ISSN 1392-3730/eISSN 1822-3605

JOURNAL of CIVIL ENGINEERING and MANAGEMENT

2024
Volume 30
Issue 8
Pages 720–737
https://doi.org/10.3846/jcem.2024.21360

COMPUTER VISION BASED EARLY FIRE-DETECTION AND FIREFIGHTING MOBILE ROBOTS ORIENTED FOR ONSITE CONSTRUCTION

Liulin KONG[1], Jichao LI[2], Shengyu GUO[2]✉, Xiaojie ZHOU[2], Di WU[3]

[1] *Faculty of Engineering, China University of Geosciences (Wuhan), Wuhan, PR China*
[2] *School of Economics and Management, China University of Geosciences (Wuhan), Wuhan, PR China*
[3] *School of water conservancy, North China University of Water Resources and Electric Power, Henan, PR China*

Article History:
- received 17 November 2022
- accepted 20 March 2024

Abstract. Fires are one of the most dangerous hazards and the leading cause of death in construction sites. This paper proposes a video-based firefighting mobile robot (FFMR), which is designed to patrol the desired territory and will constantly observe for fire-related events to make sure the camera without any occlusions. Once a fire is detected, the early warning system will send sound and light signals instantly and the FFMR moves to the right place to fight the fire source using the extinguisher. To improve the accuracy and speed of fire detection, an improved YOLOv3-Tiny (namely as YOLOv3-Tiny-S) model is proposed by optimizing its network structure, introducing a Spatial Pyramid Pooling (SPP) module, and refining the multi-scale anchor mechanism. The experiments show the proposed YOLOv3-Tiny-S model based FFMR can detect a small fire target with relatively higher accuracy and faster speed under the occlusions by outdoor environment. The proposed FFMR can be helpful to disaster management systems, avoiding huge ecological and economic losses, as well as saving a lot of human lives.

Keywords: convolutional neural network, firefighting, fire accidents prevention, mobile robot, improved YOLOv3-Tiny model, construction sites.

✉Corresponding author. E-mail: *guoshy@cug.edu.cn*

1. Introduction

Construction sites are extreme hazardous due to its dynamic, temporary and decentralized nature (Li et al., 2015). Among the numerous risks and accidents facing the construction workers, fires are the most dangerous hazards and the leading cause of death in construction. According to the Campbell's Data Report, an average of 3840 fires occurred in buildings under construction, killing 4 people, hurting 49 more and causing the direct economic loss of 0.3 billion yuan every year from 2013 to 2017 (Campbell, 2020). In China, 252 thousand fires were reported in 2020, of which the number of fires on the construction sites accounted for about 1.1% (Management & Emergency, 2020), resulting in more than 296 people injured and 259 people died (Bosheng, 2022). The construction site is usually in a state of high fire risk, which is mainly due to the following two reasons. The first one is that there is lot of hot work (such as electric welding and cutting tasks, etc.) on the construction site, especially in the crossing areas when multi-tasks are underway, which is easy to ignite combustibles and cause fire (CE Safety, 2020; Fire Safety Matters, 2020). The second reason is the difficulty of fire control and timely safety management, due to the main challenges of early fire detection and confusion created in emergency response (Su et al., 2021). For instance, a fire accident took place at the construction site of a hotel (Palmer, 2012). The construction workers attempted to put out the fire initially by themselves, so they delayed notifying fire officials. By the time firefighters arrived, the fire has spread into a fully developed stage due to the wind. These highlight that lack of fire information and the bad decision making are critical threats to firefighting and rescue operations (Zhang et al., 2022).

To promptly report the fire while it is in an early stage is an important way to minimize the damage caused by a fire accident (Ahn et al., 2023). In the beginning, early fire detection and warning methods were done by using sensor equipment (such as infrared, ion, or optical sensors, etc.) and video surveillance (Qiu et al., 2018). Current fire alarm sensors capture flame or smoke with the help of light, temperature, smoke or gas signals to judge whether a fire occurs (Hong et al., 2019). However, the sensors (like temperature detectors, smoke detectors, thermal cameras etc.) have low sensitivity and poor prediction accuracy because of the detection distance and installation position, especially in large space environment. To cope with these limitations, numerous video-based fire detection methods have been proposed and applied in this field (Muhammad et al., 2018; Jiao et al., 2019; Wu et al., 2019; Yang et al., 2019; Xie et al., 2020; Xu et al., 2020), which are motivated by the encouraging advantages such as lower cost, larger coverage of surveillance area, less human interference and more available details like fire size, location, and degree of burning (Muhammad et al., 2018). Besides, the light generated by the combustion of flame propagates faster than the smoke or heat diffusion, which is more suitable for video fire detection. Despite these advantages, they still encounter some problems, e.g., the occlusion of hot work under scaffolding, the long warning distance between fixed cameras and operation areas, the received untimely fire alarming due to construction noise and the impossibility to put out the fire instantly. For example, fixed cameras may detect a fire at first, but they cannot put it out as there is no fire extinguishing devices equipped with. While the fire spreads quickly, which may cause serious consequences. Therefore, to control the fire with minimum losses, early detection of fire and a timely response (such as early warning and firefighting) are of paramount interest and helpful to the onsite safety management.

To address these challenges, mobile robotic systems capable of autonomous navigation may provide an alternative. In the past years, a variety of integrated robotic systems that fuse visual and sensor data have been applied in the construction industry, including: (1) cleaning robot, for instance, combining a façade cleaning robot with a floor cleaning robot (Vishaal et al., 2018) and developing portable and cheap window cleaning robot (Mir-Nasiri et al., 2018) to improve the façade cleaning robots; (2) inspection robot, for example, developing facade inspection robot to detect visible cracks (So et al., 1996) or assist wall inspection (Tso & Feng, 2003); (3) construction waste recycling robot, such as a vision-based robotic system for automatic nails and screws recycling (Wang et al., 2019) or on-site construction and demolition waste sorting and recycling (Wang et al., 2020); (4) robotic welding of medium-thickness plate (Geng et al., 2023); (5) robotic navigation for surveillance and security (Di Paola et al., 2010; Park et al., 2016); (6) robotic tunnel inspection and maintenance (Victores et al., 2011), etc. However, computer vision-based robots for fire detection in construction field remain poorly explored.

To address this issue, our research presents a prototype of firefighting mobile robot (FFMR) based on computer vision technology to detect, warn and handle fire accidents, where a YOLOv3-Tiny-S model (YOLO is short for You Only Look Once) based on YOLOv3-Tiny model is developed for an accurate fire detection. Once a fire is detected, the early warning system will send warning signals instantly and the FFMR moves to the right place to fight the fire source using the extinguisher. Thus, the FFMR can accurately handle the fire accident as early as possible. The main contributions of the proposed method are as follows:

(1) A dataset with various scenarios of construction fire for the fire detection area is obtained, which can be used to learn the important features for predicting accurately and over-coming over-fitting problems.

(2) A YOLOv3-tiny-based improved fire detection approach to achieve a real time fire detection on construction sites with no obvious decline on detection accuracy, which has successfully optimized the YOLOv3-Tiny with the trade-off between computational complexity and performance.

(3) A prototype of firefighting mobile robot (FFMR) based on computer vision technology is proposed to patrol the desired territory and will constantly observe for fire-related events to make sure the camera without any occlusions.

(4) The network structure of the proposed YOLOv3-Tiny model is optimized by introducing a Spatial Pyramid Pooling (SPP) module and refining the multi-scale anchor mechanism, which has more accuracy for the small fire detection than the other detection networks with comparable speed.

The rest of this paper is organized as follows. Section 2 reviews the current applications of robots on firefighting and computer vision-based fire detection. Section 3 elucidates the research methodology. Section 4 presents the results of a case study to validate the effectiveness and feasibility of our proposed approach. Section 5 discusses the limitations of the study and potential future work. Section 6 gives the conclusions.

2. Literature review

In this section, we first review the applications of mobile robots in firefighting. Then, we critically discuss the current computer vision-based methods reported in fire detection along with their strengths and weaknesses. During the process, we briefly highlight our approach to solving the problems of some of the current methods for early fire detection and prevention.

2.1. The applications of robots on firefighting

Fire detection and prevention is an important issue in the preservation of forests, crops and buildings (Roberto et al. 2013). At the beginning, many approaches based on stationary wireless sensor networks (WSN) were proposed to detect a fire occurrence (Hefeeda & Bagheri, 2007; Liu et al., 2009). But a large number of sensors will be needed to monitor a forest or a building, which attracts the use of mobile robots due to their mobility and large area coverage. Then, a growing number of mobile robots have successfully applied on firefighting. One of the future research directions is how to make robots more intelligent (Wang et al., 2017). Some studies focused on robot components such as designing protectable leg mechanisms of multi-legged robots for firefighting (Zhang et al., 2018). Moreover, various studies provided computer vision algorithms to support the control of robots, such as localizing the water spray in a pair of infrared cameras (McNeil & Lattimer, 2017), using infrared image feedback for fire horizontal position aiming (Zhu et al., 2020a), predicting falling position of jet trajectory in fire extinguishing (Zhu et al., 2020b) and accurately classifying fire, smoke, and their thermal reflections using thermal images (Kim et al., 2016). Besides, some studies explored the application of swarm intelligence on multi-robot systems of firefighting (Innocente & Grasso, 2019). Owing to the innovations on the hardware and software, several intelligent firefighting robots were designed in the existing studies. For instance, Madhevan et al. (2017) developed a wireless automatic fire fighting surveillance robot that could traverse autonomously in the hazardous environment and locate the victim. Ando et al. (2018) proposed a novel hose type robot, which could fly directly into the fire source via a water-jet. Li et al. (2019) developed a fire reconnaissance robot to offer important fire information to fire fighters. These robots integrated technologies such as ZigBee, thermal imaging and augmented reality.

In brief, the above-mentioned studies improved the intelligence of robots in several aspects, however, computer vision-based robots for fire detection in construction field remain poorly explored. Considering the complex environment on site (e.g., narrow space, unsafe scaffolds, severe occlusions, highly dynamic situations) (Edirisinghe, 2019), fire inspection and treatment at the initial stage is also a difficult task for robots. Therefore, the application of robots on construction firefighting should be further developed in this research.

2.2. Computer vision-based flame detection

Recent advances in computer vision technology have resulted in a variety of methods for fire detection. Existing studies on image/video-based flame detection mainly focus on improving accuracy and reducing false alarms. Some research used machine learning-based methods such as extreme learning machine (Prema et al., 2018) and Gaussian mixture model (Han et al., 2017; Li et al., 2021); however, the images/videos were pre-processed with hand-crafted features (e.g. color, texture) and the computational requirements were high with the added models. Then many studies have turned to convolutional neural networks (CNNs)-based deep learning methods since a deep CNN won the champion in the famous ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition in 2012 (Krizhevsky et al., 2017). CNN not only has the ability to perform feature extraction and classification within one network (Pincott et al., 2022), but also replace the hand-designed features and learn the complete characteristics of objects (Wu et al., 2019). Thus, numerous CNN-based video fire detection methods (including R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2016), improved Faster R-CNN (Chaoxia et al., 2020), etc.) have been proposed over the past years for better detection performance. Despite their great success in the fire detection, there are still some limitations during their practical applications. The aforementioned methods utilized region-based approaches which divide the object detection into two stages. At first stage, many region proposals are generated, and then each of them will be input to a CNN model for the classification and prediction of objects (Chen et al., 2022). Therefore, the two-stage methods will cost lots of time and computing resources (Wu et al., 2019). Moreover, these methods might lead to false alarms if there were fire-like objects.

To address these limitations, some one-stage methods including YOLO (You Only Look Once) (Redmon et al., 2016; Redmon & Farhadi, 2017), SSD (Single Shot Multibox Detector) (Liu et al., 2016), etc. for object detection are proposed, where the detection task is regarded as a regression problem and it achieves end-to-end target detection without complex pipeline. To name a few, Nguyen et al. (2021) developed a real-time fire detection and alarm system based on SSD algorithm. Zhan et al. (2023) proposed an improved SSD to detect a flame for the limited detection distance, delayed reaction and high false alarm rate of previous flame detection systems. Shen et al. (2018) detected flame by optimizing YOLO network model. Sridhar and Sathiya (2021) used YOLOv2 to detect early electrical fire. Zhang et al. (2021) combined the attention mechanism with YOLOv3 to improve the accuracy of flame detection. The YOLO algorithm has been constantly improved and the latest is YOLOv3, which uses the K-means clustering method to automatically select the best initial regression frame for the data set (Yi et al., 2019). To improve both speed and accuracy, many improved fire classification and detection models based on different YOLO network models have been developed (Wu et al., 2022; Yar et al., 2023; Chen et al., 2023). For instance, Li et al. (2022) proposed an improved YOLOv4-tiny model, which obtained a higher average detection accuracy than the YOLOv4-tiny, YOLOv5-s, and YOLOv7-tiny ones. Although the accuracies of these one-stage methods are a little lower than those of the two-stage methods, they can achieve faster detection speed. In general, YOLO achieves a higher mean average precision (mAP) than other real-time systems (Chang et al., 2010).

In summary, YOLOv3 and YOLOv4 stand out for their improved accuracy, especially for small objects, making them suitable for diverse applications. While YOLOv4 introduces advanced techniques, increasing computational requirements. Considering the real-time speed, we select YOLO method for the fire detection and will make improvements based on YOLOv3-Tiny model, then, a YOLOv3-Tiny-S model is proposed to meet a real-time fire detection with satisfied accuracy.

3. Research approach

Several techniques were developed to for the improved accuracy and speed of fire detection, especially for the small fires. Here, a brief overview of the proposed method is proposed as presented in Figure 1.

As illustrated in Figure 1, Firstly, the proposed FFMR patrolled the desired territory on the construction site and captured live video sources from its fixed camera. Secondly, the input images were resized to 416×416 pixels using the improved YOLOv3-Tiny model for fire detection. Meanwhile, the FFMR sent timely early warning and then put the fire out with the fire extinguishing if a fire was detected. The scopes of this study were (1) to optimize the YOLOv3-Tiny model with the trade-off between computational complexity and performance for onsite fire detection and (2) to design a prototype of intelligent FFMR, which can accurately detect fire, send early warning and put out the fire as soon as possible during its patrol inspections. To achieve these two goals, the detailed descriptions of our proposed methodology are given in the following subsections.

3.1. YOLOv3-Tiny-S-based fire detection

YOLOv3 is an improved version of YOLO series, evolved from the YOLO (Redmon et al., 2016) and YOLOv2 (Redmon & Farhadi, 2017) networks, which is more effective for detecting small targets (Redmon &Farhadi, 2018). The detection accuracy of YOLOv3 is relatively high, while their real-time performance on low performance devices or PCs is not ideal due to their complex network structure (Xiao et al., 2019). As a simplified model of YOLOv3, YOLOv3-Tiny reduces the depth of the convolutional layer and occupies less memory, which greatly improved the running speed. Figure 2 illustrates the basic network architecture of YOLOv3-Tiny, where its backbone network has only 7 convolutional layers and 6 pooling layers. As shown in Figure 2, a small number of 1×1 and 3×3 convolutional layers are used to extract the features, and the step sizes of the first 5 pooling layers and the last 1 pooling layer are 2 and 1 to achieve dimensionality reduction, respectively. For instance, the input image is 416×416 pixels and then the output one is 13×13 pixels via 5 maximum pooling layers. In the prediction part, the feature fusion method is adopted, which is detected by two feature maps with 13×13 pixels and 26×26 pixels, respectively.

We have tested the performance of YOLOv3-Tiny for fire detection on our proposed FFMR. Results show that although the simplified network improves the fire detection speed, the detection accuracy is much lower. Firstly, the feature extraction ability of fire is not ideal. And secondly, the detection performance for small fire targets is poor. To better deal with the trade-off between the accuracy and real-time performance of YOLOv3-Tiny, the following improvements are made in the proposed YOLOv3-tiny-S model: (1) optimizing the network structure to enhance the feature extraction ability of fire; (2) adding the Spatial Pyramid Pooling (SPP) module to improve the detection accuracy of fire; (3) adopting the multi-scale anchor mechanism to increase the detection accuracy of small fire; (4) using K-means clustering method to generate suitable priori bounding boxes. In this way, the accuracy of the proposed model detection will increase.

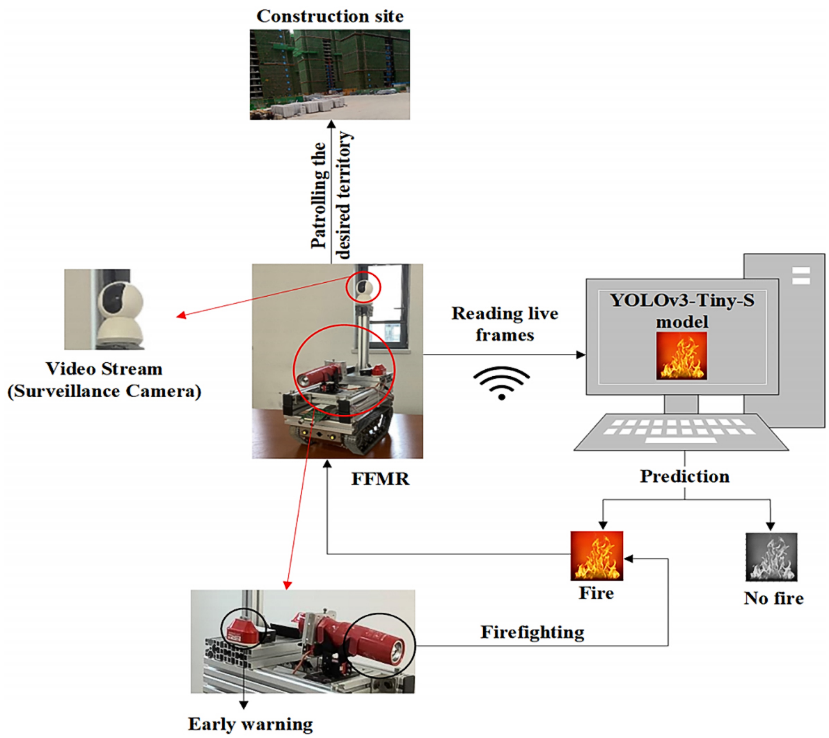

Figure 1. Overall process of the proposed method

Layer	Type	Filters	Size	Output
0	Convolutional	16	3×3/1	416×416×16
1	Maxpool		2×2/2	208×208×16
2	Convolutional	32	3×3/1	208×208×32
3	Maxpool		2×2/2	104×104×32
4	Convolutional	64	3×3/1	104×104×64
5	Maxpool		2×2/2	52×52×64
6	Convolutional	128	3×3/1	52×52×128
7	Maxpool		2×2/2	26×26×128
8	Convolutional	256	3×3/1	26×26×256
9	Maxpool		2×2/2	13×13×256
10	Convolutional	512	3×3/1	13×13×512
11	Maxpool		2×2/1	13×13×512
12	Convolutional	1024	3×3/1	13×13×1024

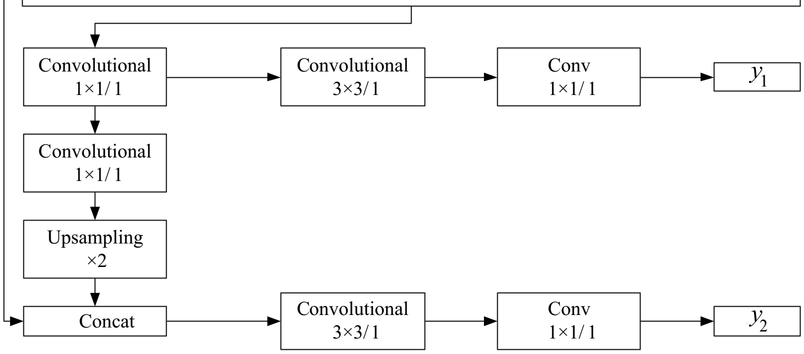

Figure 2. Network structure of YOLOv3-Tiny

3.1.1. Network structure optimization

The Max-pooling layer can reduce the image size and extract key information. However, the pooling window only retains the maximum feature of fires, which means some fire features are lost. In particularly, the complex background environment of the construction site may interfere with the flame detection. Therefore, the maximum pooling layer is replaced by a 3×3 convolutional layer with a step size of 2 to retain more fire features. By adding Batch Normalization (BN) on all convolutional layers, the vanishing gradient problem can be solved to some extent. Finally, BN is applied to the Leaky ReLu activations, as shown in Figure 3. Compared with the pooling layers, the convolutional layers can retain more fire features, which helps the network to further extract features. While the additional convolution layers will increase the scale of layer parameters, which may cause greater computation. Based on the network structure of ResNet, a 1×1 convolution kernel is introduced, which is applied to decrease the number of channels, reduce the parameters of convolution kernel, and then simplify the network model. In addition, it allowed for increasing the depth of the network and improving the representation ability of the model. Therefore, the 11th and the 12th convolutional layers in the network structure of YOLOv3-Tiny are respectively resized as 3×3 and 1×1 with a step size of 1. The output layers are also added 3×3 and 1×1 convolutional layers with a step size of 1 to improve the detection performance. The optimized network structure of YOLOv3-Tiny is shown in Figure 3.

3.1.2. Multi-scale prediction based on feature fusing

Feature fusion is one of the main methods to enhance feature information, and feature pyramid network is the most common one used to realize it. In classification/detection tasks, the feature extraction process is divided into different levels according to the depth of the network levels. Each level will generate feature maps of different scales and finally the features of each level are fused together, which are combined to form a feature pyramid. FPN (feature pyramid networks) adopted a top-down structure to combine the low-resolution but high-level semantic maps with the high-resolution but low-level semantic maps to get a feature pyramid that has rich semantics at all levels (Tesema et al., 2018) and make independent prediction on different levels of feature maps. The maps from different layers have features of different semantic levels and location information. The FPN-based methods make good use of features from different convolutional layers and have better ability to adapt to multi-scale object detection. The FPN network structure is shown in Figure 4.

In convolution neural network, the shallow network layer contains less feature semantic information, which makes it easier to locate the target location. The deep network layer has rich feature semantic information, but the target location information is rough. To make full use of feature information in the shallow network layer and improve detection performance of small target, this paper adopts the multi-scale module idea of FPN structure and adds a prediction scale to the 2-scale prediction on the

Layer	Type	Filters	Size	Output
0	Convolutional	16	3×3/1	416×416×16
1	Convolutional	16	3×3/2	208×208×16
2	Convolutional	32	3×3/1	208×208×32
3	Convolutional	32	3×3/2	104×104×32
4	Convolutional	64	3×3/1	104×104×64
5	Convolutional	64	3×3/2	52×52×64
6	Convolutional	128	3×3/1	52×52×128
7	Convolutional	128	3×3/2	26×26×128
8	Convolutional	256	3×3/1	26×26×256
9	Convolutional	256	3×3/2	13×13×256
10	Convolutional	512	3×3/1	13×13×512
11	Convolutional	256	1×1/1	13×13×256
12	Convolutional	512	3×3/1	13×13×512

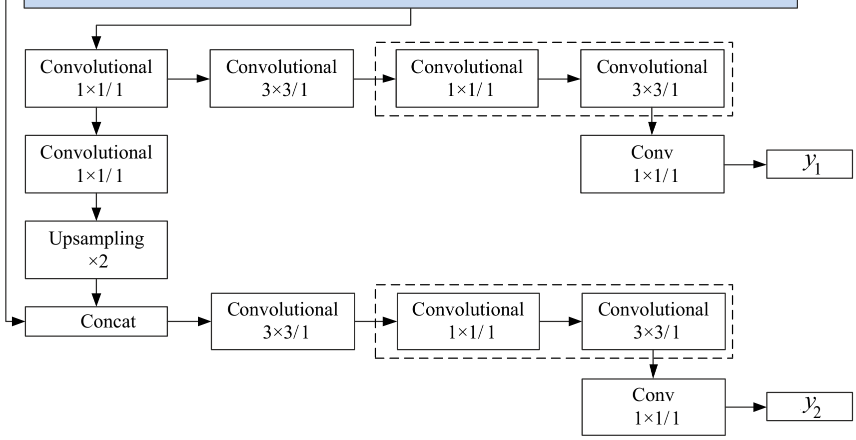

Figure 3. Optimization of YOLOv3-Tiny Network structure

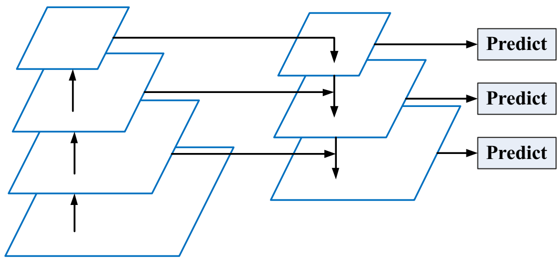

Figure 4. Network structure of the FPN

basis of the original model (as shown in Figure 5). When fusing the multi-scale feature maps, the output size needs to be consistent with each other. Thus, based on the second-branch predictive network, the dimension of the deep convolutional neural network is reduced by a 1×1 convolution kernel, twice sampled and then fused with the 6th convolutional layer, which will finally form three prediction scales of 13×13, 26×26 and 52×52 for detection of large, medium and small fire targets, respectively.

3.1.3. Spatial pyramid pooling (SPP)

SPP is a flexible solution for handling different scales, sizes, and aspect ratios (He et al., 2015). The SPP-net is a feature enhancement module, which can pool and concatenate feature maps of any size at multiple scales to improve the ability of feature expression and the robustness of detection. Due to the different distances between the cameras and the flame targets on construction site, the feature sizes of the flame image input into the convolution network are not consistent, thus the accuracies of the detections are reduced. Moreover, YOLOv3-Tiny connects the global features of different convolution networks, which ignores the multi-scale local feature interactions in the same convolution layer (Zhang et al., 2020). To solve these problems, we introduce the idea of SPP to replace the 12th convolutional layer of the YOLOv3-Tiny network to optimize the network structure. The structure of the SPP module is as shown in Figure 6. The input size can be ignored in the improved SPP. Three different scales: 5×5, 9×5 and 13×13 of the max-pooling layers are used to obtain the local feature map of fires, and then it is concatenated to the input features of SPP to obtain rich features. In this way, the local features are fused with the global ones and the expression ability of the featured maps is enriched, thus, the detection accuracy can be improved.

3.1.4. K-means anchor boxes

Appropriate prior boxes can guarantee the detection effects and speeds. K-means clustering algorithm in YOLOv3-tiny obtained 6 prior boxes on the Common Objects in Context (COCO) dataset according to the 80 annotated

Layer	Type	Filters	Size	Output
0	Convolutional	16	3×3/1	416×416×16
1	Convolutional	16	3×3/2	208×208×16
2	Convolutional	32	3×3/1	208×208×32
3	Convolutional	32	3×3/2	104×104×32
4	Convolutional	64	3×3/1	104×104×64
5	Convolutional	64	3×3/2	52×52×64
6	Convolutional	128	3×3/1	52×52×128
7	Convolutional	128	3×3/2	26×26×128
8	Convolutional	256	3×3/1	26×26×256
9	Convolutional	256	3×3/2	13×13×256
10	Convolutional	512	3×3/1	13×13×512
11	Convolutional	256	1×1/1	13×13×256
SPP				

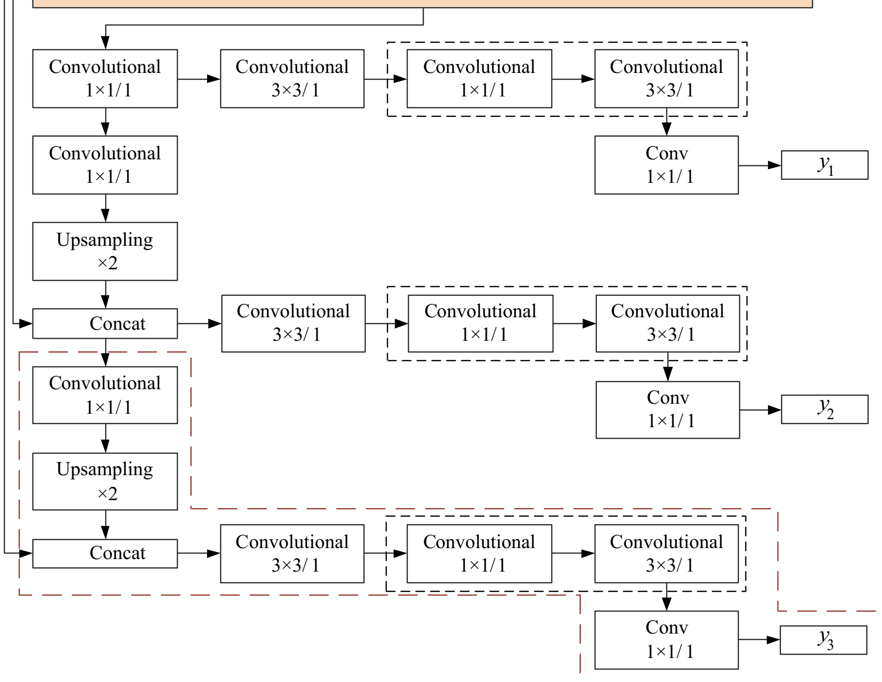

Figure 5. Network structure of YOLOv3-Tiny-S

ground truth boxes. However, due to their various target categories of the original dataset and generalized anchors, it is difficult to obtain accurate target fire information, which is not suitable for fire detection scene on construction sites. Therefore, we run K-means clustering on the fire training set to find the most suitable anchor boxes and the corresponding coordinates for each bounding box (Wu et al., 2019). Generally, the K-means algorithm uses the Euclidean distance to calculate the distance between data objects and cluster centers. However, if the Euclidean distance is calculated by anchors, calculating a large anchor box will produce more errors than the small prediction box. Then we adopted Intersection over Union (IOU) score to evaluate the clustering result (Yi et al., 2019). The purpose of clustering is to obtain greater IOU between the prior boxes and ground truth boxes, thus, the clustering distance in the K-means algorithm is modified as Eqn (1):

$$d\left(\text{box, centroid}\right) = 1 - \text{IOU}\left(\text{box, centroid}\right). \tag{1}$$

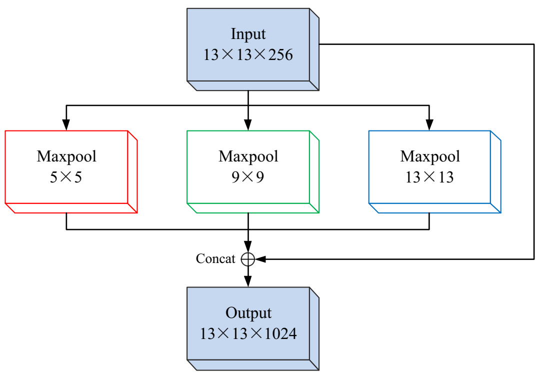

Figure 6. Structure of the SPP module

In Eqn (1), d (box, centroid) represents the clustering distance, centroid represents as the center of mass by the algorithm, box represents the other bounding boxes and IOU represents the ratio of the intersecting area of the two boxes to the combined area. Figure 7 shows the results of K-means clustering for various values of k (number of anchor boxes) and plots the average (AVG) IOU with closest centroid at k = 9. As illustrated in Figure 7, with the increase of k values, AVG IOU is increasing quickly and then it gradually levelled off when k is larger than 9. The larger value of k could improve the detection accuracy and speed, while it will also increase the computational cost of the model. Considering the trade-off between computational complexity and speed, the k value in the K-means algorithm is chosen to be 9. Thus, we selected 9 clusters and then divided them evenly across 3 different scales.

Since the initial value of the clustering center in the K-means algorithm affects the clustering effect, multiple clustering operations are carried out when the number of clustering centers k is 9, as shown in Figure 7b. When obtaining the highest value of the curve, that is, the corresponding AVG IOU value is 72.72%, the nine priori boxes are selected as shown in Table 1.

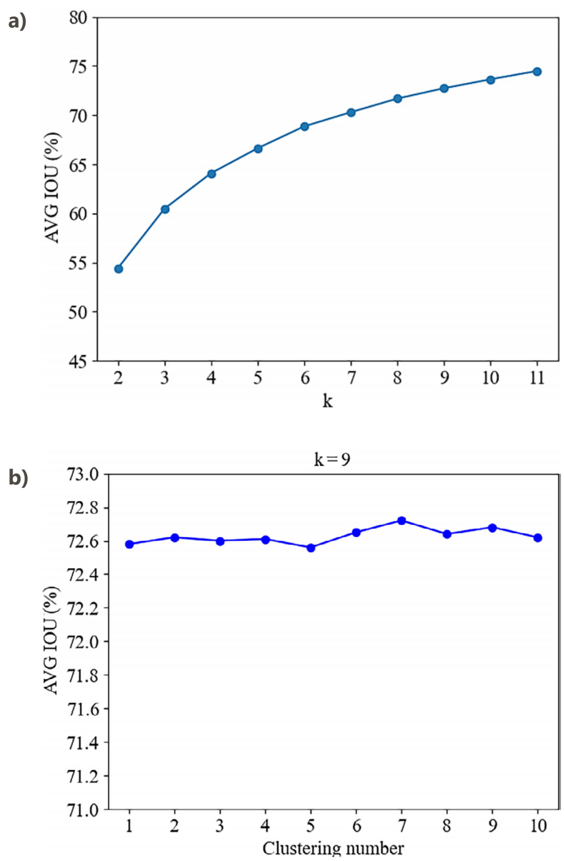

Figure 7. Results of K-means clustering: a – the relationship between k and Avg IOU; b – fluctuation trend of Avg IOU under k = 9

Table 1. The corresponding relationship between feature maps and prior boxes

Feature map	Receptive fields	Prior boxes		
13×13	Large	(82, 88)	(89, 169)	(190, 240)
26×26	Medium	(37, 80)	(51, 124)	(55 ,60)
52×52	Small	(18, 32)	(24, 56)	(34, 43)

3.2. Prototype of FFMR

The main function of the FFMR is to detect the potential fire targets when there is no patrol on the construction site and give timely and accurate early warning when the fire is detected using video sensors. Meanwhile, the FFMR moves to the right location and then puts the fire out with the fire extinguishing. When the fire in the fire extinguishing area cannot be detected, it indicates that the flame is extinguished. FFMR can eliminate fire hazards at its source, which helps managers to strengthen fire safety management on the construction site and avoids secondary damage to the working environment. The designed and physical pictures of FFMR are presented in Figure 8.

The proposed FFMR consists of two parts: the mobile platform control system and robot command system, as shown in Figure 9. The hardware and software configurations of the mobile platform control system contain robot body, crawler walking mechanism, control module, lighting module, rectangular support frame, fire extinguishing module, video/image acquisition module, alarm device and sensor module, etc. The robot command system includes user interface (as shown in Figure 9), which can realize real-time display, wireless transmission and reception of information, etc.

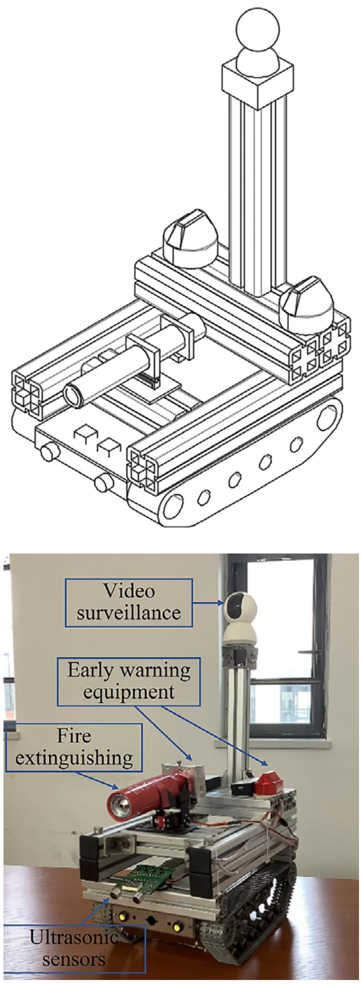

Figure 8. The designed and physical pictures of FFMR

The proposed FFMR system consist of four key steps: (1) YOLOv3-Tiny-S-based fire detection; (2) early fire warning; (3) distance measurement; and (4) fire extinguishing. Then the procedure of implementation of our proposed FFMR system is as illustrated in Figure 10. Firstly, the video/image acquisition module of FFMR system initializes the spherical camera mounted on the firefighting robot, which is used to collect the onsite video/image in real time. Meanwhile, the collected video/image is transmitted to the module of FFMR system through digital signal and detected using the proposed YOLOv3-Tiny-S model in this paper. Afterwards the detected results are returned to the mobile platform control system. Secondly, the control module analyzes the results, and sends both light and sound alarms once a fire is detected. At the same time, the firefighting robot stops moving forward. Thirdly, the distance between the fire and the firefighting robot is measured by the ultrasonic sensors mounted in front of the firefighting robot, which subtracts the range of fire extinguisher to obtain the distance that the firefighting robot should move forward. Finally, when the firefighting robot moves in the range of the fire extinguisher, it stops moving forward and the safety manager remotely control the firefighting robot to open the fire extinguisher to put out the fire.

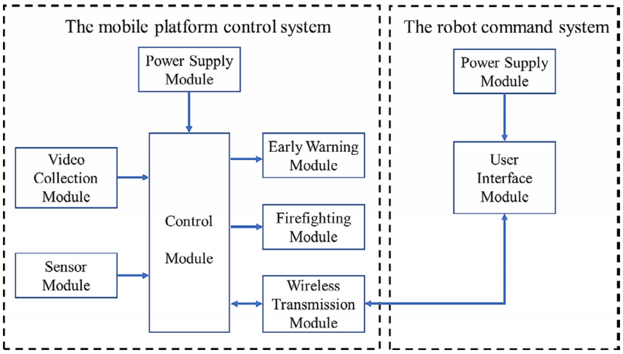

Figure 9. The general structure block diagram of FFMR

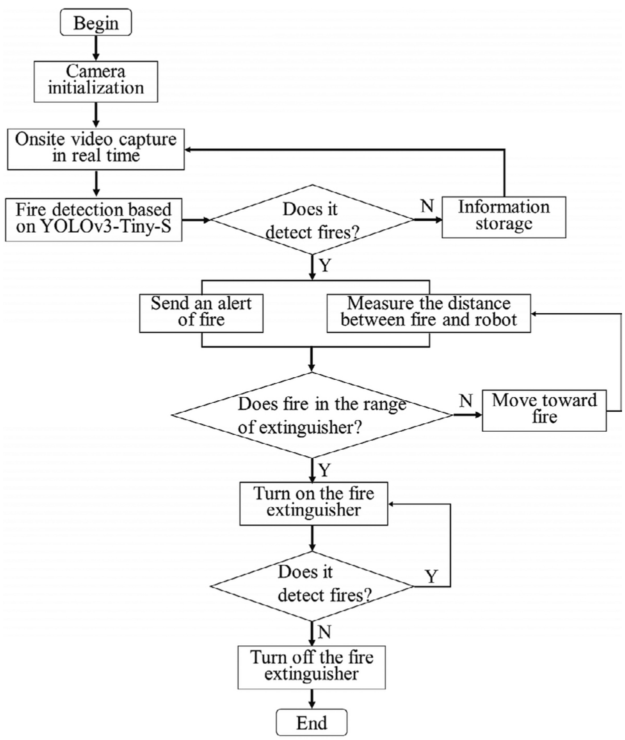

Figure 10. Workflow of FFMR

4. Experiments & results

4.1. Fire detection model training

4.1.1. Data collection

The image data used in this paper was collected from: (1) videos and images of fire accidents on the construction sites and (2) simulated fire scenes on construction sites. 3880 images of fires were initially collected according to the same time interval. In addition, 925 fire-like images were also collected, such as sunlight and night lights at the construction site. CNNs usually require a lot of data for training due to the large number of parameters needed to properly tune these networks. Insufficient training images can increase the risk of overfitting in the last fully connected layers, resulting in the performance degradation of the model in fire detection. Thus, these images were then expanded to 9567 images using data augmentation methods like flipping, mirroring, and adding noise, etc. as shown in Figure 11. Among them, 7653 images (about 80% of the total number) were randomly selected for training, the other 957 ones (10%) for verification and the rest 957 ones (10%) for testing, where the weights were saved after the training of the detection model.

4.1.2. Experimental platform

The experiments in this study were performed on the Ubuntu18.04 LTS. The video training environment was deployed on the Intel(R) Xeon(R) E5-2678 CPU, 32 GB of memory and TITAN X GPU, 12 GB of memory. The testing environment was deployed on the Intel Core i7-9750 CPU, 8 GB of memory and GTX1650 GPU, 4 GB of memory. CUDA10.2 and CUDNN7.6.5 were used to accelerate training. Under the Darknet deep learning framework, the program was written in Python 3.7.

4.1.3. Evaluation metrics

To verify the effectiveness of the improved model, Precision (P), Recall (R), Mean Average Precision (mAP) and F1 scores are used as evaluation parameters for detection accuracy and Frame Per Second (FPS) is adopted as evaluation parameters for detection speed (He et al., 2019). The "Precision" is the ratio of the number of correctly detected fires to the total number of detected ones, which is calculated as shown in Eqn (2). The "Recall" (recall rate) is the ratio of the number of correctly detected fires to the total number of fires in the data set and its calculation method is shown in Eqn (3). The "mAP" is the Mean Average Precision of all clusters, which is calculated as shown in Eqn (4).

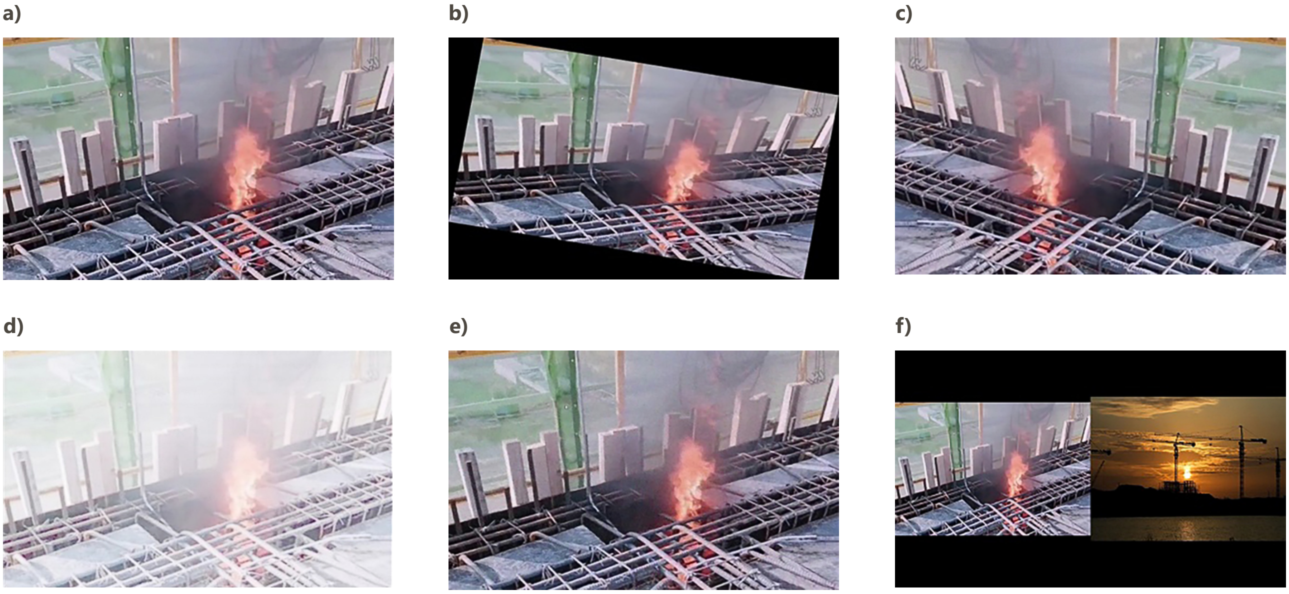

Figure 11. Data augmentation methods: a – original image; b – flipping; c – mirroring; d – brightening; e – adding noise; f – splicing

The "F1 score" is adopted as a trade-off between the recall and precision, and its definition is shown in Eqn (5).

$$\text{Precision} = \frac{TP}{TP+FP}; \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN}; \tag{3}$$

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i; \tag{4}$$

$$F1 = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}, \tag{5}$$

where *TP* (true positive) is the number of fires which have been correctly detected; *FP* (false positive) is the number of some other objects but detected as fires; *FN* (false negative) is the number of fires which are failed to be detected in the image; AP refers to the value of Average Precision; and *N* is the number of predicted target categories.

4.1.4. Training

The detailed network training parameters are set as follows: (1) the batch size is 64; (2) the number of subdivisions is 8; (3) the width and height are both 416; (4) the initial learning rate is 0.001; (5) the momentum coefficient is 0.9; and (6) the maximum iteration is 30000. In 24000~27000 iterations, the initial learning rate is multiplied by 0.1 times; and in 27000~30000 iterations, 0.1 times the current learning rate. During the training process, the input size of the model is adjusted every 10 iterations to ensure that the model has a good detection performance on images of different sizes.

The loss of YOLOV3-Tiny and YOLOV3-Tiny-S during training is shown in Figure 12. As illustrated in Figure 12, with the increase of training steps, the losses for both models show a decreasing trend. While YOLOv3-Tiny-S has faster convergence speed and better convergence results than YOLOv3-Tiny during training. The loss curve for YOLOv3-Tiny-S began to saturate after 5000 training steps and the final loss is around 0.3, which is lower than the YOLOv3-Tiny model. This shows that the performance of the proposed model is improved.

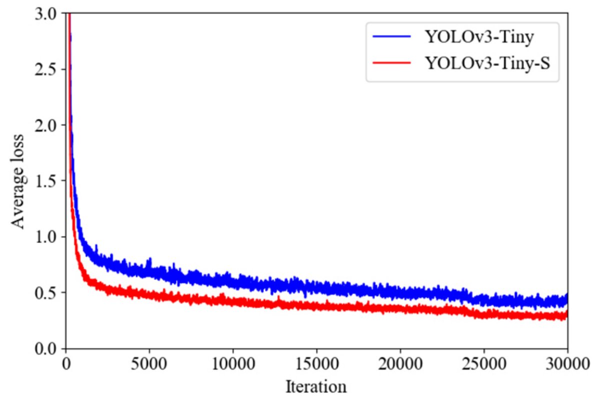

Figure 12. Loss curves of the two YOLO models

4.2. Analysis of experimental data

4.2.1. Results of fire detection

To validate the performance of the improved YOLOv3-Tiny model, other different modifications based on YOLOv3-Tiny models were evaluated for comparison. For convenience, incorporation of only the network architecture modification is called method "A"; while incorporation of the network architecture modification, three-scale prediction and K-means clustering is called method "B". The test set consisted of 957 images. Table 2 shows the Recall, Precision, F1 score, and AP of different methods, where "A" and "B" denote two different optimized models.

Table 2 shows that incorporation of network architecture modification in method "A" brought a slight rise of the Precision, Recall and mAP, consequently resulting in

Table 2. The detection results of different optimized models

Methods	YOLOv3-Tiny	A	B	YOLOv3-Tiny-S
Network architecture modification	–	√	√	√
Three-scale prediction + K-means clustering	–	–	√	√
Adding SPP module	–	–	–	√
Precision	0.90	0.91	0.92	0.92
Recall	0.80	0.82	0.81	0.82
mAP (%)	87.32	88.17	89.12	90.10
F1 score	0.85	0.86	0.86	0.87
Weight (MB)	34.7	30.3	31.6	27.4

Note: "–" denotes "without such modification"; "√" denotes "with such modification".

an improvement of the F1 score from 0.85 to 0.86 and a 12.68% decrease of the weight. This demonstrates the replacement of the maximum pooling layer with the convolution layer and the adding a small number of 1×1 convolutional layer with the step size of 1 played a role. Compared with method "A", further incorporation of three-scale prediction and K-means clustering increased the mAP to 89.12% and the weight enhanced with 4.29%. This is because the increase of multiple convolution layers leads to the increase of computation cost. When adding SPP module in YOLOv3-Tiny-S model, the local and global fire features were fused, resulting in the smallest weight with only 27.4 MB and better accuracy. The Precision, mAP, F1 score, and weight were improved with each modification. The results illustrate that the YOLOv3-Tiny-S model had better performance than YOLOv3-Tiny and the other YOLOv3-Tiny based models as listed in Table 2. Compared with the original YOLOv3-tiny, YOLOv3-Tiny-S had advantages on model performance and complexity, which improved Precision, Recall, mAP and F1 score by 2.22%, 2.50%, 2.78% and 2.35%, respectively, while decreasing the number of model parameters by 21.04%. Figure 13 shows the P-R curves of the original YOLOv3-tiny and YOLOv3-Tiny-S models on the test set. This demonstrates that the improved YOLOv3-tiny-s model is effective and better than the original YOLOv3-tiny model.

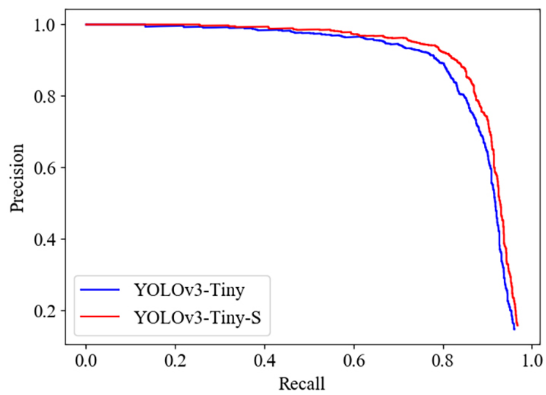

Figure 13. P-R curves of fire detection of original YOLOv3-tiny and YOLOv3-Tiny-S models

To further validate the performance of the improved YOLOv3-tiny model, other detection models using one-stage method were evaluated for comparison, as listed in Table 3. As shown in Table 3, the calculated Precision, Recall, mAP and F1 score of YOLOv3-Tiny-S model were second only to those of YOLOv3 model. However, the YOLOv3 model uses a deeper convolutional model and three size layers to predict the detection object, which requires a powerful GPU with more than 4 GB memory so that their real-time performance on low performance devices or PCs is not ideal. The YOLOv2-Tiny, YOLOv3-Tiny and YOLOv4-Tiny models had a slightly higher speed than the YOLOv3-Tiny-S model, but their detection accuracies are reduced. YOLOv2-Tiny scored the lowest Precision, mAP and F1 score, followed by SSD model. And SSD model presented the next lowest FPS, which is only second to that of YOLOv3 model. Generally speaking, the proposed YOLOv3-Tiny-S model has better detection performance.

Table 3. A comparison of different fire detection methods

Methods	Precision	Recall	mAP (%)	F1 score	FPS
YOLOv2-Tiny	0.72	0.84	83.61	0.77	45.33
YOLOv3-Tiny	0.90	0.80	87.32	0.85	46.90
YOLOv3	0.93	0.86	90.17	0.89	7.60
YOLOv4-Tiny	0.91	0.82	89.93	0.86	40.15
YOLOv3-Tiny-S	0.92	0.82	90.10	0.87	34.55
SSD	0.87	0.80	84.03	0.83	22.15

Image frame examples are presented in Figures 14–17. First, the K-means clustering is used on the training set bounding boxes to automatically find the best prior anchors boxes in the YOLOv3-Tiny-S model, which can better locate the fire's position as displayed in Figure 14d. As shown in Figure 14, the four YOLOv models scored higher accuracies than the SSD model. We observed from Figure 15 that all the five models can detect the outdoor fire with 99% accuracy. Secondly, dataset with fire and without fire is selected for testing. This is because there are many fire-like objects and situations (such as sunshine), which may be predicted as fire, making the classification more difficult. As shown in Figure 16, all the five models can differentiate between real fire and scenes with sunshine. YOLOv2-Tiny ranked highest in detection with 99% accuracy. YOLOv4-Tiny achieved 96%, followed by YOLOv3-Tiny-S, YOLOv3-Tiny and SSD. Finally, dataset with small fire under occlusion is selected for testing.

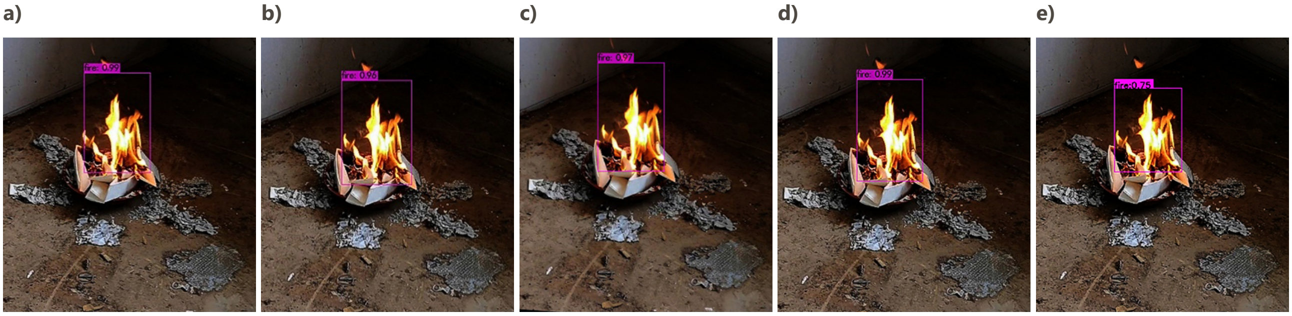

Figure 14. Fire detection results of the five models in indoor environment: a – YOLOv2-Tiny; b – YOLOv3-Tiny; c – YOLOv4-Tiny; d – YOLOv3-Tiny-S; e – SSD

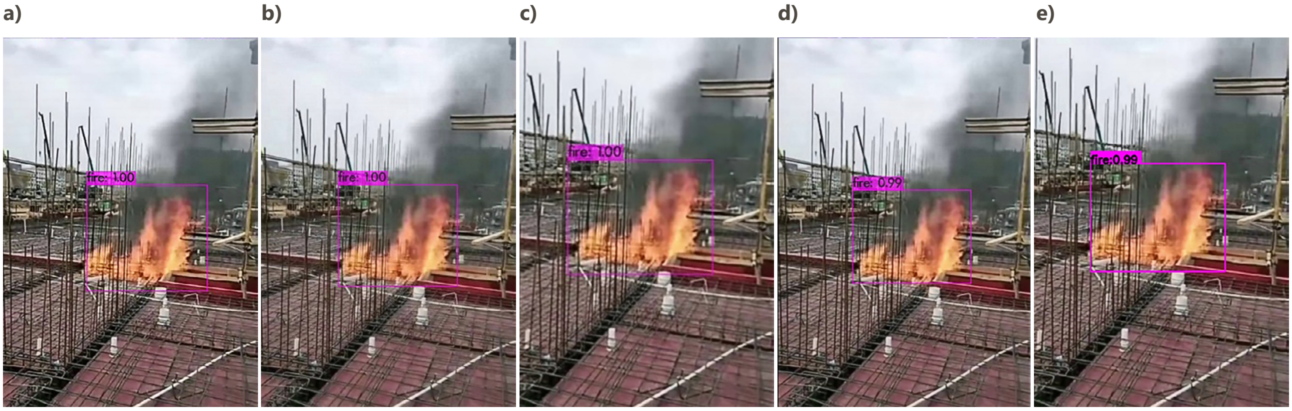

Figure 15. Fire detection results of the five models in outdoor environment: a – YOLOv2-Tiny; b – YOLOv3-Tiny; c – YOLOv4-Tiny; d – YOLOv3-Tiny-S; e – SSD

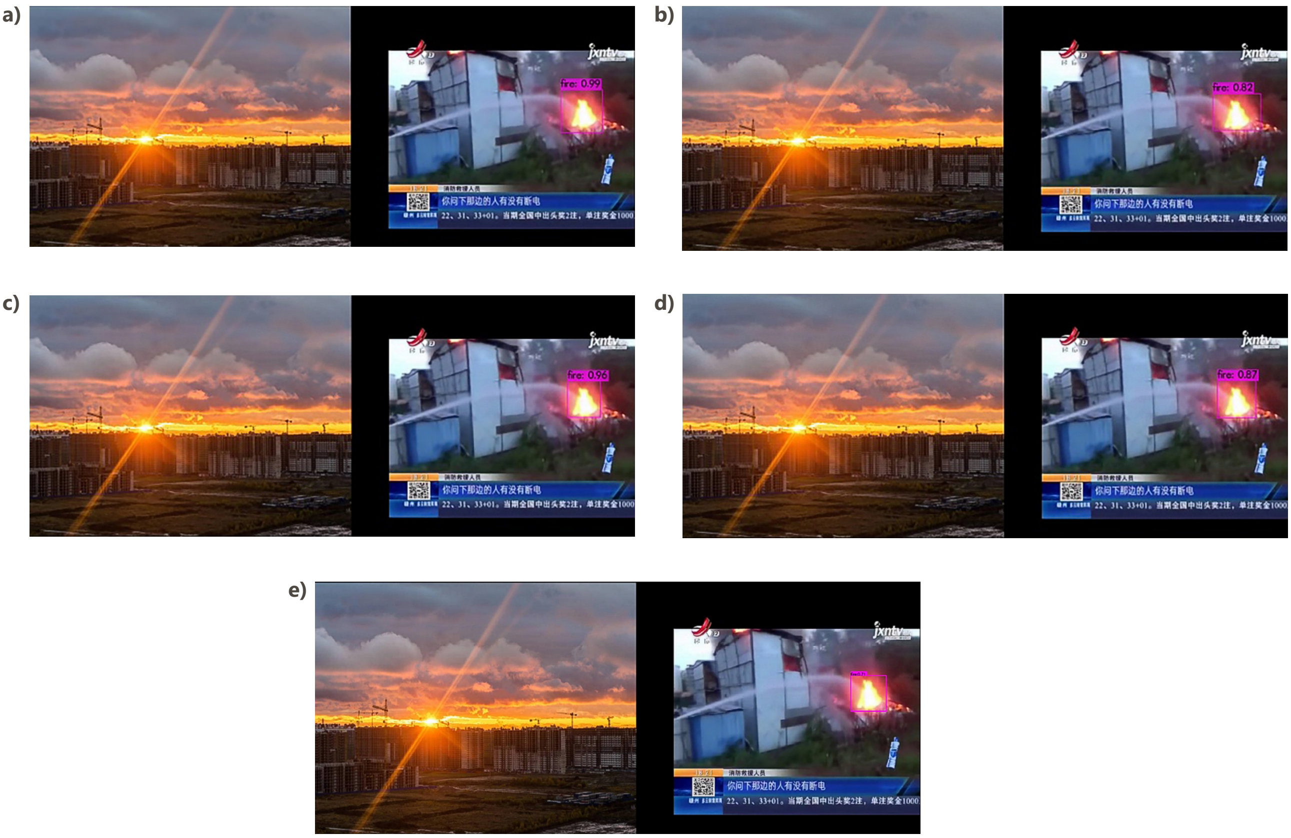

Figure 16. Detection results of the five models for real fires and fire-like objects: a – YOLOv2-Tiny; b – YOLOv3-Tiny; c – YOLOv4-Tiny; d – YOLOv3-Tiny-S; e – SSD

Figure 17. Fire detection results of the five models under outdoor occlusion: a – YOLOv2-Tiny; b – YOLOv3-Tiny; c – YOLOv4-Tiny; d – YOLOv3-Tiny-S; e – SSD

As illustrated in Figure 17, both YOLOv3-Tiny-S model and YOLOv3-Tiny model can effectively detect the small fire under certain occlusion, while the accuracy of the former is higher than that of the latter. YOLOv2-Tiny can detect the small fire under certain occlusion, but its positioning accuracy is poor. Neither YOLOv4-Tiny nor SSD models can effectively detect the small fire. This indicates that YOLOv3-Tiny-S model has the best detection performance for small fires under occlusion.

4.2.2. Application of FFMR

To test the real-time fire detection and early warning performance of the proposed FFMR, a video of fires under a safe environment on the construction site was tested with the proposed YOLOv3-Tiny-S model, which is integrated into the software end of the camera module. According to the hot work approval system of the construction site in China, the relevant personnel must apply for approval and hold the hot work certificate before hot work. Due to the strict management of hot work on the construction site, it is difficult for the robot to capture the fire in the test process. In addition, it is safer to carry out the experiment in the area outside safe working distance than in the operation area. Therefore, a safe environment away from the safe operation area on the construction site is selected to simulate the scene of fire. The video stream was collected with a resolution of 1280×720. The implementation process was presented in Figure 18.

As shown in Figure 18a, in the human-computer interaction interface of the robot command system for the real-time fire detection, the small target is detected as fire with accuracy of 93% and framed by a red box. The early fire warning is shown in Figure 18b. When the FFMR detects the fire, the alarm light (as displayed in the yellow circle)

Figure 18. The fire detection and early warning performance of the proposed FFMR: a – fire detection; b – early warning of fire; c – fire fighting; d – fire out

flashes and rings. Figure 18c shows the place where the robot starts the micro fire extinguisher to extinguish the flame; Figure 18d shows that the fire was out (as shown in the red circle), then the fire extinguishing and the alarm device system stopped working (as framed in the yellow circle). The results show that the proposed YOLOv3-Tiny-S has good detection performance of fires, and meets the requirement of real-time, which can be used in FFMR. Beside this, through the FFMR to view the patrol situation of the construction site in real time, the early-warning system will send light and sound signals once the targets are detected as fire, and then the fire extinguishing device is remotely controlled to deal with the fire at its early stage. Therefore, the proposed FFMR can be helpful to disaster management systems, avoiding huge ecological and economic losses, as well as saving a lot of human lives.

5. Discussions

Fire detection and its early warning are an important part in the fire safety management of the construction site. However, the traditional fire detection and early warning equipment on the construction site are difficult to meet the requirements of the new social environment. For example, the sensor equipment is easy to be affected by the detection distance and installation position, resulting in the poor accuracy of early warning. The fixed camera has problems such as large monitoring blind area, long early warning distance and inability to carry out early fire extinguishing. To overcome the above shortcomings, the increasing maturity of computer vision technology and the continuous emergence of multi-functional firefighting robots provide new solutions. Therefore, this paper proposed a FFMR system by combining the computer vision technology and the firefighting robot, which can realize the functions of onsite inspection by robot, fire detection, early warning and early fire extinguishing. The proposed FFMR can help prevent the frequent occurrence of fire accidents on construction sites and assist managers to strengthen onsite fire management. The main contributions of this study are as follows:

(1) This research proposed a YOLOv3-Tiny-S model for onsite fire detection due to the limitations of the YOLOv3 series, which can detect the fire with higher accuracy in varying indoor and outdoor environments with occlusion. Firstly, because of the lack of open data sets of fire videos and images on the construction site, this paper constructs the fire dataset through online acquisition and field simulation, and uses a variety of data enhancement technologies to expand its number to meet the requirements of network training. Based on YOLOv3-Tiny model, by optimizing the network structure, multi-scale prediction, adding the SPP module and K-means clustering, the mAP of fire detection by YOLOv3-Tiny-S reached 90.10%, and the model scale decreased by 21.04%. This indicates the improved model is effective. Finally, compared with other models, the calculated Precision, mAP and F1 score of YOLOv3-Tiny-S model were second only to those of YOLOv3 model, while the detection speed of the former was much superior to that of the latter, which reached 34.55 FPS. In addition, compared with the YOLOv2-Tiny, YOLOv3-Tiny and YOLOv4-Tiny models, although the detection speed of YOLOv3-Tiny-S is reduced, the accuracy of fire detection is improved. This demonstrates that our proposed YOLOv3-Tiny-S can better deal with the trade-off between the accuracy and real-time performance of fire detection.

(2) A FFMR suitable for fire detection and early warning in construction site is proposed. Due to the complexity of the construction site and the limitations of the traditional fire early detection and early warning equipment, this study proposed the FFMR to do site safety patrol, fire detection, early warning and firefighting, which are realized by the mobile platform control system and command system of FFMR. In the process of patrolling the construction site, the firefighting robot is equipped with a camera to collect the video images in the route in real time. Once a fire is detected, an early warning is sent. At the same time, the operator can remotely control the firefighting robot to go to the fire until the fire is in the range of fire extinguisher (approximately 1.5 m) according to the distance measured by the ultrasonic sensors equipped in front of the robot. Then, open the micro fire extinguisher for early fire extinguishing, which improves the efficiency of emergency response. As the camera used in this FFMR can rotate horizontally and vertically, then, it can detect fire in other floors. The FFMR using the crawler-type walking mechanism is better applicable to the uneven site conditions on the construction site, which also can easily fit in the tight spaces (greater than 500 mm × 300 mm × 500 mm). In addition, the micro fire extinguisher has the advantages of light weight and small volume. When it is equipped on the mobile firefighting robot, it can enhance the flexibility of movement in a narrow space (such as the scaffold area), improve the cruise ability and increase the patrol time. This is because there are two lithium batteries of 12 V used in the FFMR, where the less the battery consumption for the firefighting, the longer the patrol time. One lithium battery at overcharged can support about half an hour of patrol time, then one hour of patrol time can be obtained if no fire is detected. The proposed FFMR has the slightly cost-effective nature by incorporating of a mobile robotic system capable of autonomous navigation (20000 yuan RMB) and several low-cost embedded devices (around 1000 yuan RMB), such as three sensors (99 yuan RMB), two early warning lights (50 yuan RMB), a camera (300 yuan RMB), a micro fire extinguisher (200 yuan RMB), two lights (20 yuan RMB), a remote control (300 yuan RMB), which enables it to be adopted practically on sites.

Although the proposed FFRM has achieved some success in the computer vision-based fire detection and firefighting robot, it has some limitations. For example, in terms of dataset, this study mainly obtains the fire video images of multiple construction scenes downloaded from the Internet and simulated from construction sites. However, due to the variety of construction sites, it is necessary to build a highly targeted dataset in a specific environment to increase the robustness of detection. Moreover, when a fire is detected, it needs the operator to remotely control the robot to open the micro fire extinguisher to put out the fire, which cannot be opened automatically. Therefore, the firefighting robot needs to be more intelligent to automatically turn on the fire extinguisher to extinguish the fire once the occurrence of fire is detected. Finally, the fire detection range of the proposed FFRM should be no more than 10 m considering the camera and the training algorithm used in this research. If a greater distance is needed, then many factors should be considered as a whole, such as: (1) the resolution of the camera, where the higher the resolution, the longer distance can be detected; (2) the training effect of the algorithm used, where the higher the detection rate, the easier the fire target can be detected; (3) the external environment, where the relatively poor light, the far distance of fire can be detected.

6. Conclusions

The occurrence of fire accidents on the construction site can cause huge ecological and economic losses, as well as a lot of human lives. Therefore, this study proposed a FFMR combining computer vision-based fire detection methods and firefighting robot to overcome the shortcomings of traditional fire detection and its early warning on construction sites. Incorporating YOLOv3-Tiny-S model in the camera mounted on the firefighting robot, we showed that a small fire target can be detected with higher accuracy and faster speed under the occlusion by outdoor environment. Through optimizing the network structure, adding the SPP module, adopting the multi-scale anchor mechanism etc. the proposed YOLOv3-Tiny-S model achieved good detection performance, which can help perfecting the research of CNN models. Then, the autonomous response of light and sound early warning is made once a fire is detected. In addition, the reliable communication between the operator and the firefighting robot is ensured by the human-computer interaction interface of the robot command system, where the operator can remotely control the firefighting robot to put out the fire at its early stage. On the other hand, the reliability of accurate fire detection and early response to the fire accident management is helpful to establish a systematic mechanism of fire detection and early warning response on the construction sites.

The proposed FFMR can assist the manager to patrol constantly the desired territory on construction sites, which reduces the detection blind area by fixed cameras and improves the overall effect of fire safety management. In the process of manual inspection, managers are prone to delay emergency response or make wrong response strategies due to negligence or lack of experience. The use of firefighting robot instead of managers to patrol the potential fire area of the construction site can give real-time early warning and timely response to fire safety management when the fire is detected. Therefore, the proposed FFMR can detect and deal with fire at its early stage, which is helpful to effectively prevent the spread of fire, reduce the occurrence of fire accidents, and realize fire safety management. Furthermore, this research can be extended to other high-risk industries, such as petrochemical industry and coal mine, so as to promote the applicability of the research.

Data availability statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under grant No. 72301256.

Funding

This work was supported by the National Natural Science Foundation of China under Grant (72301256) and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) under Grant (CUG2642022006).

Author contributions

LK and SG conceived the study and were responsible for the design and development of the data analysis. JL, XZ and DW were responsible for data collection and analysis. JL, LK and SG were responsible for data interpretation. LK wrote the first draft of the article and SG revised the draft.

Disclosure statement

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

References

Ahn, Y., Choi, H., & Kim, B. S. (2023). Development of early fire detection model for buildings using computer vision-based CCTV. *Journal of Building Engineering*, *65*, Article 105647. https://doi.org/10.1016/j.jobe.2022.105647

Ando, H., Ambe, Y., Ishii, A., Konyo, M., Tadakuma, K., Maruyama, S., & Tadokoro, S. (2018). Aerial hose type robot by water jet for fire fighting. *IEEE Robotics and Automation Letters*, *3*(2), 1128–1135. https://doi.org/10.1109/LRA.2018.2792701

Bosheng, S. (2022, November 25). *Heavy construction fire accidents reported in 2020* (in Chinese). https://www.163.com/dy/article/GU0RDHF00552HS5R.html

Campbell, R. (2020). *Fires in structures under construction or renovation*. National Fire Protection Association.

CE Safety. (2022). *Fires in buildings under construction caused by hot work*. https://cesafety.co.uk/news/how-to-reduce-the-risk-of-fires-caused-by-hot-works/

Chang, C. K., Siagian, C., & Itti, L. (2010). Mobile robot vision navigation & localization using gist and saliency. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 4147–4154). IEEE.

Chaoxia, C., Shang, W., & Zhang, F. (2020). Information-guided flame detection based on faster r-cnn. *IEEE Access*, *8*, 58923–58932. https://doi.org/10.1109/ACCESS.2020.2982994

Chen, C., Guorun, Y., Chenyu, W., Sotirios, G., & Shaohua, W. (2022). Enhancing the robustness of object detection via 6G vehicular edge computing. *Digital Communications and Networks*, *8*(6), 923–931. https://doi.org/10.1016/j.dcan.2022.10.013

Chen, X., Xue, Y. P., Hou, Q. S., Fu, Y. & Zhu, Y. L. (2023). RepVGG-YOLOv7: A Modified YOLOv7 for fire smoke detection. *Fire*, *6*(10), Article 383. https://doi.org/10.3390/fire6100383

Di Paola, D., Milella, A., Cicirelli, G., & Distante, A. (2010). An autonomous mobile robotic system for surveillance of indoor environments. *International Journal of Advanced Robotic Systems*, *7*(1), 19–26. https://doi.org/10.5772/7254

Edirisinghe, R. (2019). Digital skin of the construction site: Smart sensor technologies towards the future smart construction site. *Engineering*, *Construction and Architectural Management*, *26*(2), 184–223. https://doi.org/10.1108/ECAM-04-2017-0066

Fire Safety Matters. (2020). *One-fifth of construction industry-related fires in England caused by hot work*. https://www.fsmatters.com/Causes-of-construction-sector-fires-revealed

Geng, Y., Lai, M., Tian, X. C., Xu, X. L., Jiang, Y., & Zhang, Y. K. (2023). A novel seam extraction and path planning method for robotic welding of medium-thickness plate structural parts based on 3D vision. *Robotics and Computer-Integrated Manufacturing*, *79*, Article 102433. https://doi.org/10.1016/j.rcim.2022.102433

Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440–1448). IEEE. https://doi.org/10.1109/ICCV.2015.169

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580–587). IEEE. https://doi.org/10.1109/CVPR.2014.81

Han, X. F., Jin, J. S., Wang, M. J., Jiang, W., Gao, L., & Xiao, L. P. (2017). Video fire detection based on Gaussian Mixture Model and multi-color features. *Signal*, *Image and Video Processing*, *11*(8), 1419–1425. https://doi.org/10.1007/s11760-017-1102-y

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(9), 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824

He, W., Huang, Z., Wei, Z., Li, C., & Guo, B. (2019). TF-YOLO: An improved incremental network for real-time object detection. *Applied Sciences*, *9*(16), Article 3225. https://doi.org/10.3390/app9163225

Hefeeda, M., & Bagheri, M. (2007, October). Wireless sensor networks for early detection of forest fires. In *2007 IEEE International Conference on Mobile Adhoc and Sensor Systems*. IEEE. https://doi.org/10.1109/MOBHOC.2007.4428702

Hong, X., Wang, W., & Liu, Q. (2019, June). Design and realization of fire detection using computer vision technology. In *2019 Chinese Control and Decision Conference* (*CCDC*) (pp. 5645–5649). IEEE. https://doi.org/10.1109/CCDC.2019.8832897

Innocente, M. S., & Grasso, P. (2019). Self-organising swarms of firefighting drones: Harnessing the power of collective intelligence in decentralised multi-robot systems. *Journal of Computational Science*, *34*, 80–101. https://doi.org/10.1016/j.jocs.2019.04.009

Jiao, Z., Zhang, Y., Xin, J., Mu, L., Yi, Y., Liu, H., & Liu, D. (2019, July). A deep learning based forest fire detection approach using UAV and YOLOv3. In *2019 1st International Conference on Industrial Artificial Intelligence* (*IAI*). IEEE. https://doi.org/10.1109/ICIAI.2019.8850815

Kim, J. H., Jo, S., & Lattimer, B. Y. (2016). Feature selection for intelligent firefighting robot classification of fire, smoke, and thermal reflections using thermal infrared images. *Journal of Sensors*, *2016*, Article 8410731. https://doi.org/10.1155/2016/8410731

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Li, H., Lu, M., Hsu, S. C., Gray, M., & Huang, T. (2015). Proactive behavior-based safety management for construction safety improvement. *Safety Science*, *75*, 107–117. https://doi.org/10.1016/j.ssci.2015.01.013

Li, S., Feng, C., Niu, Y., Shi, L., Wu, Z., & Song, H. (2019). A fire reconnaissance robot based on SLAM position, thermal imaging technologies, and AR display. *Sensors*, *19*(22), Article 5036. https://doi.org/10.3390/s19225036

Li, Z., Mihaylova, L., & Yang, L. (2021). A deep learning framework for autonomous flame detection. *Neurocomputing*, *448*, 205–216. https://doi.org/10.1016/j.neucom.2021.03.019

Li, S., Wang, Y. H., Feng, C. Y., Zhang, D., Li, H. Z., Huang, W. & Shi, L. (2022). A thermal imaging flame-detection model for firefighting robot based on YOLOv4-F model. *Fire*, *5*(5), Article 172. https://doi.org/10.3390/fire5050172

Liu, S., Tu, D., & Zhang, Y. (2009, November). Multiparameter fire detection based on wireless sensor network. In *2009 IEEE International Conference on Intelligent Computing and Intelligent Systems* (Vol. 3, pp. 203–206). IEEE. https://doi.org/10.1109/ICICISYS.2009.5358197

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Lecture notes in computer science: Vol. 9905. Computer vision – ECCV 2016* (pp. 21–37). Springer, Cham. https://doi.org/10.1007/978-3-319-46448-0_2

Madhevan, B., Sakkaravarthi, R., Singh, G. M., Diya, R., & Jha, D. K. (2017). Modelling, simulation and mechatronics design of a wireless automatic fire fighting surveillance robot. *Defence Science Journal*, *67*(5), Article 572. https://doi.org/10.14429/dsj.67.10237

Management & Emergency. (2020). *Nationalwide fire and fire rescue reported in 2020* (in Chinese). https://www.119.gov.cn/article/3xBeEJjR54K

McNeil, J. G., & Lattimer, B. Y. (2017). Robotic fire suppression through autonomous feedback control. *Fire Technology*, *53*(3), 1171–1199. https://doi.org/10.1007/s10694-016-0623-1

Mir-Nasiri, N., Siswoyo J. H., & Ali, M. H. (2018). Portable autonomous window cleaning robot. *Procedia Computer Science*, *133*, 197–204. https://doi.org/10.1016/j.procs.2018.07.024

Muhammad, K., Ahmad, J., & Baik, S. W. (2018). Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing*, *288*, 30–42. https://doi.org/10.1016/j.neucom.2017.04.083

Nguyen, A. Q., Nguyen, H. T., Tran, V. C., Pham, H. X. & Pestana, J. (2021). A visual real-time fire detection using single shot MultiBox detector for UAV-based fire surveillance. In 8th *IEEE International Conference on Communications and Electronics* (*IEEE ICCE*), Vietnam. https://doi.org/10.1109/ICCE48956.2021.9352080

Palmer, A. (2012). Hotel construction site catches fire. *Top Stories*, *Midland Reporter-Telegram*. Texas.

Park, J., Cho, Y. K., & Martinez, D. (2016). A BIM and UWB integrated mobile robot navigation system for indoor position tracking applications. *Journal of Construction Engineering and Project Management*, *6*(2), 30–39. https://doi.org/10.6106/JCEPM.2016.6.2.030

Pincott, J., Tien, P. W., Wei, S., & Calautit, J. K. (2022). Development and evaluation of a vision-based transfer learning approach for indoor fire and smoke detection. *Building Services Engineering Research and Technology*, *43*, 319–332. https://doi.org/10.1177/01436244221089445

Prema, C. E., Vinsley, S., & Suresh, S. (2018). Efficient flame detection based on static and dynamic texture analysis in forest fire detection. *Fire Technology*, *54*(1), 255–288. https://doi.org/10.1007/s10694-017-0683-x

Qiu, X., Xi, T., Sun, D., Zhang, E., Li, C., Peng, Y., Wei, J., & Wang, G. (2018). Fire detection algorithm combined with image processing and flame emission spectroscopy. *Fire Technology*, *54*(5), 1249–1263. https://doi.org/10.1007/s10694-018-0727-x

Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7263–7271). IEEE. https://doi.org/10.1109/CVPR.2017.690

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv*. https://doi.org/10.48550/arXiv.1804.02767

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779–788). IEEE. https://doi.org/10.1109/CVPR.2016.91

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. https://doi.org/10.1109/TPAMI.2016.2577031

Roberto, G. F., Branco, K. C., Machado, J. M., & Pinto, A. R. (2013, April). Local data fusion algorithm for fire detection through mobile robot. In *2013 14th Latin American Test Workshop-LATW*. IEEE. https://doi.org/10.1109/LATW.2013.6562667

Shen, D., Chen, X., Nguyen, M., & Yan, W. Q. (2018, April). Flame detection using deep learning. In *2018 4th International conference on control*, *automation and robotics* (*ICCAR*) (pp. 416–420). IEEE. https://doi.org/10.1109/ICCAR.2018.8384711

So, A. T., Lo, T. Y., & Chan, W. L. (1996). An autonomous robotic cladding inspector for high-rise buildings in Hong Kong. *HKIE Transactions*, *3*(2), 37–45. https://doi.org/10.1080/1023697X.1996.10667701

Sridhar, P., & Sathiya, R. (2021). Computer vision based early electrical fire-detection in video surveillance oriented for building environment. *Journal of Physics: Conference Series*, *1916*, Article 012024. https://doi.org/10.1088/1742-6596/1916/1/012024

Su, Y., Mao, C., Jiang, R., Liu, G., & Wang, J. (2021). Data-driven fire safety management at building construction sites: Leveraging CNN. *Journal of Management in Engineering*, *37*(2), Article 04020108. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000877

Tesema, F. B., Lin, J., Ou, J., Wu, H., & Zhu, W. (2018). Feature fusing of feature pyramid network for multi-scale pedestrian detection. In *2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing* (*ICCWAMTIP*) (pp. 10–13). IEEE. https://doi.org/10.1109/ICCWAMTIP.2018.8632614

Tso, S. K., & Feng, T. O. N. G. (2003, September). Robot assisted wall inspection for improved maintenance of high-rise buildings. In *20th International Symposium on Automation and Robotics in Construction* (pp. 449–455). https://doi.org/10.22260/ISARC2003/0071

Victores, J. G., Martínez, S., Jardón, A., & Balaguer, C. (2011). Robot-aided tunnel inspection and maintenance system by vision and proximity sensor integration. *Automation in Construction*, *20*(5), 629–636. https://doi.org/10.1016/j.autcon.2010.12.005

Vishaal, R., Raghavan, P., Rajesh, R., Michael, S., & Elara, M. R. (2018). Design of dual purpose cleaning robot. *Procedia Computer Science*, *133*, 518–525. https://doi.org/10.1016/j.procs.2018.07.065

Wang, Y., Xing, J. P., Guo, H., & Wang, L. J. (2017). Key technologies of tunnel firefighting robots. *IETE Technical Review*, *34*(1), 3–10. https://doi.org/10.1080/02564602.2016.1139475

Wang, Z., Li, H., & Zhang, X. (2019). Construction waste recycling robot for nails and screws: Computer vision technology and neural network approach. *Automation in Construction*, *97*, 220–228. https://doi.org/10.1016/j.autcon.2018.11.009

Wang, Z., Li, H., & Yang, X. (2020). Vision-based robotic system for on-site construction and demolition waste sorting and recycling. *Journal of Building Engineering*, *32*, Article 101769. https://doi.org/10.1016/j.jobe.2020.101769

Wu, H., Wu, D., & Zhao, J. (2019). An intelligent fire detection approach through cameras based on computer vision methods. *Process Safety and Environmental Protection*, *127*, 245–256. https://doi.org/10.1016/j.psep.2019.05.016

Wu, Z., Xue, R., & Li, H. (2022). Real-time video fire detection via modified YOLOv5 network mode. *Fire Technology*, *58*(4), 2377–2403. https://doi.org/10.1007/s10694-022-01260-z

Xiao, D., Shan, F., Li, Z., Le, B.T., Liu, X., & Li, X. (2019). A target detection model based on improved Tiny-Yolov3 under the environment of mining truck. *IEEE Access*, *7*, 123757–123764. https://doi.org/10.1109/ACCESS.2019.2928603

Xie, Y., Zhu, J., Cao, Y., Zhang, Y., Feng, D., Zhang, Y., & Chen, M. (2020). Efficient video fire detection exploiting motion-flicker-based dynamic features and deep static features. *IEEE Access*, *8*, 81904–81917. https://doi.org/10.1109/ACCESS.2020.2991338

Xu, Z., Guo, Y., & Saleh, J. H. (2020). Tackling small data challenges in visual fire detection: A deep convolutional generative adversarial network approach. *IEEE Access*, *9*, 3936–3946. https://doi.org/10.1109/ACCESS.2020.3047764

Yang, H., Jang, H., Kim, T., & Lee, B. (2019). Non-temporal lightweight fire detection network for intelligent surveillance systems. *IEEE Access*, *7*, 169257–169266. https://doi.org/10.1109/ACCESS.2019.2953558

Yar, H., Khan, Z. A., Ullah, F. U. M., Ullah, W. & Baik, S. W. (2023). A modified YOLOv5 architecture for efficient fire detection in smart cities. *Expert Systems with Applications*, *231*, Article 120465. https://doi.org/10.1016/j.eswa.2023.120465

Yi, Z., Yongliang, S., & Jun, Z. (2019). An improved TINY-YOLOV3 pedestrian detection algorithm. *Optik*, *183*, 17–23. https://doi.org/10.1016/j.ijleo.2019.02.038

Zhan, H. W., Pei, X. Y., Zhang, T. H. & Zhang, L. Q. (2023). Research on flame detection method based on improved SSD algorithm. *Journal of Intelligent & Fuzzy Systems*, *45*, 6501–6512. https://doi.org/10.3233/JIFS-232645

Zhang, J., Jin, Z., & Feng, H. (2018). Type synthesis of a 3-mixed-DOF protectable leg mechanism of a firefighting multi-legged robot based on GF set theory. *Mechanism and Machine Theory*, *130*, 567–584. https://doi.org/10.1016/j.mechmachtheory.2018.08.026

Zhang, H., Qin, L., Li, J., Guo, Y., Zhou, Y., Zhang, J., & Xu, Z. (2020). Real-time detection method for small traffic signs based on Yolov3. *IEEE Access*, *8*, 64145–64156. https://doi.org/10.1109/ACCESS.2020.2984554

Zhang, H., Wang, Z., Chen, M., Peng, Y., Gao, Y., & Zhou, J. (2021). An improved YOLOv3 algorithm combined with attention mechanism for flame and smoke detection. In X. Sun, X. Zhang, Z. Xia, & E. Bertino (Eds.), *Lecture notes in computer science: Vol. 12736. Artificial intelligence and security. ICAIS 2021* (pp. 226–238). Springer, Cham. https://doi.org/10.1007/978-3-030-78609-0_20

Zhang, T., Wang, Z., Zeng, Y., Wu, X., Huang, X., & Xiao, F. (2022). Building artificial-intelligence digital fire (AID-Fire) system: A real-scale demonstration. *Journal of Building Engineering*, *62*, Article 105363. https://doi.org/10.1016/j.jobe.2022.105363

Zhu, J., Li, W., Lin, D., Cheng, H., & Zhao, G. (2020a). Intelligent fire monitor for fire robot based on infrared image feedback control. *Fire Technology*, *56*(5), 2089–2109. https://doi.org/10.1007/s10694-020-00964-4

Zhu, J., Pan, L., & Zhao, G. (2020b). An improved near-field computer vision for jet trajectory falling position prediction of intelligent fire robot. *Sensors*, *20*(24), Article 7029. https://doi.org/10.3390/s20247029