol. 77, no. 12, pp. 15075–15092, 2018. doi: 10.1007/s11042-017-5090-2.

[33] H. A. H. Mahmoud, A. H. Alharbi, and N. S. Alghamdi, "Time-efficient fire detection convolutional neural network coupled with transfer learning," *Intell. Autom. Soft Comput.*, vol. 31, no. 3, pp. 1393–1403, 2022. doi: 10.32604/iasc.2022.020629.

[34] Z. Yin, B. Wan, F. Yuan, X. Xia, and J. Shi, "A deep normalization and convolutional neural network for image smoke detection," *IEEE Access*, vol. 5, pp. 18429–18438, 2017. doi: 10.1109/ACCESS.2017.2747399.

[35] G. H. Lin, Y. M. Zhang, G. Xu, and Q. X. Zhang, "Smoke detection on video sequences using 3D convolutional neural networks," *Fire Technol.*, vol. 55, no. 5, pp. 1827–1847, 2019. doi: 10.1007/s10694-019-00832-w.

[36] J. Zeng, Z. Lin, C. Qi, X. Zhao, and F. Wang, "An improved object detection method based on deep convolutional neural network for smoke detection," in *Proc. ICMLC*, Chengdu, China, 2018, pp. 184–189.

[37] Q. X. Zhang, G. H. Lin, Y. M. Zhang, G. Xu, and J. J. Wang, "Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images," *Procedia Eng.*, vol. 211, no. 3, pp. 441–446, 2018. doi: 10.1016/j.proeng.2017.12.034.