IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 2 July 2022, accepted 9 July 2022, date of publication 14 July 2022, date of current version 9 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3190867

METHODS

Characteristics Based Fire Detection System Under the Effect of Electric Fields With Improved Yolo-v4 and ViBe

SHUAI ZHAO[1], BOYUN LIU[1], ZHENG CHI[1], TAIWEI LI[2], AND SHENGNAN LI[1]
[1]School of Power Engineering, Naval University of Engineering, Wuhan 430033, China
[2]School of Weaponry Engineering, Naval University of Engineering, Wuhan 430033, China
Corresponding author: Shuai Zhao (157999945@qq.com)

ABSTRACT To address slow image-based detection of fires under the effect of electric fields and its limitation to static fire characteristics, this paper proffersa video-based fire detection system with improved you only look once version 4 (Yolo-v4) and visual background extractor (ViBe) algorithms. The proposed system uses a simplified weighted bi-directional feature pyramid network (Bi-FPN) in place of the path aggregation network (PANet) as a feature fusion network in Yolo-v4. Using multiple dynamic fire characteristics, it can eliminate falsely detected frames. The ViBe algorithm is improved to consider the sudden change of light triggered by fire flickering. Compared with other fire detection algorithms, the proposed system achieves 98.9% fire detection accuracy with a false detection rate of 2.2%. It can extract target fires by adjusting to sudden changes of light using no more than 16 frames. Moreover, the system achieves fire detection with more dynamic fire characteristics compared with the image-based fire detection under the effect of electric fields.

INDEX TERMS Fire detection, Yolo-v4, ViBe, bi-directional feature pyramid, dynamic characteristics, sudden change of light.

I. INTRODUCTION

In fuel combustion, fires, as weakly ionized plasmas [1], generate a variety of charged and neutral particles. However, the particle trajectory and ion collision frequency of fires may change under the effect of external electric fields. The change will further affect their chemical reaction rate and morphological characteristics and even lead to the extinguishing of combustion [2], [3]. A fire extinguishing technique based on electric fields has therefore been developed [4], but it still faces a number of unsolved problems, for example, accelerated, complicated, and uncertain fires in the combustion affected by imposed external electric fields. To further explore the mutual effect of electric fields and fires, the dynamic characteristics of fires must be detected under the effect of electric fields.

The first factor that fire detection relied on was color. In [5], Celik *et al.* proposed an algorithm to effectively partition the area in terms of fire color, but misjudgment was easily caused by indistinct color between light and fire in actual environment since it was limited to a specific feature of fire. In [6], a dynamic fire analysis and detection method in red green blue (RGB)/ hue saturation intensity (HSI) color space based on decision rule was put forward. It took into account the increase of pixels resulting from the randomness of fire in detection, but failed to discriminate moving area from fire since it depended on the difference between frames. Apart from color model, Borges [7] proposed some features of fire in the lower area including skewness, color, roughness, and area size to identify the change between frames, and combined them with the Bayesian classifier for fire detection. In order to enhance the accuracy of fire recognition, some researchers included the motion features of fire besides color features. In [8], a feature vector was extracted from the physical features of light stream and fire to distinguish fire from rigid objects. Presently, fire detection focuses on how to algorithm optimization to detect fire from other objects, and how to improve its efficiency. In terms of practical effect,

The associate editor coordinating the review of this manuscript and approving it for publication was Zeev Zalevsky.

some fire detection approaches may require little time to implement but be troubled by low accuracy. Some methods may achieve good accuracy of fire detection, but take more time to operate. Muhammad introduced the convolutional neural network (CNN) model into fire detection in [9], which guaranteed the accuracy with short-time implementation, but its accuracy could be further improved.

For this reason, you only look once version 4 (Yolo-v4) and visual background extractor (ViBe) algorithms are chosen and improved to propose a monocular visual recognition fire detection system. The false detection rate is lowered by adding the bi-directional feature pyramid network (Bi-FPN) into the feature fusion network (PANet) in Yolo-v4. A number of dynamic, characteristic fire detection mechanisms are integrated to eliminate the frames with check error enhancing the accuracy to 98.9%. The ViBe threshold adaptive mechanism adjusts along with the sudden changes of light triggered by fire flickering. The proposed algorithm can achieve the target recognition of different-size fires, accurate extraction of target fires, and detection of unsteady fires with multiple dynamic characteristics.

II. RELATED WORKS

In recent years, many Chinese and foreign scholars [10]–[17] have explored the combustion characteristics, chemical reactions, and ionic wind effects of fires under the effect of electric fields. However, they often used the photos taken by intensified charge coupled device or complementary metal oxide semiconductor cameras in experiments followed by artificial processing. This approach achieves higher accuracy only in processing the static morphological information of small steady fires, but it experiences slow and inaccurate data processing while facing unsteady fires above middle size.

Deep learning networks such as convolutional neural networks [18], recurrent neural networks [19], and deep belief networks [20] have recently been applied in object detection, image processing, and positioning. Frequently used in video-based fire detection, in particular, the convolutional neural network algorithm facilitates successful image recognition. For instance, a nine-layer convolutional neural network was devised by [20], while a two-level structural convolutional neural network was designed by [21]. From [22], a 12-layer convolutional neural network was developed to considerably increase fire detection rates and lower the false detection rate, but the detection was still quite slow. One-stage object detection algorithms represented by Yolo algorithms [24]–[27] can achieve a detection rate exceeding 45 frames per second, which is actually real-time. However, these algorithms do not perform satisfactorily in the detection of small objects. Among them, Yolo-v4 and Yolo-v5 are the improved algorithms of Yolo-v3. Moreover, Yolo-v5 is an algorithm launched by a different team from that of Yolo-v4, but it has not been published yet. The two algorithms Yolo-v4 and Yolo-v5 have similar performance since both of them use the cross stage partial dark network (CSPDarknet) and path aggregation network (PANet) as their backbone feature extraction network and feature fusion network, respectively. Nevertheless, Yolo-v4 is more customized. In recent experiments [28], it has been revealed that Yolo-v4 is still the most accurate among these Yolo algorithms. Considering the abundance of materials available, better customization, and higher mAP, Yolo-v4 is therefore taken as the object detection algorithm in this paper.

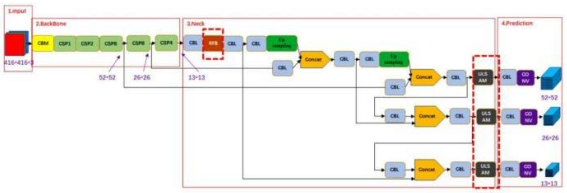


FIGURE 1. Diagram of Yolo-v4 network structure.

Moving objects are detected using image or video sequences in such methods as frame differencing [29], background subtraction and optical flow. Among them, background subtraction is the most common method for the detection of moving objects [30]. In background subtraction, the visual background extractor (ViBe) algorithm [31] has been widely adopted due to less computation, faster processing, and good performance. However, it is still poorly adapted to background during the sudden changes of light in fixed-point fire detection [32]. These algorithms do not perform well in the fire detection based on dynamic morphological characteristics, so they are not suitable for dynamic characteristic detection when different sizes of fires exist with violent flickering under the influence of electric fields. Therefore, this paper integrates Yolo-v4 and ViBe algorithms for greater improvement.

III. ALGORITHMS FOR THE SYSTEM

A. YOLO-V4 ALGORITHM FOR DETERMINING REGION OF INTEREST (ROI)

As shown in Fig. 1, Yolo-v4 is a neural network algorithm that uses the cross stage partial dark network 53 (CSPDarknet-53) as its backbone network and takes spatial pyramid pooling (SPP), feature pyramid network (FPN), and personal area network (PAN) as its enhanced feature extraction networks. As an end-to-end object detection algorithm, it uses a neural network integrating four stages, that is, object region proposal generation, object feature extraction with backbone network, enhanced feature extraction network for integration of shallow and deep information, and verification of detected object proposals.

The backbone feature extraction network of Yolo-v4 is CSPDarknet53. After an image is processed in the backbone network, Yolo-v4 will output the feature maps of scales 13×13, 26×26, and 52×52. Each scale of feature map contains the semantic information of different dimensions. In the feature fusion section, the feature map of scale 13×13 will enter the SPP structure [33], which treats the new feature map

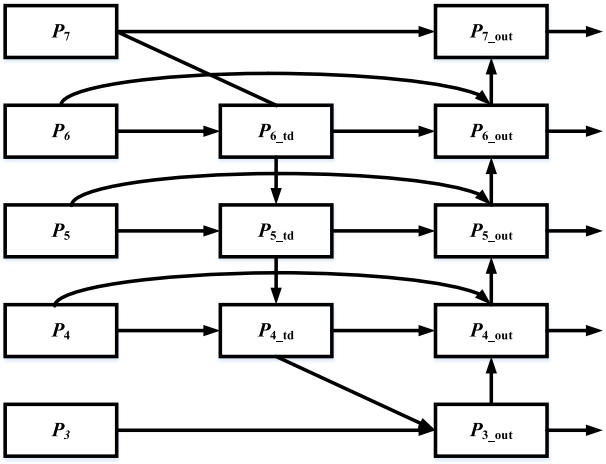


FIGURE 2. Bi-FPN structure.

and the original feature map by stacking and convolution, and then outputs it into the feature fusion network PAN [34]. The feature map of scale 13 × 13 is upsampled twice in the PAN, and then stacked with the feature maps of scales 26 × 26 and 52 × 52, respectively, for convolution. Subsequently, downsampling, stacking and convolution are carried out in a similar way from bottom to top, in order to sufficiently fuse the features in the feature maps of three scales. In the end, three Yolo detection heads, i.e. 13 × 13, 26 × 26, and 52 × 52, are output.

The region containing the fires to be detected takes up a very small part of a captured image; therefore, the Yolo-v4 object detection algorithm is employed in this paper to identify target fires and demarcate the region of interest (ROI). Formed by a cluster of data, the ROI contains the coordinate information of the selected region. After preliminary image processing, information useful for the experiment will be retained without altering image size. In this way, the useful information is extracted from the image, and only a part of the target region is kept for subsequent processing. This can greatly shorten the duration of image processing and enhance the subsequent filtering efficiency.

B. BI-FPN

A Bi-directional feature pyramid network (Bi-FPN) has a typical structure of complicated bi-directional feature fusion, as shown in Fig. 2.To some extent, it achieves a simplified structure by removing two intermediate points in the structure of the traditional FPN (e.g. PANet), that is, highest-dimension feature layer and lowest-dimension feature layer, and adding a residual edge connecting input feature map with an output feature map in each intermediate feature layer.

At each point for feature fusion, different input feature maps should make different contributions to the output. Therefore, Bi-FPN uses a weight for training to adjust the contribution of different inputs to the output feature map.

In the selection of weight, Bi-FPN uses fast normalization fusion to improve the rate by 30% while keeping similar performance compared with the optimization based on softmax. The fast normalized fusion is defined by Equation (1):

$$o = \sum_i \frac{\omega_i}{\varepsilon + \sum_j \omega_j} \cdot \boldsymbol{I}_i \tag{1}$$

where i and j represent the number of input feature maps at two feature fusion points, and $i = j$; I_j is the matrix of input feature maps; ε is 10^{-4}, a constant to keep the denominator from being zero; ω_i and ω_j are the weights of each input feature map, initially set in the range $0 < \omega_i < 1$ and $0 < \omega_j < 1$. The guaranteed value of the function activated by rectified linear unit (ReLU) is constantly greater than zero after each weight training. After multiple trainings, the input feature maps at each feature fusion point get the weights for the best performance of the object detection algorithm.

C. VIBE ALGORITHM FOR OBJECT CAPTURING

Visual background extractor (ViBe), a background modeling algorithm first proposed by Barnich *et al.* [35], is based on pixel models. In the algorithm, the samples of background frames are used to constitute the background models of pixels. Subsequently, the pixel values of input frames are matched with the background sample set to judge whether pixels belong to the background by virtue of a preset threshold. The matched pixels are then used to update the background model. The algorithm is mainly implemented in three steps, that is, background modeling, foreground object detection, and background model updating. With less memory used in operation, it can achieve more real-time and more accurate detection.

IV. IMPROVEMENT OF ALGORITHMS

A. FEATURE FUSION NETWORK

The feature fusion network takes the feature maps of scales 13 × 13, 26 × 26, and 52 × 52 from the 7 th, 13 th, and 16 th layers of the backbone network as its input. Among them, the feature map of the scale 13 × 13 will enter the spatial pyramid pooling (SPP) structure, and treated by max-pooling of three scales with the pooling kernel size of 13 × 13, 9 × 9, and 5 × 5, respectively. The output is input into the Bi-FPN-Lite structure together with the feature maps of the scales 52 × 52 and 26 × 26.

In the structure, the simplified weighted bi-directional feature pyramid network (Bi-FPN) is used in place of the path aggregation network (PANet) feature fusion network used for neck in Yolo-v4.

In the structure, the simplified weighted bi-directional feature pyramid network (Bi-FPN) is used in place of the path aggregation network (PANet) feature fusion network used for neck in Yolo-v4. The original Bi-FPN structure has outputs to the feature fusion network after continuously downsampling an image five times. However, the Yolo-v4 algorithm can output feature maps of only three scales after the backbone

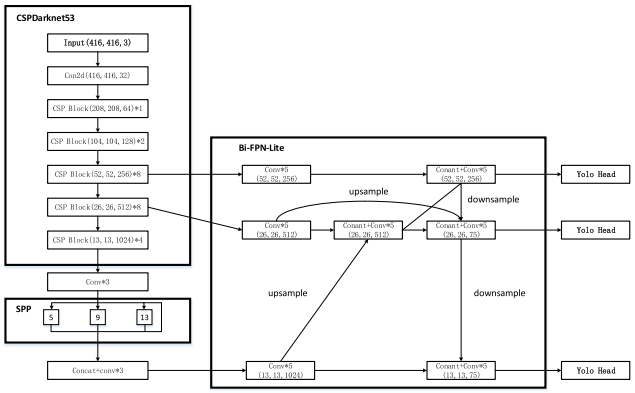


FIGURE 3. Framework of the improved Yolo-v4 network structure.

feature extraction network. It is also difficult to advance downsampling after the feature maps of the scale 13 × 13.

As shown in Fig. 3, two feature layers are reduced from the Bi-FPN algorithm used in this paper to make it match with the Yolo-v4 model. Therefore, feature fusion happens only to the feature layers of three scales, i.e., 13 × 13, 26 × 26, and 52 × 52. This reduction is equivalent to adding a residual edge pointing from P_4^{in} to P_4^{out} on the basis of PANet. Meanwhile, the attention mechanism is also imposed at the points where multiple feature maps are involved in feature fusion, e.g., P_4^{td}, P_4^{out}, P_3^{out}, and P_5^{out}. Unlike the weight training of feature maps in a cross stage partial (CSP) block, Bi-FPN-Lite uses the attention mechanism to directly multiply each output feature map at a feature fusion point by a weight ω and performs the weight training with fast normalized fusion. Upsample is directly taken at P_5^{in} and used in feature fusion with P_4^{in} at the pointP_4^{td}. Then, feature fusion is performed directly using P_4^{out} and the downsampling results atP_4^{in}, P_4^{td} and P_4^{out}. The input feature matrix at the points P_4^{td} and P_4^{out} are substituted into Equation (1) to obtain Equations (2) and (3):

$$\boldsymbol{P}_4^{\text{td}} = Conv\left(\frac{\omega_1 \cdot \boldsymbol{P}_4^{\text{in}} + \omega_2 \cdot Resize\left(\boldsymbol{P}_5^{\text{in}}\right)}{\omega_1 + \omega_2 + \varepsilon}\right) \tag{2}$$

$$\boldsymbol{P}_4^{\text{out}} = Conv\left(\frac{\omega_1 \cdot \boldsymbol{P}_4^{\text{in}} + \omega_2 \cdot \boldsymbol{P}_4^{\text{td}} + \omega_3' \cdot Resize\left(\boldsymbol{P}_3^{out}\right)}{\omega_1' + \omega_2' + \omega_3' + \varepsilon}\right) \tag{3}$$

where Conv stands for convolution; Resize represents the feature maps from upsampling or downsampling in other layers; ω denotes the weight assigned to each feature map for training, which is initially a random number in the range $0 < \omega < 1$ and determined as a reasonable value after each training; ε is 10^{-4}, a constant used to prevent the denominator from being zero; and the plus sign in the denominator represents the stacking of feature maps. The Bi-FPN-Lite structure will output three Yolo detection heads, i.e., 13 × 13, 26 × 26, and 52 × 52, which are subsequently used to generate the prediction boxes with the Yolo-v4 decoding algorithm.

B. PROCESSING OF FALSELY DETECTED FRAMES WITH MULTIPLE FIRE CHARACTERISTICS

Fires are continuous and always changing, so that they must be judged by virtue of their characteristics including flicker frequency, rate of area change, and circularity variation. When the actual size of a combustor is known, we can calculate the distance between the pixels for the edges of the combustor in a video and thus obtain the proportion of pixels to its actual size. In this way, we can determine the size of fires. The fires are identified again in the missed frames based on multiple fire characteristics to reduce false detection rate.

1) FLICKER FREQUENCY

Normally, the flicker frequency of fires ranges 7∼10Hz [36] and may be even higher under the effect of external electric fields. In this paper, the change of pixel light in neighboring multi-frame images of a video is used to calculate the flicker characteristics of fires. The results of fire detection are analyzed in terms of flicker characteristics to effectively eliminate misjudgments in detection. We use the cumulative difference of light between neighboring frames to determine flicker frequency and construct the flicker count matrix $M(x, y, t)$ and gray scale matrix $I(x, y, t)$. The equation for calculation is as follows:

$$\begin{cases} M(x, y, t) = \begin{cases} M(x, y, t-1) + 1, \Delta I > T \\ M(x, y, t-1), else \end{cases} \\ \Delta I = |I(x, y, t) - I(x, y, t-1)| \end{cases} \tag{4}$$

where $I(x, y, t)$ is the grayscale of the pixel (x, y) at the time t; $M(x, y, t)$ is the flicker count of the pixel (x, y) at the time t; T is the threshold of light difference between two neighboring frames. If ΔI exceeds the threshold T, the flicker count increases by 1. Otherwise, it remains unchanged. At the time t, the flicker count at a pixel within a time period T is used to judge whether the region it belongs to flickers. Normally, T is 1s, that is, frame rate. If the flicker count M varies beyond the threshold T_M (set to 10) [29] within the time period of 1s, it is judged that the region flickers.

$$M(x, y, t) - M(x, y, t - T) > T_M \tag{5}$$

Finally, the flicker frequency of images within a time period is calculated by

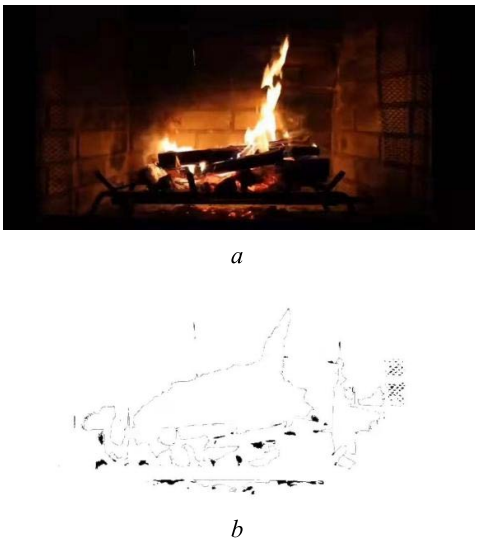
$$F = \frac{M(x, y, t_n) - M(x, y, t_1)}{t_n - t_1} \tag{6}$$

2) RATE OF AREA CHANGE

The area of fires changes continuously. In images, the area occupied by suspected fires varies constantly, and is calculated by

$$A_{\Delta k} = \frac{S_{t+k} - S_t}{k} \tag{7}$$

where $A_{\Delta k}$ represents the rate of area change within the time period k; S_{k+t} is the area occupied by suspected fires in the k^{th} frame of an image after the time t; and S_t is the area occupied by suspected fires in an image at the time t.



a

b

FIGURE 4. ViBe algorithm for sudden changes of light (a) Original image of video, (b)Segmented image with sudden changes of light.

3) CIRCULARITY

Circularity is a concept representing the complexity of an object's shape. Compared with the objects that also flicker but have a stable shape, such as neon lights and lamps, fires have a clearly more difficult shape. Circularity is therefore taken as a basis for recognizing fires. It is calculated by

$$C_t = \frac{P_t^2}{4\pi A_t}(k = 1, 2, \cdots, n) \tag{8}$$

where C_t is the circularity of the area occupied by suspected fires in an image at the time t; P_t represents the circumference of the area occupied by suspected fires; A_t is the area occupied by suspected fires; n is the total size of dataset.

C. IMPROVEMENT FOR SUDDEN CHANGE OF LIGHT

The original ViBe algorithm uses a fixed global threshold, which applies a unified threshold in the binarization of an entire image. However, threshold selection depends over much on subjective experience. High-frequency shake of fires may cause a fast-changing background with sudden changes of light. In this case, a fixed threshold may lead to massive disturbance. The threshold should therefore be increased to adjust for the sudden changes of light in the background. As shown in Fig. 4, fires cause sudden changes of light and create a brighter background. With a fixed threshold, the algorithm cannot adjust to the changing background, which generates a lot of white noises in a segmented image.

The algorithm must be adjusted to the change of detection environment in a timely manner. In this paper, an adaptive threshold is added and improved in the following steps while segmenting the region of moving fires.

Firstly, an initiation interference mechanism is established. It is assumed that an image contains$_X$ foreground pixels and Y background pixels, and Z indicates the proportion of foreground pixels to the total pixels of the image. Then there is

$$Z = \frac{X}{X + Y} \tag{9}$$

The sudden change of light leads to continuous frames, causing the variation of Z. If the difference of Z between two neighboring frames exceeds the set T (which is 75% herein), it is judged that a sudden change of light occurs. In this case, the adaptive threshold is initiated to 20 frames.

Secondly, the distance threshold set for the background is updated. The distance is adjusted to the change of the background, and the minimum distance set $D(x)$ is defined by

$$D(x) = \{D_1(x), \mathrm{K}, D_k(x), \mathrm{K}, D_N(x)\} \tag{10}$$

Thirdly, in the above equation, $D_k(x) = \min\{\mathrm{dist}(I_i(x), v_i(x))\}$ refers to the minimum Euclidian distance from the selected pixel $I_i(x)$ to $v_i(x)$ in the sample; min indicates the minimum distance to the sample numbered i. The average of the N values of $D_k(x)$, that is, $d_{\min}(x)$, represents the dynamic motion of the background and is defined by

$$d_{\min}(x) = \frac{1}{N}\sum_k D_k(x) \tag{11}$$

In other words, $d_{\min}(x)$ is obtained and recorded after each successful match of $I_i(x)$ and $v_i(x)$.

Fourthly, when the background is static, $d_{\min}(x)$ changes slightly. If any dramatic change happens to the background, $d_{\min}(x)$ will increase significantly. The variation of $d_{\min}(x)$ can realize the adaptive update of $R(x)$ as follows:

$$R(x) = \begin{cases} R(x)g\,(1 - \alpha_{dkc}) & if\ R(x) > d_{\min}(x) \cdot \zeta \\ R(x)g\,(1 + \alpha_{inx}) & \text{else} \end{cases} \tag{12}$$

where a_{dkc}, a_{inx}, and ζ are fixed parameters, which are 0.05, 5, and 1.1, respectively. When the background is subject to dynamic interference, $R(x)$ will increase slowly and adjust to such interference.

A parameter $L(x)$ is defined to describe the dispersion level of samples as follows:

$$L_i(x) = \frac{1}{N}\sum_{j=1}^{N} K\left(v_i(x), v_j(x)\right) \tag{13}$$

$$K\left(v_i(x), v_j(x)\right) = \begin{cases} 1 & dist\left(v_i(x), v_j(x)\right) \le R \\ 0 & \text{else} \end{cases} \tag{14}$$

where $v_i(x)$ and $v_j(x)$ are two samples in the background model $M(x)$; N is the number of samples. For every sample $v_i(x)$ in $M(x)$, we calculate its dispersion level from other samples and then determine the best substitute.

The spatial consistency is judged as follows: The $k \times k$ region at any pixel $x(x_i, y_i)$ in the video frame I is defined by

$$N_x = \left\{y = \left(y_i, y_j\right) \in I : |x_i - y_i| \le k, \left|x_j - y_j\right| \le k\right\} \tag{15}$$

The set Ωx is defined as the pixels in N_x which match with the background model:

$$\Omega_x = \{y \in N_x : \#\{M(y) \cap S_R(I(y))\} < \#_{\min}\} \tag{16}$$

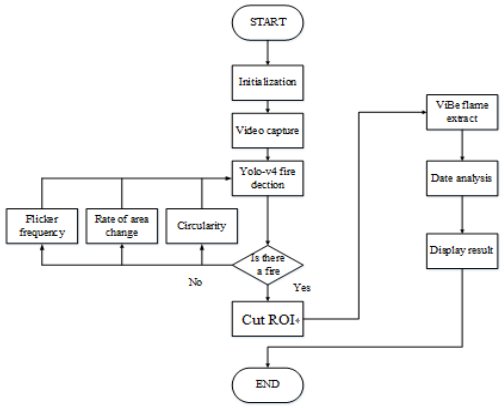


FIGURE 5. Process flow of a detection system.

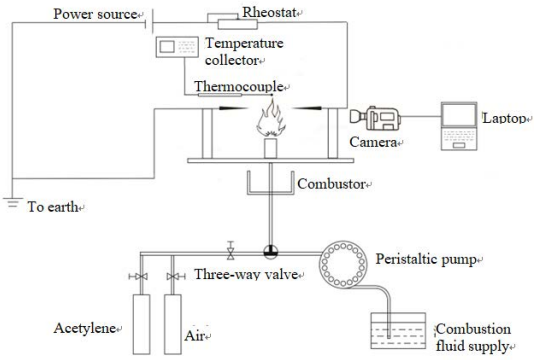


FIGURE 6. Schematic diagram of experiment system.

V. DETECTION EXPERIMENT

The process flow of a detection system presented in Fig. 5 is detailed as follows:

Step 1: Start the system operation and initialize the configuration of parameters;

Step 2: Capture a video of fires;

Step 3: Use the improved Yolo-v4 algorithm for detection in the video and identify the frames with fires;

Step 4: Identify the frames without fires based on dynamic fire characteristics;

Step 5: Demarcate the region of interest;

Step 6: Use the ViBe foreground detection algorithm to highlight fires and binarize the images;

Step 7: Calculate the actual size of fires based on the relationship between pixel size and actual size and analyze the characteristics;

Step 8: Output the information on fire characteristics and end the operation.

A. EXPERIMENT EQUIPMENT AND SETTING

As shown in Fig. 6, the experiment equipment contained mainly a combustion system, an electric field application system, and a detection system. Among them, the video-based detection system was equipped with a high-speed industrial camera (NPX-GS6500UM) and a workstation with an Intel(R)Xeon(R)CPU, 3080Ti×2GPU, 128G memory, and WIN10 operating system. The platform for testing algorithms was a laptop with an Intel(R) Core(TM) i7-11800HQ CPU, RTX3060-6G GPU, and 16G memory.

TABLE 1. Description of datasets.

Dataset	Images with fires (piece)	Images without fires (piece)	Total (piece)
Test set	3800	1300	5100
Validation set	1000	400	1400
Training set	20333	0	20333
Total	25133	1700	26833

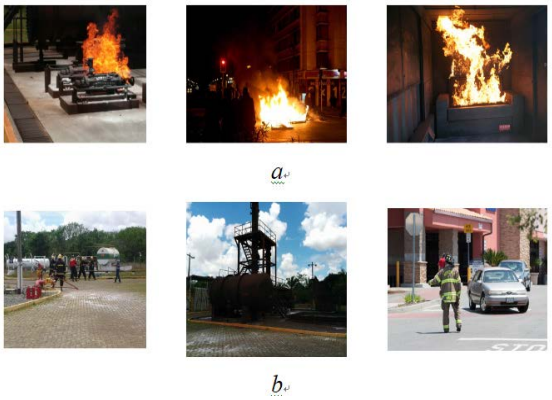


FIGURE 7. Some test data (a) Data with fires, (b) Data without fires.

B. EXPERIMENT DATASETS

The datasets used in the experiment are presented in Table 1.

Firstly, the training set for this experiment was formed by 20,333 pieces of images taken from the disclosed fire datasets ImageNet, and BoWFire, and [37].

Secondly, the videos both with and without fires were selected from the fire video database disclosed by the Bilkent University [38], the videos provided in [39], and from those taken in the experiment. From these videos, 5,500 pieces of images were extracted per frame, and 1,000 pieces of fire images were downloaded from the internet, so as to constitute the test set and validation set in this experiment, as shown in Fig. 7.

C. IMPROVED YOLO-V4 ALGORITHM FOR VIDEO-BASED FIRE DETECTION

It is defined that accuracy A is the probability of correctly judging whether there is a fire or not, false detection rate P_f is the probability of misjudging fires when there are no fires, and miss rate N_f is the probability of missing fires in the detection. They are calculated by the following:

$$\begin{cases} A = \frac{T_P + T_N}{N_{\text{pos}} + N_{\text{neg}}} \times 100\% \\ P_f = \frac{F_P}{N_{\text{neg}}} \times 100\% \\ N_f = \frac{F_N}{N_{\text{pas}}} \times 100\% \end{cases} \tag{17}$$

TABLE 2. Ablation experiment with the improved YOLO-V4.

Algorithm	mAP/%	Parameter	False detection rate
YOLOv4	82.98	64,040,001	2.5
YOLOv4+Bi-FPN-Lite	84.51	64,439,425	1.7

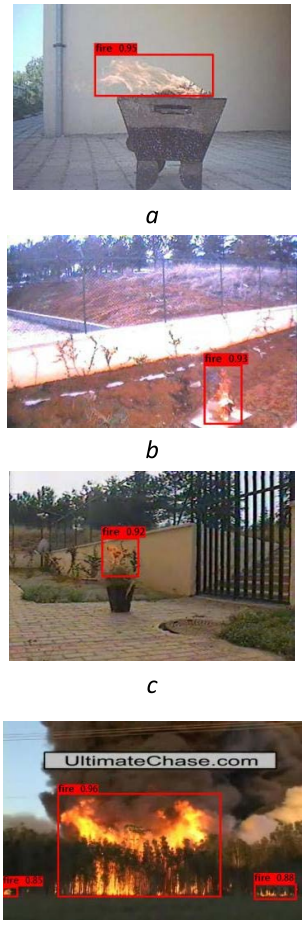


FIGURE 8. Some video-based fire detection results with improved Yolo-v4 algorithm (a) Video 1, (b) Video 2, (c) Video 3, (d) Video 4.

where T_P and T_N are the number of correctly detected images with and without fires, respectively; F_P is the number of images falsely detected to contain fires when there are none; F_N is the number of images missed when there are fires; N_{pos} and N_{neg} are the number of images in the samples with and without fires, respectively.

The video-based detection results with the improved Yolo-v4 algorithm are presented in Table 2. The proposed algorithm achieves 98.9% accuracy, a falsely detection rate of 4.8%, and a miss rate of 1.1%. With such high accuracy, the algorithm can be applied in the detection system.

TABLE 3. Detection results of videos with and without fires.

Detection results of videos with fires					
No.	Total frames	Number of correctly detected frames	Accuracy/%	Number of missed frames	Miss rate/%
1	1,315	1,306	99.3	9	0.7
2	2,307	2,282	98.9	27	1.1
3	2,496	2,481	99.4	15	0.6
4	1,500	1,476	98.4	24	1.6
5	2,995	2,953	98.6	42	1.4
6	1,769	1,741	98.4	28	1.6
7	1,562	1,546	99.0	16	1
8	5,857	5,804	99.1	53	0.9
9	954	947	99.3	7	0.7
10	850	841	98.9	9	1.1
Average	2,161	2,138	98.9	23	1.1

Detection results of videos without fires			
No.	Total frames	Number of falsely detected frames	False detection rate/%
11	3,075	141	4.6
12	2,325	116	5.0
13	850	43	5.1
14	568	27	4.7
15	937	45	4.6
Average	776	37	4.8

The results are also illustrated in Fig. 8. The number in the upper left corner of a red box indicates the probability of the model judging fires in the region. In this paper, dynamic fire characteristics are used to rule out images without fires to lower the false detection rate and optimize the accuracy of fire detection.

D. ABLATION EXPERIMENT

An ablation experiment is presented in this paper to demonstrate what role the modules of the improved YOLO-v4 play in the optimization of YOLO-v4. The PAN structure of feature pyramid in YOLO-v4 is replaced by the Bi-FPN-Lite structure in this paper to compare the improved YOLO-v4 and YOLO-v4 algorithms. Table 3 presents the results of the ablation experiment with the improved YOLO-v4. In the table, Bi-FPN-Lite implies that the Bi-FPN-Lite structure is used in place of the PAN feature fusion network in YOLO-v4. The ablation experiment is performed to compare the algorithms in such two structures in terms of mAP, parameter, and false detection rate. It is evident that the improved YOLO-v4 has the PAN structure in place of the Bi-FPN-Lite structure, which improves mAP by around 1.5% but causes very tiny change to the model size. It is therefore concluded that

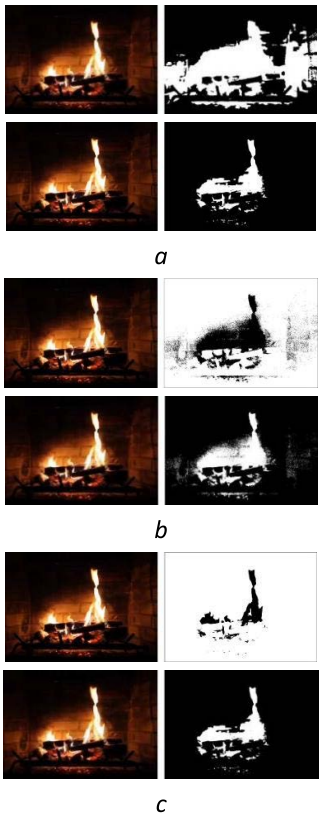


FIGURE 9. Comparison of three algorithms in terms of segmentation after adjusting to sudden changes of light (a) Mixtures of Gaussians, (b) ViBe, (c) Improved ViBe(the first frame taken when fires were detected at the top; the first frame after adjusting to sudden changes of light at the bottom.

Bi-FPN-Lite can certainly improve the mAP of algorithm and reduce the false detection rate with little influence on model size.

E. REDUCTION OF FALSE DETECTION RATE WITH DYNAMIC FIRE CHARACTERISTICS

The frames judged to contain fires in the detection results were detected again considering dynamic fire characteristics. The parameters for dynamic fire characteristics were calculated within the time period in which these frames were detected. Taking Videos 8, 9, 11 and 12 as examples, a video was taken for a time period of five seconds in which the frames were detected to contain fires. These five-second videos were used to analyze the change of flicker frequency. In Video 8, flicker frequency varied between 8 Hz and 10 Hz, which fell into the normal flicker frequency range of fires, 7∼10 Hz, so that it was judged that the video did contain fires. In Video 9, flicker frequency was enhanced under the effect of external electric fields and exceeded the normal range, so that it was judged that the video did contained fires. In Video 11, flicker frequency fell into the normal range, but fires showed a low rate of area change, so that it was determined that there were no fires. In Video 12, fires flickered in the normal frequency but had lower circularity, so that it was determined that there were no fires. In this way, the interference caused by the objects similar to fires to detection can be eliminated to effectively reduce the false detection rate and enhance detection accuracy. The results of the optimized video-based fire detection are given in Table 3.

In video-based fire detection, the improved algorithm with dynamic fire characteristics helped achieve 98.9%accuracy, a false detection rate of 1.7%, and a miss rate of 1.1%. In order to further prove its accuracy in fire detection, the proposed algorithm was compared with several other algorithms including the nine-layer convolutional neural network in [21], the two-level structural convolutional neural network in [22], the 12-layer convolutional neural network in [23], and other Yolo algorithms including Yolo-v1, Yolo-v2, Yolo-v3, Yolo-v4, and Yolo-v5. The twelve videos used in this experiment were also employed in the comparative analysis. The detection results with these algorithms are presented in Table 4.

Table 3 reveals that the proposed algorithm has better accuracy and much higher detection speed than the nine-layer, two-level structural, or 12-layer convolutional neural networks. Among the Yolo algorithms, the proposed algorithm is more accurate than Yolo-v1, Yolo-v2, Yolo-v3, Yolo-v4, and Yolo-v5 by 17.2%, 13.5%, 2.7%, 1.0%, and 6.3%, respectively, and it realizes a lower false detection rate than the others by 13.5%, 7.7%, 4.6%, 0.8%, and 6.5%, respectively. In terms of speed, it is slightly slower than Yolo-v4 and Yolo-v5, but its speed of 55 frames per second still meets the requirements for fast fire detection.

F. EXTRACTION OF FIRE TARGETS WITH IMPROVED VIBE ALGORITHM

The first image detected to contain fires was taken as the first frame. It was compared with the first frame captured after adjusting to sudden changes of light to establish the threshold adaptation speed of these algorithms.

As shown in Fig. 9, the mixtures of Gaussians algorithm achieved good image segmentation while facing sudden changes of light but experienced slight interference and less accurate segmentation after adaptation. Comparatively, the ViBe algorithm was more accurate than the mixtures of Gaussians algorithm but caused white noise. The improved ViBe algorithm proposed in this paper realized the fastest adaptation, that is, 0.69s, as given in Table 5, which was much better than the other two algorithms.

G. FIRE CHARACTERISTICS UNDER THE EFFECT OF ELECTRIC FIELDS

In Fig. 10, the parameters for fire characteristics are presented under the effect of different electric fields. In the experiment, we used a ceramic tube as the combustor and imposed adjustable external electric fields with an electrode spacing of 55 mm and ethanol flow 63 ml/h. The curves of these parameters reveal that the fire height decreases with the

TABLE 4. Results of optimized video-based fire detection.

No.	Total frames	Number of falsely detected frames	False detection rate
11	550	3	0.6
12	668	10	1.5
13	850	14	1.7
14	568	12	2.1
15	937	24	2.6
Average	715	13	1.7

TABLE 5. Comparison of detection results with different algorithms.

Algorithm	Accuracy /%	False detection rate /%	Speed/(frame·S^{-1})
Ref. [21]	81.6	17.8	29
Ref. [22]	88.7	18.7	27
Ref. [23]	90.3	19.9	21
Yolo-v1	82.7	15.2	43
Yolo-v2	85.4	9.4	55
Yolo-v3	96.1	6.3	62
Yolo-v4	97.4	2.5	65
Yolo-v5s	92.6	7.9	89
The proposed algorithm	98.9	1.7	55

TABLE 6. Comparison of algorithms.

Algorithm	Number of frames for adaptation	Adaptation time
Mixtures of Gaussians	33	1.32
ViBe	47	2.01
Improved ViBe	16	0.69

increasing voltage within the voltage range in the experiment. Thus, the electric field can compress fires in combustion. Direct-current (DC) electric fields can compress fires more strongly than alternating-current (AC) fields. Moreover, DC electric fields affect the circularity of fires, i.e., stable combustion, more significantly than AC electric fields. The change of fires under the effect of electric fields follows the same rules as concluded in [2], [3], and [17]. The data obtained from image processing in the system is consistent with that from the artificial image processing frame by frame. Therefore, it can provide theoretical support for further developing the extinguishing techniques of fires under the effect of electric fields.

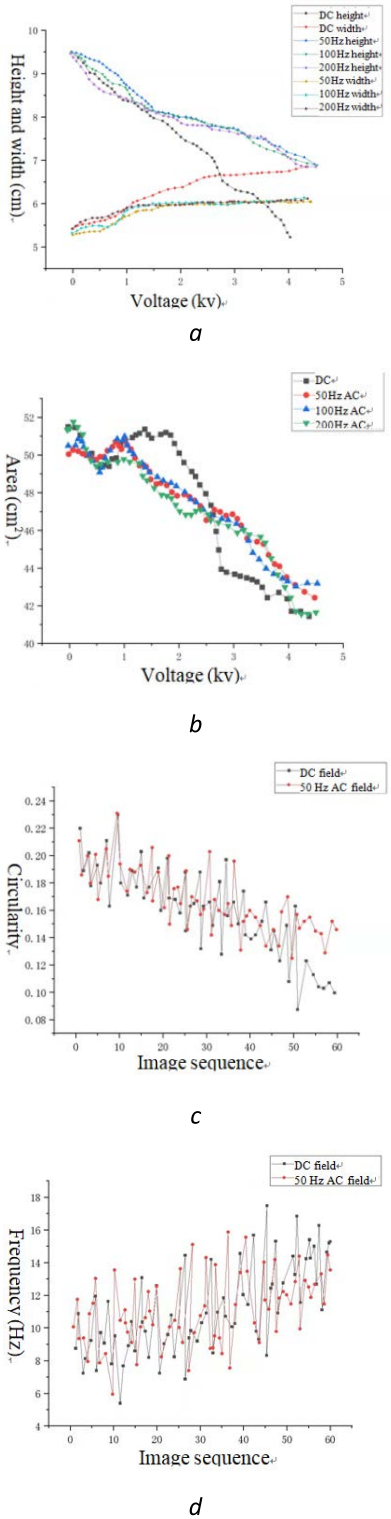


FIGURE 10. Fire characteristics under the effect of different electric fields (a)Fire width and height under the effect of different electric fields, (b)Fire area under the effect of different electric fields, (c) Fire circularity under the effect of different electric fields, (d) Fire flicker frequency under the effect of different electric fields.

VI. CONCLUSION

As surveillance systems become more intelligent, exploring fire detection with surveillance videos based on dynamic

fire characteristics grows in significance. The traditional image-based fire detection methods and existing fire detection algorithms do not perform well in detecting unsteady fires because of their dynamic characteristics. For this reason, a monocular visual recognition fire detection system is designed with the improved Yolo-v4 and ViBe algorithms. By using the simplified Bi-FPN with attention mechanism in place of the PANet feature fusion network in the Yolo-v4 algorithm, the system can perform better at fire detection. Meanwhile, dynamic fire characteristics are considered in fire detection to significantly eliminate misjudgments. The threshold update mechanism of the ViBe algorithm is adjusted to cope with the sudden change of light caused by frequent fire flickering under the effect of external electric fields. In this way, unsteady fires can be accurately detected and tested with dynamic characteristics. Experiments have been carried out to verify the efficacy of the designed detection system. Compared with the techniques available, this fire detection system can more effectively detect unsteady fires with multiple dynamic characteristics under the effect of electric fields.

REFERENCES

[1] M. Šícha, "Measurement of the electron energy distribution function in flame plasma at atmospheric pressure," *Czechoslovak J. Phys.*, vol. 29, no. 6, pp. 640–645, Jun. 1979.
[2] O. Imamura, K. Yamashita, S. Nishida, G. Ianus, M. Tsue, and M. Kono, "Two-droplet combustion of N-Octane in a direct current electric field under microgravity," *Combustion Sci. Technol.*, vol. 183, no. 8, pp. 755–763, Apr. 2011.
[3] D. G. Park, S. H. Chung, and M. S. Cha, "Bidirectional ionic wind in nonpremixed counterflow flames with DC electric fields," *Combustion Flame*, vol. 168, pp. 138–146, Jun. 2016.
[4] E. Sher, G. Pinhasi, A. Pokryvailo, and R. Bar-On, "Extinction of pool flames by means of a DC electric field," *Combustion Flame*, vol. 94, no. 3, pp. 244–252, 1993.
[5] T Celik, H. Demirel, H. Ozkaramanli, and M. Uyguroglu, "Fire detection using statistical color model in video sequences," *Image Represent.*, vol. 18, no. 2, pp. 176–185, 2007.
[6] T.-H. Chen, P.-H. Wu, and Y.-C. Chiou, "An early fire-detection method based on image processing," in *Proc. Int. Conf. Image Process.*, 2004, pp. 1707–1710.
[7] P. V. K. Borges and E. Izquierdo, "A probabilistic approach for vision-based fire detection in videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 5, pp. 721–731, May 2010.
[8] M. Mueller, P. Karasev, I. Kolesov, and A. Tannenbaum, "Optical flow estimation for flame detection in videos," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2786–2797, Jul. 2013.
[9] K. Muhammad, J. Ahmad, and S. W. Baik, "Early fire detection using convolutional neural networks during surveillance for effective disaster management," *Neurocomputing*, vol. 288, pp. 30–42, May 2018, doi: 10.1016/j.neucom.2017.04.083.
[10] J. Kuhl, T. Seeger, L. Zigan, S. Will, and A. Leipertz, "On the effect of ionic wind on structure and temperature of laminar premixed flames influenced by electric fields," *Combustion Flame*, vol. 176, pp. 391–399, Feb. 2017.
[11] A. Ata, J. S. Cowart, A. Vranos, and B. M. Cetegen, "Effects of direct current electric field on the blowoff characteristics of bluff-body stabilized conical premixed flames," *Combustion Sci. Technol.*, vol. 177, no. 7, pp. 1291–1304, Jul. 2005, doi: 10.1080/00102200590950476.
[12] S. Marcum and B. Ganguly, "Electric-field-induced flame speed modification," *Combustion Flame*, vol. 143, nos. 1–2, pp. 27–36, Oct. 2005.
[13] X. Meng, X. Wu, C. Kang, A. Tang, and Z. Gao, "Effects of direct-current (DC) electric fields on flame propagation and combustion characteristics of premixed $CH_4/O_2/N_2$ flames," *Energy Fuels*, vol. 26, no. 11, pp. 6612–6620, Nov. 2012.
[14] Y. Luo, Z. Jiang, Y. Gan, J. Liang, and W. Ao, "Evaporation and combustion characteristics of an ethanol fuel droplet in a DC electric field," *J. Energy Inst.*, vol. 98, pp. 216–222, Oct. 2021.
[15] Y. Luo, Y. Gan, and Z. Jiang, "An improved reaction mechanism for predicting the charged species in ethanol-air flame," *Fuel*, vol. 228, pp. 74–80, Sep. 2018.
[16] Y. Luo, Y. Gan, J. Xu, Y. Yan, and Y. Shi, "Effects of electric field intensity and frequency of AC electric field on the small-scale ethanol diffusion flame behaviors," *Appl. Thermal Eng.*, vol. 115, pp. 1330–1336, Mar. 2017.
[17] T. Z. Ge, "Experimental and numerical simulation of electrical characteristics of laminar flame quenching," M.S. thesis, Harbin Inst. Technol., Harbin, China, 2018.
[18] J. Huyan, W. Li, S. Tighe, Z. Xu, and J. Zhai, "CrackU-net: A novel deep convolutional neural network for pixelwise pavement crack detection," *Struct. Control Health Monitor.*, vol. 27, no. 8, p. 2551, Aug. 2020.
[19] W. Sun, A. R. C. Paiva, P. Xu, A. Sundaram, and R. D. Braatz, "Fault detection and identification using Bayesian recurrent neural networks," *Comput. Chem. Eng.*, vol. 141, Oct. 2020, Art. no. 106991.
[20] R. Sonobe, Y. Hirono, and A. Oi, "Non-destructive detection of tea leaf chlorophyll content using hyperspectral reflectance and machine learning algorithms," *Plants*, vol. 9, no. 3, p. 368, Mar. 2020.
[21] S. Frizzi, R. Kaabi, M. Bouchouicha, J.-M. Ginoux, E. Moreau, and F. Fnaiech, "Convolutional neural network for video fire and smoke detection," in *Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2016, pp. 877–882.
[22] Q. Zhang, J. Xu, L. Xu, and H. Guo, "Deep convolutional neural networks for forest fire detection," in *Proc. Int. Forum Manage., Educ. Inf. Technol. Appl.*, 2016, pp. 568–575.
[23] T. J. Fu, C. E. Zheng, Y. Tian, Q. M. Qiu, and S. J. Lin, "Forest fire recognition based on deep convolutional neural network under complex background," *Comput. Modernization*, vol. 3, pp. 52–57, Jan. 2016.
[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[25] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6517–6525.
[26] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[27] A. Bochkovskiy, C. Y. Wang, and H. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[28] G. Jocher. *YOLOv5*. [Online]. Available: https://github.com/ultralytics/yoloV5/issues/6
[29] S. S. Sengar and S. Mukhopadhyay, "Foreground detection via background subtraction and improved three-frame differencing," *Arabian J. Sci. Eng.*, vol. 42, no. 8, pp. 3621–3633, Aug. 2017.
[30] Q. U. Zhong and X. L. Huang, "Algorithm of enhanced visual background extraction for eliminating ghost," *Comput. Sci.*, 2017.
[31] C. D. Yang and Q. Meng, "Algorithm of micro-motion object detection based on ViBe and multi-feature extraction," *Comput. Sci.*, vol. 44, no. 2, pp. 309–312, 316, 2017.
[32] Q. Zhang, W. Lu, C. Huang, W. Lian, and X. Yang, "An adaptive vibe algorithm based on dispersion coefficient and spatial consistency factor," *Autom. Control Comput. Sci.*, vol. 54, no. 1, pp. 80–88, Jan. 2020.
[33] H. Kaiming, Z. Xiangyu, R. Shaoqing, and S. Jian, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Jan. 2015.
[34] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," *CoRR*, vol. abs/1803.01534, pp. 8759–8768, Jun. 2018.
[35] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.
[36] C. Chen, "Fire detection based on parallel computing of motion and flicker frequency feature," *J. Data Acquisition Process.*, 2017.
[37] T. Toulouse, L. Rossi, A. Campana, T. Celik, and M. A. Akhloufi, "Computer vision for wildfire research: An evolving image dataset for processing and analysis," *Fire Saf. J.*, vol. 92, pp. 188–194, Sep. 2017.
[38] A. Cetin. *Fire Detection Samples*. [Online]. Available: http://www.Signal.ee.bilkent.edu.tr/VisiFire
[39] C. R. Steffens, S. S. D. C. Botelho, and R. N. Rodrigues, "A texture driven approach for visible spectrum fire detection on mobile robots," in *Proc. 13th Latin Amer. Robot. Symp., 4th Brazilian Robot. Symp. (LARS/SBR)*, Oct. 2016, pp. 257–262.

SHUAI ZHAO was born in Qinhuangdao, Hebei, China, in 1994. He is currently pursuing the master's degree with the School of Power Engineering, Naval University of Engineering, Wuhan, China.

His research interest includes electric field driven fire dynamics in vessels.

BOYUN LIU was born in Fuqing, Fujian, China, in 1977. He received the Ph.D. degree.

He is currently an Associate Professor and a Supervisor for graduate students. He took part in the research and development of safety equipment for multiple vessel models, including respirator attached inflator system and quick fire smoke removal system. He completed more than ten key technical research and breakthrough projects, and received three prizes for progress in military science and technology. He is the author of more than 90 published papers and nine national patents. His research interests include teaching and research of vessel safety technology and engineering.

ZHENG CHI was born in Wuhan, Hubei, China, in 2000. He is currently pursuing the bachelor's degree with the School of Power Engineering, Naval University of Engineering, Wuhan, China.

His research interests include fire detection and characteristic test.

TAIWEI LI was born in Dalian, Liaoning, China, in 1999. He is currently pursuing the bachelor's degree with the School of Weaponry Engineering, Naval University of Engineering, Wuhan, China.

His research interests include computer vision and innovative fire extinguishing technology.

SHENGNAN LI, photograph and biography not available at the time of publication.