

Research Article

Real-Time Fire Detection Method Based on Computer Vision for Electric Vehicle Charging Safety Monitoring

Yuchen Gao ¹, Qing Yang ¹, Shiyu Zhang ² and Dexin Gao ²

¹School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao, Shandong 266061, China

²School of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao, Shandong 266061, China

Correspondence should be addressed to Qing Yang; 03390@qust.edu.cn

Received 16 November 2022; Revised 27 March 2023; Accepted 2 August 2023; Published 17 August 2023

Academic Editor: Gongping Yang

Copyright © 2023 Yuchen Gao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the process of charging and using electric vehicles, lithium battery may cause hazards such as fire or even explosion due to thermal runaway. Therefore, a target detection model based on the improved YOLOv5 (You Only Look Once) algorithm is proposed for the features generated by lithium battery combustion, using the K-means algorithm to cluster and analyse the target locations within the dataset, while adjusting the residual structure and the number of convolutional kernels in the network and embedding a convolutional block attention module (CBAM) to improve the detection accuracy without affecting the detection speed. The experimental results show that the improved algorithm has an overall mAP evaluation index of 94.09%, an average F1 value of 90.00%, and a real-time detection FPS (frames per second) of 42.09, which can meet certain real-time monitoring requirements and can be deployed in various electric vehicle charging stations and production platforms for safety detection and will provide a guarantee for the safe production and development of electric vehicles in the future.

1. Introduction

As people's awareness of environmental protection increases, the proportion of new-energy electric vehicles is also increasing and their supporting infrastructure construction, such as charging stations, is also increasing year by year. The power source of new energy electric vehicles is mainly lithium battery, which has the advantages of high storage energy density, long service life, and lightweight and is very popular in all kinds of new energy vehicle products. However, in the process of rapid charging of new energy electric vehicles at charging stations, the internal lithium battery may leak, catch fire, or even explode due to thermal runaway and other reasons [1]. Therefore, it becomes especially important to monitor the safety of electric vehicles in the charging process.

At present, deep learning-based target detection methods have gradually become the mainstream, including one-stage series algorithms and two-stage series algorithms.

The two-stage series algorithms include R-convolutional neural network (R-CNN), faster R-CNN, and other networks. In the first step, region proposal network (RPN) is trained, and in the second step, the classification and location information of the target are predicted by convolutional neural network [2]. The one-stage series of algorithms, mainly, single shot multibox detector (SSD) algorithms [3], adopt the idea of mathematical regression, omitting RPN, and directly regressing to obtain the class probability and location coordinates of the object, with slightly lower accuracy, but significantly improving the detection speed. The combustion of electric vehicle battery cells is complex and variable, with no exact rule. Once the internal stable state of the lithium battery due to collision, overcharging, and other situations is established, thermal runaway will quickly lead to the battery electrolyte decomposition and other reactions, resulting in the release of a large amount of heat and rapid heating of the battery, but also generating a lot of hydrogen, methane, and other white smoke. The lithium battery fire

process is rapid, often resulting in booming combustion and even explosion phenomena. The fire will usually ignite the surrounding electrical equipment, causing smoke and continuous burning phenomena [4–6]. Experiments in the literature [7] proved that the main way to affect the surrounding environment when the battery compartment explosion of electric vehicles is high temperature, the temperature in the battery compartment will rise to 2158K in 0.12 s, the high temperature will spread around horizontally along the pressure relief hole, and it is very easy to cause the surrounding charging pile or other vehicles to burn. In addition, in the case of large-scale outdoor charging stations, it is difficult to cover all the scenarios using traditional physical sensors and is susceptible to the influence of the surrounding environment, such as driver smoking and restaurant smoking. If not detected in time, more damage will occur.

Earlier traditional detection methods were based on image feature-based recognition judgments. Smoke and flame combustion features are diverse, and the color, texture, and motion features are extremely complex. Xie et al. [8] proposed a method for early detection of fires in indoor enclosed environments based on the reflective properties of firelight, while developing a highly sensitive foreground identification method for flame detection by using strategic background updates and block binarisation thresholds, but it is difficult to apply to complex scenes with multiple reflections outdoors. Liu et al. [9] presented the $Y_aU_aV_a$ colour model to analyse the colour changes and motion trajectories of smoke in adjacent frames, thus roughly filtering out blocks of images suspected of having smoke. Du et al. [10] improved the ViBe algorithm based on the color features of smoke to extract smoke features. Chen et al. [11] introduced a convolutional network to extract smoke texture information and combine it with the static texture information of the original image for detection. Zhao et al. [12] built a classification model of flame elements by YCbCr color space and formulated new rules to reduce the interference of image brightness. Wu et al. [13] performed multithreshold segmentation based on the image grayscale entropy criterion and used an improved particle swarm optimization algorithm to select thresholds as a way to quickly segment flame targets and background regions. The Krawtchouk torch was introduced to construct the feature vectors of flame images as a way to construct support vector machines for detection [14]. All the above methods are based on traditional image-based detection methods, which have low accuracy and slow speed. Nowadays, many deep learning-based target detection methods are applied in video flame detection with better results. The accuracy of faster R-CNN on flame detection task is improved by using color-guided anchoring-based strategy to constrain the generation region of anchor frames, but the detection speed still needs to be improved [15]. The detection efficiency of small flame regions is particularly improved by improving the prior frame of YOLOv3 network and combining the features of flame flicker to reduce the false detection [16], but the detection speed of YOLOv3 algorithm is slow and not applicable to video stream monitoring. To track ship fires, Wu et al. [17]

modified the YOLOv4 algorithm with the SE attention mechanism module. Cai et al. [18] improved the YOLOv4 algorithm by replacing the network structure and pruning operations to achieve real-time object detection on an in-vehicle platform. Deep separable convolution was applied to a YOLOv4 network by Huo et al. [19] to enhance the algorithm's ability to detect smoke, but the algorithm was too old and lacked the capability to detect early flames. Wu et al. [20] proposed a flame detection model using the YOLOv5 network, but it could not detect the intense smoke phenomenon in the early stages of lithium battery combustion and lacked the timeliness of hazard prediction. Li et al. [21] applied the YOLOv5 algorithm to the field of remote sensing and proposed the TCS-YOLO method by adding convolutional layers and replacing activation functions to improve the efficiency of identifying global oil storage tanks. Wang et al. [22] applied a structurally reparameterised adaptation of the re-param visual geometry group (RepVGG) model to the conventional CenterNet to achieve object detection in mobile driving scenarios. The YOLO [23] family of algorithms is also commonly used in applications such as marine, biomedical, and autonomous driving and is extremely versatile and stable [24–30]. In recent years, the percentage of fire accidents in electric vehicles that occur in the parked state and during charging can be more than half of the accidents [31–33]. Therefore, the safety monitoring of electric vehicles during charging is very important.

We propose an enhanced YOLOv5-based electric vehicle charging safety monitoring algorithm in this paper for lithium battery combustion and fire characteristics, such as white smoke and fire light from deflagration, that may be brought on by thermal runaway during the charging process. The method suggested in this article can be directly applied to the existing video surveillance equipment, and it is less expensive, more universal, and easier to implement than the traditional detection methods. However, it also has a higher monitoring accuracy and speed when compared to the unimproved algorithm. The Methods section of the paper describes the algorithms used as well as the innovations and improvements made to address the issue at hand. Following this, the Experimental Procedure and Results Analysis sections present experiments and comparisons based on the improved approach. Finally, we present experimental conclusions and affirmations for future applications in real-world scenarios. Figure 1 depicts the flowchart's overall structure. This paper's main contributions are as follows:

- (1) In order to solve the problem that the original algorithm has poor detection capability for small targets in the complex scenario of electric vehicle combustion, the number of convolution kernels inside the algorithm is increased and more residual components are stacked in the feature extraction part to improve the detection capability of the algorithm for small targets.
- (2) To address the uncertainty and complexity of the target locations in the flame smoke dataset, the K-means clustering algorithm was introduced to recluster the target locations in the dataset to obtain

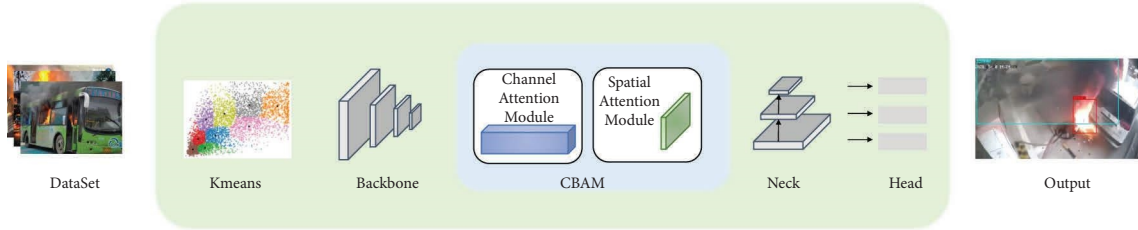


FIGURE 1: Structural diagram of algorithm.

the most suitable anchor frame for this dataset and improve the training speed and accuracy of the algorithm.

- (3) To enhance the extraction capability of the method for the flame smoke features generated by lithium battery combustion as well as to improve the generalization capability and robustness of the method. The CBAM [34] is added after the backbone feature extraction network to significantly improve the method's ability to detect flames and smoke with only a small amount of code added.

2. Methods

2.1. K-Means Algorithm-Based Flame Smoke Anchor Frame Planning. The first step of the YOLO series algorithm in the process of target detection is to generate candidate regions (anchor box). In the combustion process of lithium battery, the state changes are complex and dramatic, with great uncertainty, so this method uses the K-means algorithm to recluster the target location of the features in the dataset to get the anchor box based on the lithium battery combustion feature dataset and used for training. Firstly, the cluster centres are selected, and K samples are then counted as cluster centres $\{c_1, c_2, \dots, c_k\}$ from the dataset. Then, for each sample x_i in the dataset, we calculate the distance d_{ik} to each of the K cluster centres and assign it to the category corresponding to the cluster centre with the smallest distance. For each category c_k , we recalculate the coordinates of its cluster centre, i.e., the centre of the mass of all samples belonging to that category. We repeat the above steps until the cluster centre positions no longer change.

$$d_{ik} = 1 - \text{IOU}(\text{bboxes}, \text{anchors}), \quad (1)$$

where d_{ik} is the distance from sample x_i to the cluster centre K and IOU is the intersection ratio between the two samples, which has a smaller error compared to the traditional Euclidean formula.

The clustering results are shown in Figure 2. The maximum values of the horizontal and vertical coordinates represent the image input size for this algorithm. Clustering is based on all the marked flame positions in the dataset, and 9 clustering centres are obtained, which are represented by asterisks.

Table 1 shows the preselected box positions obtained after clustering the target positions within the dataset using the K-means algorithm, compared to the preselected box positions obtained from training based on the COCO

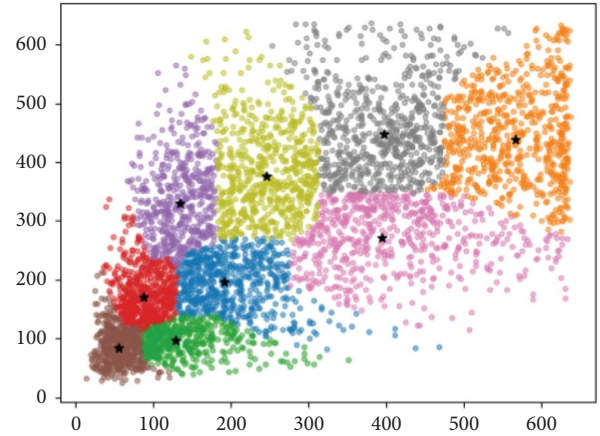


FIGURE 2: Location of K-means cluster centre.

dataset. The use of clustered preselected boxes is more beneficial to improve the training accuracy and detection results.

2.2. YOLOv5 Network Model. The YOLO series target detection algorithms solve the problem of target detection by regressing the anchor frames into which the images are divided and have good real-time performance as an end-to-end detection algorithm. The network structure of the unimproved YOLOv5 algorithm is shown in Figure 3.

The YOLOv5 target detection algorithm can be divided into four parts: input, backbone, neck, and head. Using Mosaic data enhancement in the input section, four images are stitched together to form a new image by random scaling, cropping, and random rows. This reduces the GPU memory required for training while greatly enriching the dataset and improving the robustness of the network. The backbone part is the backbone network for extracting the burning features of lithium battery during the charging process. The YOLOv5 algorithm adds a focus structure and a cross-stage partial network (CSPNet) structure, which extracts a picture based on its width and height, one pixel apart, and finally obtains four independent feature layers, which are then stacked to concentrate the picture's width and height information into the channel space, quadrupling the input channel without losing the picture information, as shown in Figure 4. The CSPNet structure splits the feature map into two parts: one part is convolved and the other part is spliced and fused. It enhances the feature extraction and maintains accuracy while being lightweight, while reducing computational bottlenecks and memory costs.

TABLE 1: Comparison of anchor frame positions after clustering of flame and smoke dataset K-means.

	K-means clustering before and after anchor box position																	
	10	13	16	30	33	23	30	61	62	45	59	119	116	90	156	198	373	326
COCO	10	13	16	30	33	23	30	61	62	45	59	119	116	90	156	198	373	326
Fire and smoke	48	90	102	87	89	175	177	164	157	332	307	243	302	409	469	326	522	466

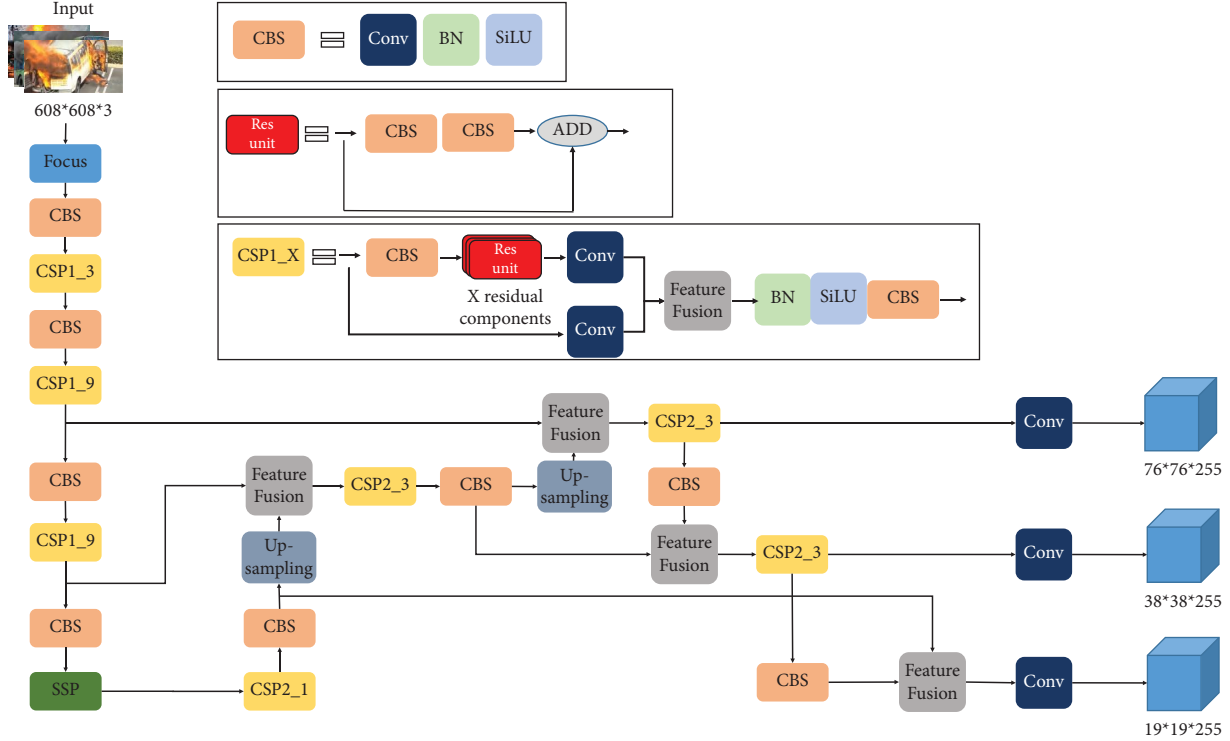


FIGURE 3: YOLOv5 network structure diagram.

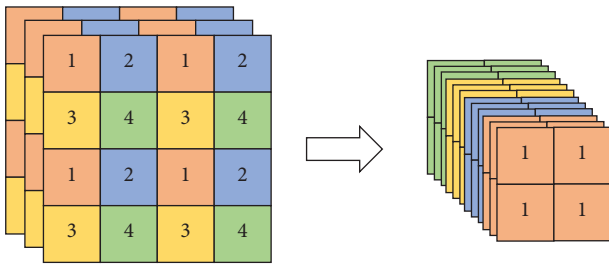


FIGURE 4: Schematic diagram of focus structure.

The neck part consists of path aggregation network (PANet) and spatial pyramid pooling (SPP) module. The PANet enhances the feature representation of the feature extraction network by fusing bottom-up and bottom-down paths. The SPP structure can use different sizes of pooling and tensor stitching to effectively increase the model perceptual field and separate significant contextual features. It avoids image distortion caused by cropping and scaling of image regions and solves the problem of repeated feature extraction, saving computational costs. The head part will output three feature layers of different depths depending on the detection target of the dataset.

2.3. CBAM. Adding an attention mechanism module is a common optimization process in the field of deep learning that allows the model to simulate the focus information observed by the human eyes. CBAM is used to assign different weights to the picture features extracted from the backbone network, which suppresses useless information and improves the utilization of effective features in the neck part.

The CBAM is a simple and effective feedforward convolutional neural network attention mechanism module, which innovatively accesses the spatial attention module after the channel attention module compared to squeeze-and-excitation network (SENet) and efficient channel attention module networks (ECANet) and uses the summation and stacking of maximum pooling and average pooling to make the feature map obtain adaptive feature refinement with corresponding weight proportion. CBAM incorporates effective suppression of the flame and smoke background information and emphasizes the target feature information. In this paper, after the backbone network, the output feature vectors are fed into the CBAM, which is a lightweight general-purpose module that can effectively improve the detection accuracy with little impact on the detection speed. The CBAM is shown in Figure 5.

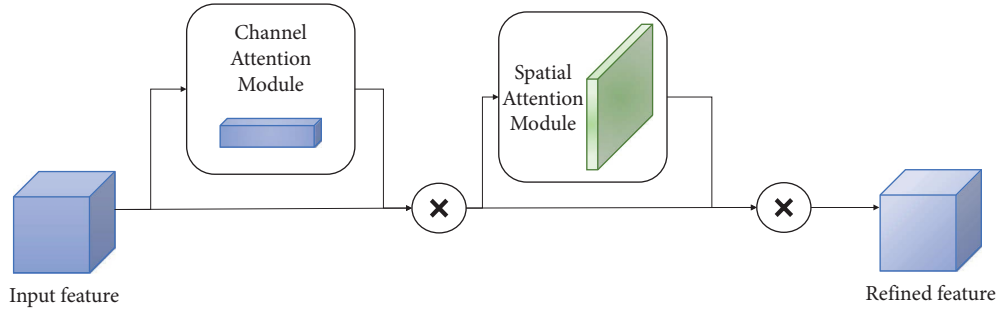


FIGURE 5: Schematic diagram of CBAM.

The channel attention mechanism is to pass the input feature information through global maximum pooling and global average pooling, respectively, through the multilayer perceptron multilayer perceptron (MLP) and then add the results obtained from the two pooling counterparts and output the results after multiplying the corresponding weights with the original input feature map by the Sigmoid function, as shown in Figure 6. Equation (1) is given by

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))), \quad (2)$$

where F is the input feature map, $\text{MLP}()$ is the multilayer perceptron, $\text{AvgPool}()$ is the mean pooling, $\text{MaxPool}()$ is the global pooling, and σ is the Sigmoid operation.

Among them, the spatial attention mechanism is to use the feature map output from the channel attention mechanism module as the input feature map. Firstly, global maximum pooling and global average pooling based on the number of channels are performed. Then, the two results are concatenated based on the number of channels. After a convolution operation, they are reduced to one channel, and finally, the results are passed through the Sigmoid function to generate the channel attention features and multiplied with the input feature map to obtain the spatial generated features, with the following equation. The spatial attention mechanism in CBAM is shown in Figure 7.

$$M_s(F) = \sigma(f^{7*7}([\text{AvgPool}(F); \text{MaxPool}(F)])), \quad (3)$$

where σ is the Sigmoid operation and $7*7$ denotes the convolution kernel size.

2.4. Improvement in the Number of Residual Components and Convolutional Kernels. The YOLOv5 algorithm is divided into networks with different weights and detection capabilities by using different numbers of residual component modules and different numbers of convolutional kernels, so that its networks have different depths and widths. For the YOLOv5 algorithm with different widths and depths, the depth and width of the network as well as the number of residual components and the number of convolutional kernels are controlled by two parameters, namely, `Depth_multiple` and `Width_multiple`, which can be classified as YOLOv5-s, YOLOv5-m, YOLOv5-l, and YOLOv5-x

according to the commonly used parameters, and the corresponding parameters are shown in Table 2.

2.5. Improved YOLOv5 Algorithm. Finally, we adjusted the number of residual structures and the number of convolution kernels in the YOLOv5 algorithm to select the parameters most suitable for lithium battery combustion feature detection, achieving better recognition results compared to the unimproved algorithm. A CBAM attention mechanism module was also added after the feature extraction network to increase the weights of valid features, allowing the algorithm to focus more on important features and suppress unnecessary features, eventually greatly improving the detection accuracy without affecting the detection speed.

The improved network structure is shown in Figure 8. Compared with Figure 3, the differences between the improved algorithm and the unimproved algorithm are marked in Figure 8, with a blue background.

3. Experimental Procedure and Results Analysis

3.1. Dataset Acquisition and Preprocessing. The initial characteristics of electric vehicle lithium battery fires are often dominated by white smoke, open flames of booming combustion, and then gradual ignition of other body structures which produces black smoke, continuous flames, etc. Despite being extremely complex, the features and background share some similarities, so we focus on selecting images with the above features to build the dataset. Figure 9 shows the schematic diagram of some of the images in the flame smoke dataset, where (a) represents the image of an electric vehicle catching fire during charging at a charging station; (b) and (c) represent the images of a sudden lithium battery fire during the use of an electric vehicle; and (d) represents the image of an electric vehicle when it catches fire at night. LabelImg software is used to annotate the flame and smoke parts in the above dataset and generate XML files with target location information to form a dataset following the VOC annotation specification. In the real scenario, there are situations such as partial masking of flames and fusion of smoke with the sky background. For different situations, rectangular boxes of different sizes are used for annotation to improve the accuracy of the dataset content. The final composition consists of 3391 images, as the experimental dataset.

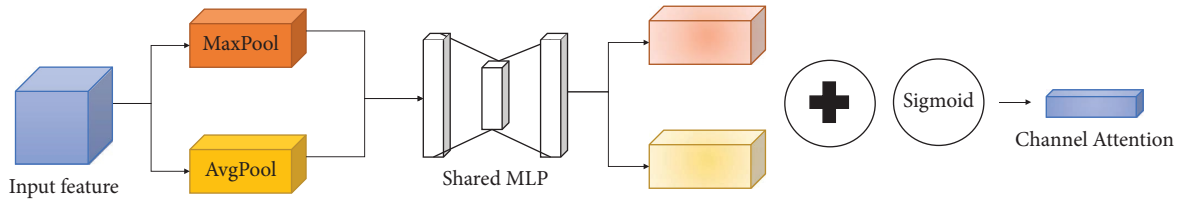


FIGURE 6: Schematic diagram of channel attention mechanism.

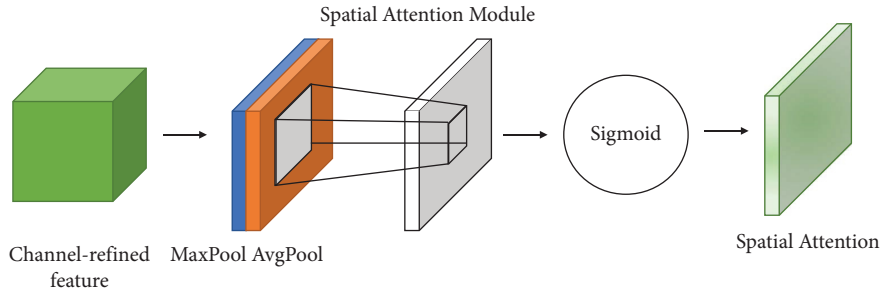


FIGURE 7: Schematic diagram of spatial attention mechanism.

TABLE 2: Comparison table of YOLOv5 algorithm structure differences under different parameters.

	Depth_multiple	Width_multiple	Residual component bases	Convolution kernel bases
YOLOv5-s	0.33	0.50	1	32
YOLOv5-m	0.67	0.75	2	48
YOLOv5-l	1.00	1.00	3	64
YOLOv5-x	1.33	1.25	4	80

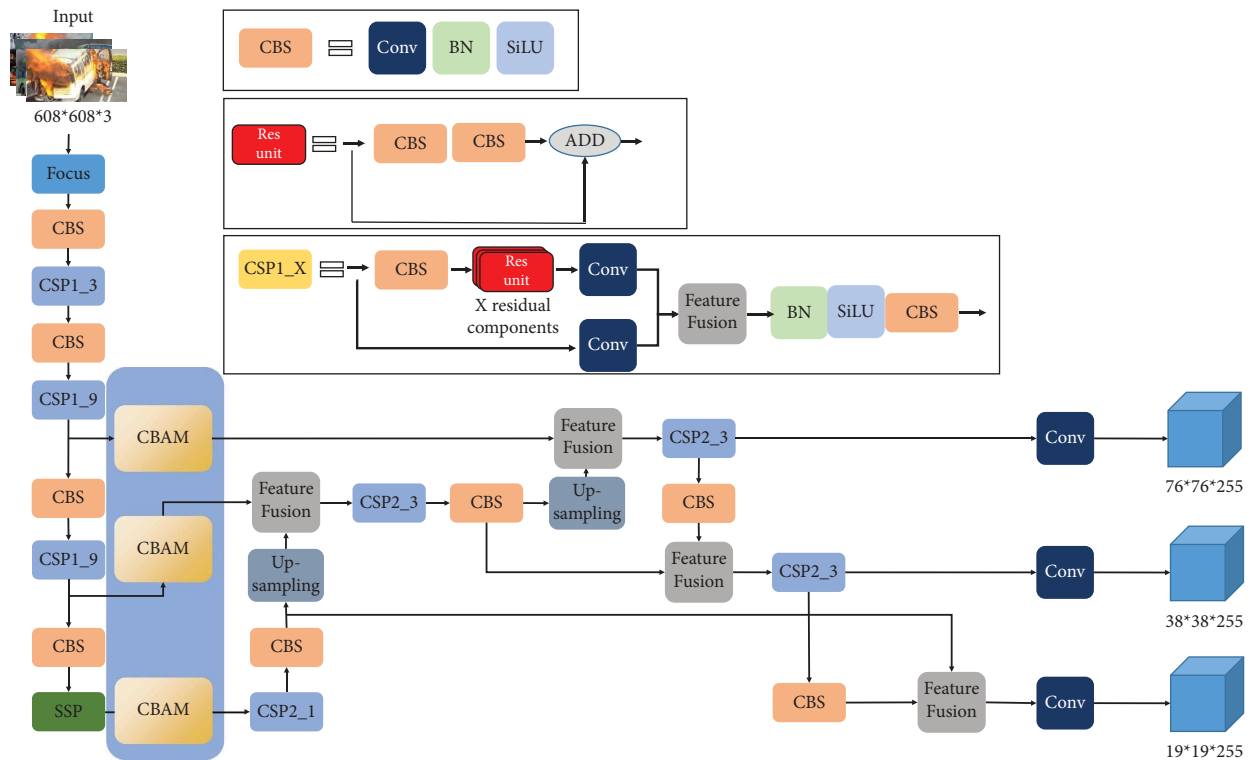


FIGURE 8: Schematic diagram of improved YOLOv5 network structure.



FIGURE 9: Schematic diagram of flame and smoke dataset. (a) Fire during charging, (b) electric vehicle fire side, (c) electric vehicle fire front, and (d) fire at night.

3.2. *Experimental Platform and Model Training.* This experiment is conducted in the same environment with an Intel(R) Xeon(R) Gold 6130H CPU, an NVIDIA RTX3060 GPU (24GB), 32 GB of running memory, and a Windows software environment with PyTorch deep learning framework. To verify the practicality and effectiveness of this algorithm, the YOLOv5 algorithm with different widths and depths, different backbone networks, and different attention mechanisms is compared, and the progress of the algorithm is verified compared to the previous generation YOLOv4 algorithm. For the classification problem of detecting targets, the samples can be classified into four cases: true positive (TP), false positive (FP), true negative (TN), and false negative (FN), by the true classification situation of the target and the classification situation obtained by the model detection. The mean average precision (mAP), accuracy (Precision), recall (Recall), and the summed mean of accuracy and recall (F1) were used as the evaluation metrics for this experiment:

$$\text{Precision} = \frac{TP}{(TP + FP)},$$

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

$$F1 = \frac{\text{Precision} * \text{Recall}}{2 * (\text{Precision} + \text{Recall})}.$$

In the initial stage of weight model training, the training weights obtained by YOLOv5 based on the COCO dataset are used for migration training, which can improve the

convergence speed of the model for the flame smoke dataset, reduce the model training time, and improve the training results. The whole experiment lasts for 100 training rounds (epoch), the confidence (confidence) is set to 0.5, in the first 50 training rounds, the batch size is 8; in the last 50 training rounds, the batch size is 4, the data input size is 416×416 , and Adam is used as the optimizer.

3.3. *Comparison of Different Widths and Depths.* The same validation set was tested during experiments using the different algorithms mentioned above to detect combustion feature targets in electric vehicles. The results are shown in Table 3. The comparison of YOLOv5 detection results under different parameters is shown in Figure 10.

From the data in Table 3 and from Figure 10, it is known that when Depth_multiple and Width_multiple are 1.00 and 1.00, respectively, the CSP1_X structure with 3, 9, and 9 residual components, the CSP2_3 structure with 3 residual components, and the network with 64, 128, 256, 512, and 1,024 convolutional kernels in the CBL structure have the best feature extraction effect. It has the best detection effect and the highest accuracy for the intense smoke and flame burning phenomena generated by electric vehicle combustion.

3.4. *Comparison of Detection Effects of Different Backbone Feature Extraction Networks.* Based on the experiments in Section 3.3, YOLOv5-l was chosen as the basis for network improvement, and then the YOLOv5 algorithm with different backbone networks was used to compare and test the

TABLE 3: Comparison of YOLOv5 algorithm detection indicators under different parameters.

Network models	Precision		Recall		F1		mAP (%)
	Fire (%)	Smoke (%)	Fire (%)	Smoke (%)	Fire	Smoke	
YOLOv5-s	92.27	93.73	80.88	91.22	0.86	0.92	92.67
YOLOv5-m	89.96	93.85	83.07	93.13	0.86	0.93	93.77
YOLOv5-l	89.74	94.12	83.67	91.60	0.87	0.93	93.99
YOLOv5-x	83.48	90.28	76.49	85.11	0.80	0.88	87.08

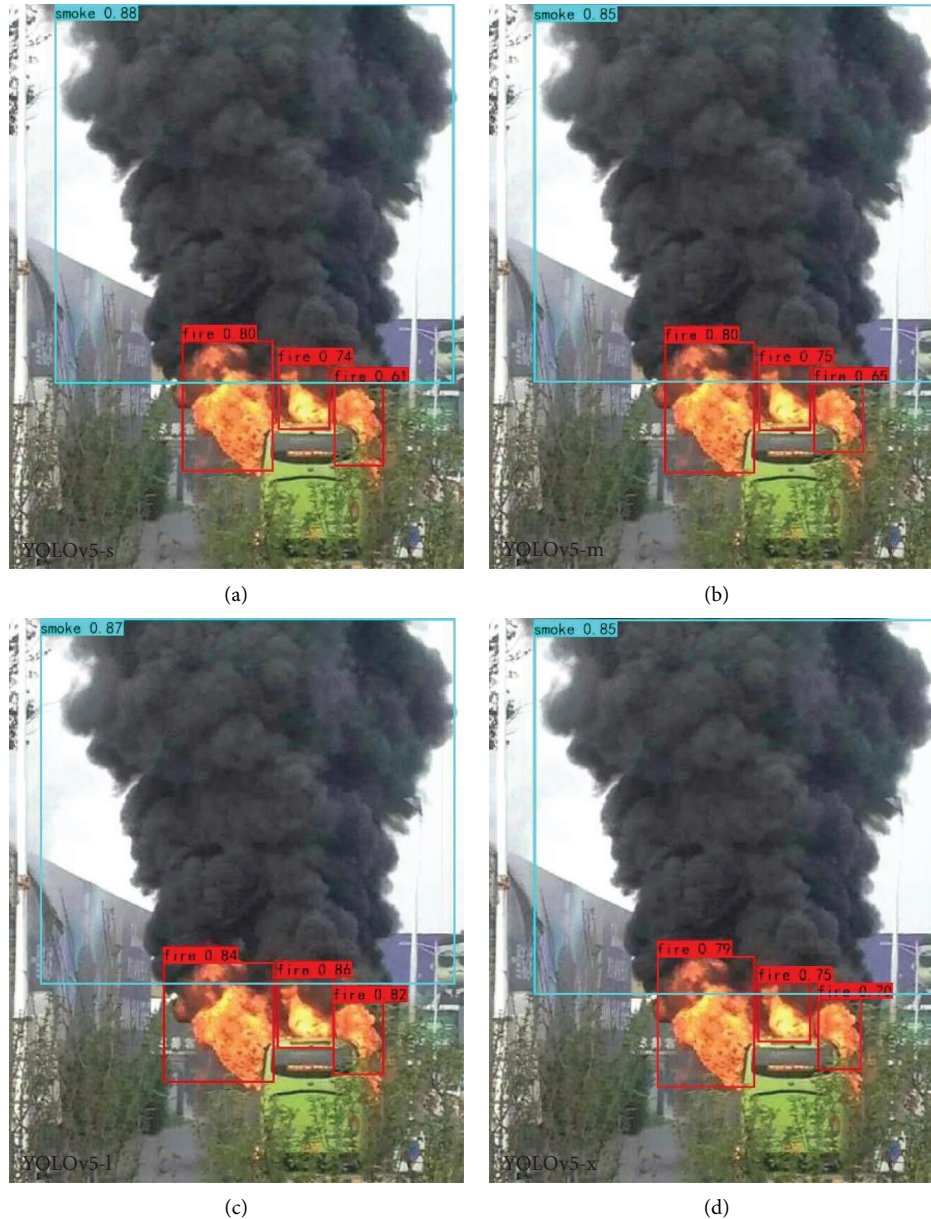


FIGURE 10: Comparison of YOLOv5 detection results under different parameters. (a) YOLOv5-s. (b) YOLOv5-m. (c) YOLOv5-l. (d) YOLOv5-x.

effect of feature extraction when the lithium battery of an electric vehicle burned. The comparison of YOLOv5 detection results under different backbone networks is shown in Figure 11.

According to the data in Table 4, the average FPS is only 10.86. When using CSPDarknet as the backbone feature extraction network, the mAP is only 2.11 percentage points lower, but the average FPS is significantly higher at 25.04,

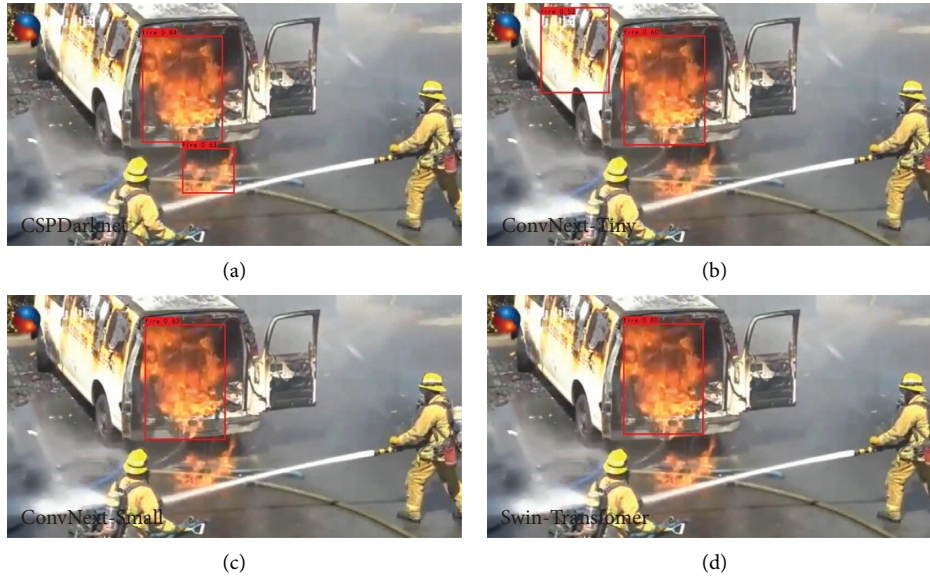


FIGURE 11: Comparison of YOLOv5 detection results under different backbone networks. (a) CSPDarknet. (b) ConvNext-Tiny. (c) ConvNext-Small. (d) SwinTransformer.

TABLE 4: Comparison table of evaluation index parameters of different backbone feature extraction networks.

Backbone network	Precision		Recall		F1		mAP (%)	FPS
	Fire (%)	Smoke (%)	Fire (%)	Smoke (%)	Fire	Smoke		
CSPDarknet	83.48	90.28	76.49	85.11	0.80	0.88	87.97	25.04
ConvNext-Tiny	88.54	95.63	55.38	66.79	0.68	0.79	83.19	27.58
ConvNext-Small	83.33	94.34	59.76	57.25	0.70	0.71	82.55	19.69
SwinTransformer-Tiny	88.44	92.95	70.12	80.53	0.78	0.86	90.08	10.86

which can extract more effective and comprehensive feature information and has better real-time detection performance.

In Figure 11, the YOLOv5 algorithm, which uses CSPDarknet as the backbone network, has the best detection performance, detecting smaller targets and being more resistant to interference than others.

3.5. Comparison of YOLOv5 Algorithms Using Different Attention Mechanisms. From the comparative experiments in Section 3.4, it is clear that the YOLOv5-l algorithm based on CSPDarknet as the backbone network has the best recognition results, so further improvements are made based on this. The addition of attention mechanisms allows the model to locate interesting information and suppress useless information. The commonly used attention mechanisms include SENet, CBAM, ECANet, etc. In this paper, we embed each of these three attention mechanisms at the same location to improve the detection accuracy during the experiment. Before the neck part, the three sizes of image feature information output from the backbone network are input into the attention mechanism module so that their features are given weight information. The YOLOv5 algorithm with the addition of CBAM increases the number of computational parameters by less than 0.01%, has the highest detection accuracy, and has the best detection effect on the features generated during the combustion of lithium

battery. The comparison of picture detection results using SENet, ECANet, and CBAM attention mechanisms is shown in Figure 12.

The data in Table 5 obtained by comparing the algorithm models based on different backbone networks, different residual structure techniques, and convolutional kernel bases and adding different attention mechanisms using the same dataset in the same experimental environment are shown in Table 5, which shows that the YOLOv5 algorithm based on the CSPDarknet backbone network with residual structure base 3, convolutional kernel base 64, and embedded CBAM is the best. The network model has the best results for monitoring electric vehicle charging safety.

3.6. Comparison of Improvement Results. Due to the rapid reaction of lithium batteries when burning, the temperature rises sharply, which can easily ignite the surrounding materials and cause the whole car to catch fire. Therefore, we have selected images of the relevant cars burning to conduct comparative experiments again and to test the universality and robustness of the algorithm.

Figure 13 shows a comparison of the detection results of the original, unimproved, and improved algorithms, respectively. The improved algorithm can more accurately mark the location of the flames and detect smoke features



FIGURE 12: Comparison of picture detection results using SENet, ECANet and CBAM. (a) SENet. (b) ECANet. (c) CBAM.

TABLE 5: Comparison table of evaluation index parameters of different backbone feature extraction networks.

Attentional mechanisms	Precision		Recall		F1		mAP (%)	Params
	Fire (%)	Smoke (%)	Fire (%)	Smoke (%)	Fire	Smoke		
SENet	91.81	92.95	84.86	92.75	0.88	0.93	93.88	47.229M
ECANet	89.57	94.16	82.07	92.37	0.86	0.93	94.00	47.057M
CBAM	88.51	95.31	82.87	93.13	0.87	0.94	94.09	47.401M

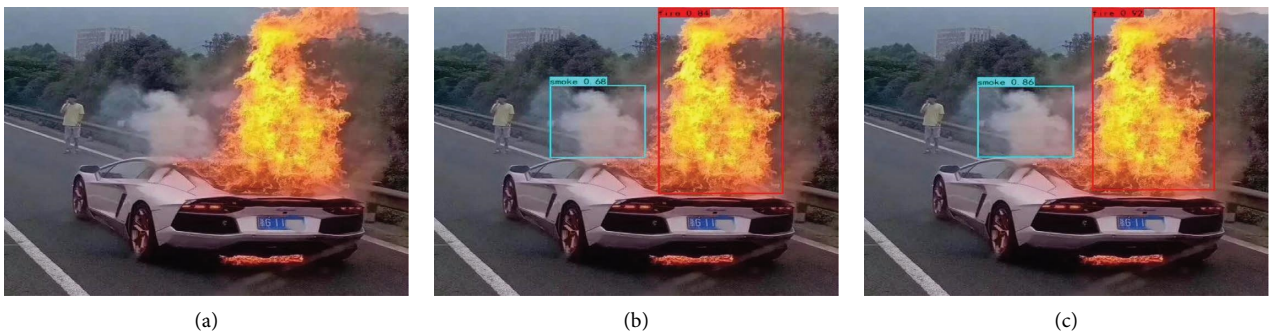


FIGURE 13: Continued.

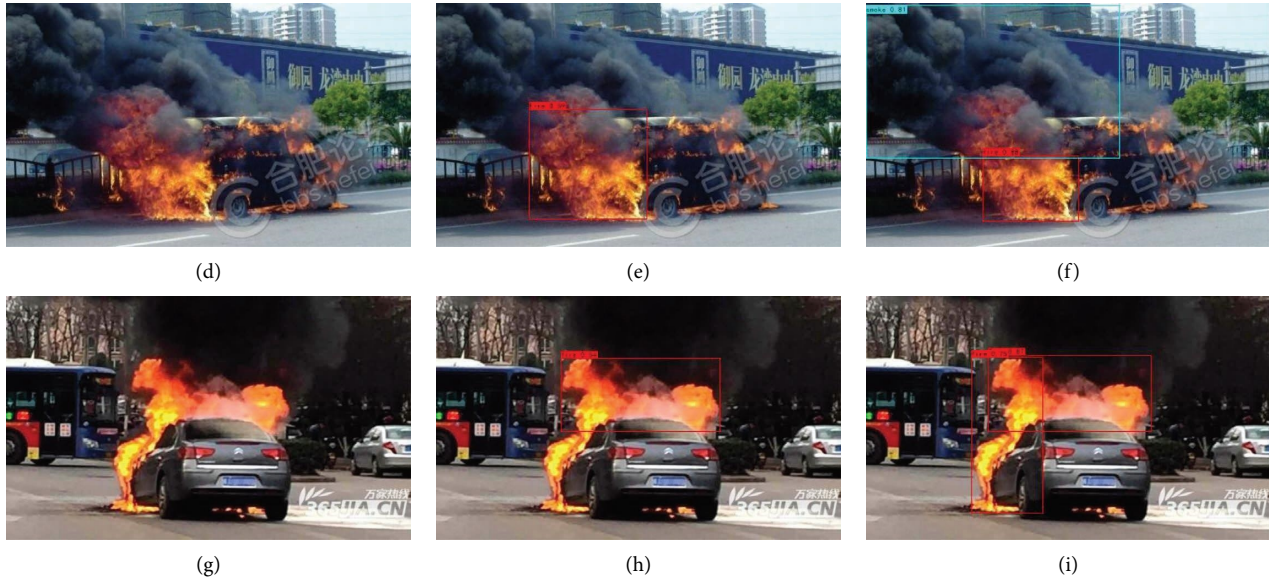


FIGURE 13: Comparison of the results before and after algorithm improvement.

that the original algorithm could not detect. The accuracy of detection is also significantly improved compared to the unimproved algorithm.

4. Conclusions

A feature target detection algorithm that realises the real-time monitoring of targets including flame and smoke in complex scenes of the charging process of electric vehicles is proposed for the potential safety issues with electric vehicle charging. In addition, the best target detection model for EV charging safety monitoring scenarios is derived after experimental comparison and analysis, and CBAM is added to the model to improve it. Finally, the enhanced algorithm mAP is able to surpass a number of well-known target detection algorithms in terms of evaluation index performance, detection accuracy (94.09%), anti-interference capability, and real-time performance. It can be inexpensively ported to mobile devices for real-time monitoring, providing a creative and practical solution for the safe operation of electric vehicle charging stations.

In addition, the algorithm can also be applied in the future to unmanned charging stations, the production, transport and storage of lithium battery, and other real-time safety monitoring scenarios, as well as the burning characteristics of lithium battery in public areas to provide security for lithium battery-related use scenarios.

Abbreviations

YOLO:	You Only Look Once
CBAM:	Convolutional block attention module
mAP:	Mean average precision
F1:	F1-score
FPS:	Frames per second
R-CNN:	R-convolutional neural network
RPN:	Region proposal network

SSD:	Single shot multibox detector
RepVGG:	Re-param visual geometry group
CSPNet:	Cross-stage partial network
PANet:	Path aggregation network
SPP:	Spatial pyramid pooling
SENet:	Squeeze-and-excitation networks
ECANet:	Efficient channel attention module networks
MLP:	Multilayer perceptron.

Data Availability

The data used to support the findings of this paper are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (grant no. U1806201), the Major Basic Natural Science Foundation of Shandong Province (grant no. ZR2021ZD12), and the Shandong Provincial Natural Science Foundation of China (grant nos. ZR2022ME194 and ZR2020MF087).

References

- [1] T. Shan, Z. Wang, J. Hong et al., "Overview of "Mechanical abuse-thermal runaway" of electric vehicle power battery and its safety prevention and control technology," *Journal of Mechanical Engineering*, vol. 58, no. 14, pp. 252–275, 2022.
- [2] Z. Xia, X. Pan, S. Song, L. E. Li, and G. Huang, "Vision transformer with deformable attention," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4794–4803, New Orleans, LA, USA, June 2022.

- [3] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 21–37, Springer, Amsterdam, The Netherlands, September 2016.
- [4] Y. Cui, J. Liu, B. Cong, X. Han, and S. Yin, "Characterization and assessment of fire evolution process of electric vehicles placed in parallel," *Process Safety and Environmental Protection*, vol. 166, pp. 524–534, 2022.
- [5] Y. Cui, B. Cong, J. Liu, M. Qiu, and X. Han, "Characteristics and hazards of plug-in hybrid electric vehicle fires caused by lithium-ion battery packs with thermal runaway," *Frontiers in Energy Research*, vol. 10, p. 431, 2022.
- [6] H. Li, W. Peng, X. Yang, H. Chen, J. Sun, and Q. Wang, "Full-scale experimental study on the combustion behavior of lithium ion battery pack used for electric vehicle," *Fire Technology*, vol. 56, no. 6, pp. 2545–2564, 2020.
- [7] Z. Niu, X. Jiang, B. Xie, and Y. Jin, "Study on simulation and safety protection of electric vehicle overcharge and explosion accident," *Transactions of China Electrotechnical Society*, vol. 37, no. 1, pp. 36–47, 2022.
- [8] Y. Xie, J. Zhu, Y. Guo, J. You, D. Feng, and Y. Cao, "Early indoor occluded fire detection based on firelight reflection characteristics," *Fire Safety Journal*, vol. 128, Article ID 103542, 2022.
- [9] T. Liu, J. Cheng, H. Honghu, X. Luo, and B. Cheng, "Video smoke detection method combining $Y^dU^aV^a$ color model and improved MobileNetV3," *Journal of National University of Defense Technology*, vol. 43, no. 5, pp. 80–85, 2021.
- [10] L. Du, Y. Xu, and W. Zhang, "Phased smoke detection algorithm using dual network fusion," *Journal of Xidian University*, vol. 47, no. 4, pp. 141–148, 2020.
- [11] J. Chen, H. Liao, X. Y. Huang, and C. M. Xie, "Importance of fractional exhaled nitric oxide in diagnosis of bronchiectasis accompanied with bronchial asthma," *Journal of Thoracic Disease*, vol. 8, no. 5, pp. 992–999, 2016.
- [12] H. Zhao, Q. Geng, F. Yu, and C. Wang, "New algorithm of flame detection based on color features," *Journal of Jilin University*, vol. 44, no. 6, pp. 1787–1792, 2014.
- [13] Y. Wu, Y. Song, and H. Zhou, "State identification of boiler combustion flame images based on gray entropy multiple thresholding and support vector machine," *Proceedings of the CSEE*, vol. 33, no. 20, pp. 66–73, 2013.
- [14] Y. Wu, L. Zhu, and H. Zhou, "State recognition of flame images based on krawtchouk moment and support vector machine," *Proceedings of the CSEE*, vol. 34, no. 5, pp. 734–740, 2014.
- [15] J. Huang, C. Chaoxia, X. Dong et al., "Faster R-CNN based color-guided flame detection," *Journal of Computer Applications*, vol. 40, no. 5, pp. 1470–1475, 2020.
- [16] Y. Zhao, J. Zhu, Y. Xie, W. Li, and Y. Guo, "A real-time video flame detection algorithm based on improved YOLO-v3," *Geomatics and Information Science of Wuhan University*, vol. 46, no. 3, pp. 326–334, 2021.
- [17] H. Wu, Y. Hu, W. Wang, X. Mei, and J. Xian, "Ship fire detection based on an improved YOLO algorithm with a lightweight convolutional neural network model," *Sensors*, vol. 22, no. 19, p. 7420, 2022.
- [18] Y. Cai, T. Luan, H. Gao et al., "YOLOv4-5D: an effective and efficient object detector for autonomous driving," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.
- [19] Y. Huo, Q. Zhang, Y. Jia et al., "A deep separable convolutional neural network for multiscale image-based smoke detection," *Fire Technology*, vol. 58, no. 3, pp. 1445–1468, 2022.
- [20] Z. Wu, R. Xue, and H. Li, "Real-time video fire detection via modified YOLOv5 network model," *Fire Technology*, vol. 58, no. 4, pp. 2377–2403, 2022.
- [21] X. Li, R. Te, F. Yi, and G. Xu, "TCS-YOLO model for global oil storage tank inspection," *Optics and Precision Engineering*, vol. 31, no. 2, pp. 246–262, 2023.
- [22] H. Wang, Y. Xu, Z. Wang, Y. Cai, L. Chen, and Y. Li, "CenterNet-Auto: a multi-object visual detection algorithm for autonomous driving scenes based on improved CenterNet," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 3, pp. 742–752, 2023.
- [23] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, Las Vegas, NV, USA, June 2016.
- [24] Y. Li, X. Zhang, and Z. Shen, "YOLO-Submarine Cable: an improved YOLO-V3 network for object detection on submarine cable images," *Journal of Marine Science and Engineering*, vol. 10, no. 8, p. 1143, 2022.
- [25] K. Liu, "STBi-YOLO: a real-time object detection method for lung nodule recognition," *IEEE Access*, vol. 10, pp. 75385–75394, 2022.
- [26] C. Dewi, R. C. Chen, X. Jiang, and H. Yu, "Deep convolutional neural network for enhancing traffic sign recognition developed on Yolo V4," *Multimedia Tools and Applications*, vol. 81, no. 26, pp. 37821–37845, 2022.
- [27] B. Yan, P. Fan, M. Wang, S. Shi, X. Lei, and F. Yang, "Real-time apple picking pattern recognition for picking robot based on improved yolov5m," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 53, no. 9, pp. 28–38, 2022.
- [28] H. Li, F. Kong, and Y. Lin, "Based on the improved YOLOv5s infrared ship detection algorithm," *Systems Engineering and Electronic Technology*, vol. 1, p. 10, 2022.
- [29] J. Yu, B. Liu, D. Yin, W. Gao, and Y. Xie, "Intelligent identification and measurement of bridge fractures based on yolov5 and u-net3+," *Journal of Hunan University*, vol. 50, no. 5, pp. 65–73, 2023.
- [30] S. Lang, X. Cao, Y. Zhang, R. Gao, and Y. Gong, "Automatic measurement method of image total station fusing improved YOLOv5 algorithm," *Chinese Journal of Instrumentation*, vol. 43, no. 5, pp. 120–127, 2022.
- [31] Z. Chen, R. Xiong, and F. Sun, "Analysis and research status of battery safety accidents in electric vehicles," *Journal of Mechanical Engineering*, vol. 55, no. 24, pp. 93–104+116, 2019.
- [32] Q. Zhang, T. Liu, C. Hao, Y. Qu, J. Niu, and Q. Wang, "Rapid detection and hazard method of thermal runaway gas in lithium-ion battery," *Journal of Beihang University*, vol. 49, no. 1, pp. 17–22, 2023.
- [33] P. Sun, R. Bisschop, H. Niu, and X. Huang, "A review of battery fires in electric vehicles," *Fire Technology*, vol. 56, no. 4, pp. 1361–1410, 2020.
- [34] S. Woo, J. Park, J. Lee, and I. S. Kweon, "Cbam: convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, Cham Switzerland, October 2018.