SMOKE VIDEO DETECTION BASED ON DOUBLE SPECTRUM

Changlin Song,* Chuanfu Liu,* Tingping Feng,* Junmin Li,* Yanlin Pan,* and Simon X. Yang**

Abstract

Smoke is a noticeable feature when a fire occurs. Therefore, the detection of smoke becomes an important research direction for fire prevention. At present the research on smoke detection is mainly based on the Convolutional Neural Network (CNN) under the visible light condition, which is not suitable for the changeable outdoor environment. Therefore, a dual-spectral video smoke detection system based on an improved Faster R-CNN is proposed. Firstly, the fusion network of the VGG16 network and the improved Convolutional Block Attention Module (CBAM) network is used to extract features, which is used to solve the problem that it is difficult to extract smoke colour features, shape features and texture features in real-time. Secondly, the non-maximum suppression (NMS) algorithm of Faster R-CNN is improved to solve the precision degradation problem. Finally, the infrared thermal imaging camera is used to identify the heat source after selecting the smoke, which solves the problem that white clouds, fog and other similar objects are misidentified as smoke and improves the identification system's accuracy. The experimental results show that the improved model is superior to the visible light detection method only and the method can be well used for smoke detection.

Key Words

Convolutional neural network, faster R-CNN, Infrared thermal imaging video smoke detection

1. Introduction

A fire is a disaster caused by uncontrolled combustion in time or space. Among all types of disasters, fire is one of the most frequent and highly destructive ones. The occurrence of fire is not only a great threat to people's lives and property, but also a great damage to the ecological environment. Currently, the mainstream method uses smoke sensors to monitor smoke concentration to prevent fires, but it requires smoke to enter the sensor and be identified only when the concentration reaches a certain level. Therefore, it is difficult to use such a sensor in an open outdoor space.

In the outdoor environments, computer vision-based smoke detection methods are often used because they are not limited by space. This method also has the characteristics of large coverage area and low cost, which is one of the main directions of outdoor smoke detection research [1]. Therefore, in this paper, based on deep learning and image processing theory, this paper proposes a dual-spectrum based video smoke detection method considering both visible and infrared thermal image information of smoke.

The main contributions are as follows:

Firstly, a fusion network is proposed based on the two-stage object detection algorithm Faster R-CNN. The smoke features are extracted by fusing the VGG16 network with the improved CBAM module.

Secondly, the NMS algorithm is improved to eliminate large frames over small frames and intersecting frames with IOU less than the threshold.

Finally, an algorithm combining visible light image and infrared image is proposed. The frames are processed by comparing the frame coordinates of visible light detection and the frame coordinates of infrared image heat source detection.

2. Related Work

In the past decades, a large number of smoke detection methods based on traditional image processing have been explored, and the detection accuracy has been improved to some extent. Smoke detection is inseparable from the feature extraction of smoke, and traditional methods generally use color features, shape features, texture features, and dynamic features as the basis for smoke detection. In terms of color features, Emmy Prema *et al.* [2] proposed an algorithm to convert the image from RGB color space to YUV color space and then perform image segmentation in order to reduce the effect of luminance on smoke color. In terms of shape features, T.-H. Chen *et al.* [3] used the ratio of the perimeter of the smoke region to its area as an indicator for detecting smoke irregularities. In terms of texture features of the smoke representation. Dong LF *et al.* [4] used local binary pattern (LBP)

* School of Mechanical Engineering, Xihua University, Chengdu 610039, P.R. China; e-mail: {345095081, 823863504, 1365615549, 553904755}@qq.com; lijunmin1975@163.com
** Advanced Robotics and Intelligent Systems Laboratory, School of Engineering, University of Guelph, Guelph, Canada ONN1G2W1; e-mail: syang@uoguelph.ca.
Corresponding author: Junmin Li, Tingping Feng
Co-first author: Changlin Song, Yanlin Pan

Recommended by Huaicheng Yan
(DOI: 10.2316/J.2024.206-1020)

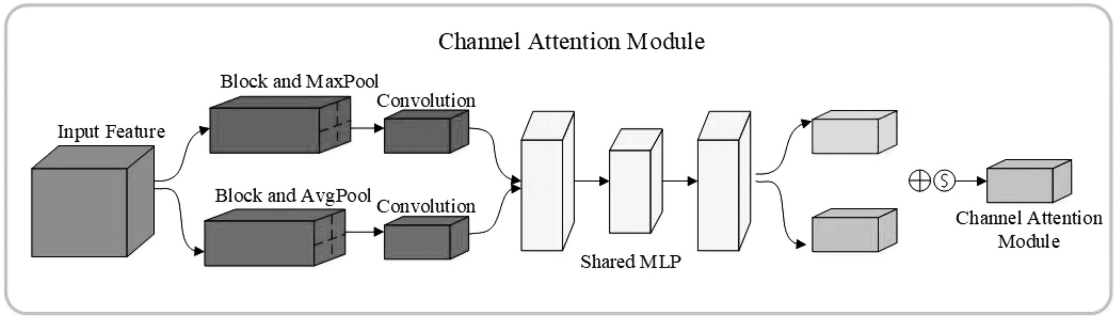


Figure 1. Channel attention module.

extraction features as one of the features to discriminate the smoke.

In recent years, Convolutional Neural Networks (CNNs) [5] have achieved better results in object detection. Yin *et al.* [6] constructed novel networks based on the classical ZF-Net. By replacing the convolutional layer with a normalized convolutional layer, thus building a deeply normalized convolutional neural network. Hu *et al.* [7] proposed an enhanced network structure with a multi-task learning strategy, thus capturing features of both intra-frame appearance and inter-frame motion of video. hu *et al.* [8] proposed an end-to-end smoke detection and smoke presence prediction framework. yuan *et al.* [9] proposed a smoke segmentation of the network structure. The network is divided into two routes, the first route is a deep neural network that extracts global contextual information about smoke and the second route is a shallow neural network that extracts detailed information about smoke. Zhang *et al.* [10] used a Faster R-CNN network to train synthetic data and compared the differences between synthetic smoke and real smoke used for training. Lin *et al.* [11] proposed Faster R-CNN and 3D CNN joint detection framework. Li *et al.* [12] proposed high-resolution remote sensing image segmentation method based on SReLU. Li *et al.* [13] proposed segmentation method of high-resolution remote sensing image for fast target recognition. Liu *et al.* [14] proposed the improved YOLOv3 for an engineering vehicle detection. Tran *et al.* [15] proposed the application of deep images and neural networks in robot 3D maps.

3. Our Method

In this paper, a detection method based on fused deep learning and infrared thermography is proposed to improve the detection accuracy of smoke in video. First, we input the visible video image into the fusion network of VGG16 and improved CBAM to extract features. Secondly, an improved NMS (non-maximum suppression) algorithm is applied to the smoke detection box. Finally, we compare the object boxes of visible light and infrared thermal imaging detection to get the final detection.

3.1 Fusion Network

Attention mechanisms can improve the extraction of object detail features by neural networks. CBAM [16] contains networks with attention in the spatial domain and attention in the channel domain. Its channel attention network structure is based on an improvement of SEnet [17]. VGG16 is a commonly used network with a simple structure that integrates well with the improved CBAM. Due to the use of maximum pooling method among the sub-blocks of VGG16 network, some important features of the original smoke images with rich details may be lost. For this reason, this paper improves the VGG16 network by introducing an improved CBAM module, which enhances the extraction of detailed features. Compared with the individual networks of the VGG16 network, the proposed fusion network is able to extract detail features and enhance the distinction between smoke features and non-smoke features.

3.1.1 *Improved CBAM Module*

In CBAM networks, maximum pooling and average pooling of channel and space can lose the detailed feature. Thus, inspired by ROI Pooling, we propose a novel channel attention and spatial channel attention module (shown in Fig. 1).

In the channel attention module, the input feature is first chunked to two blacks, and perform MaxPool and AvgPool operations respectively, then using 1×1 conv to encode the feature after pooling. Further, the features are fed to a shared multilayer perception (Shared MLP) for processing, which contains two fully connected layers that reduce and then raise the number of channels. The first fully-connected layer reduces the output channel number to 1/r of the original channel number; the second fully-connected layer scales the channels to r and adjusts back to the original channel number. Finally, the results of the two feature layers are summed, and then the channel attention weights are generated by the sigmoid gate mechanism. The formula for calculating the maximum value or mean value of the block is as follows:

$$h = \left[\frac{H}{n_h}\right] \tag{1}$$

$$w = \left[\frac{W}{n_w}\right] \tag{2}$$

$$f(n_h, n_w) = \max\left(B(n_h, n_w)\right) \tag{3}$$

$$f(n_h, n_w) = \text{mean}\left(B(n_h, n_w)\right) \tag{4}$$

where H and W are the height and width of the input feature map, and h and w are the height and width of the

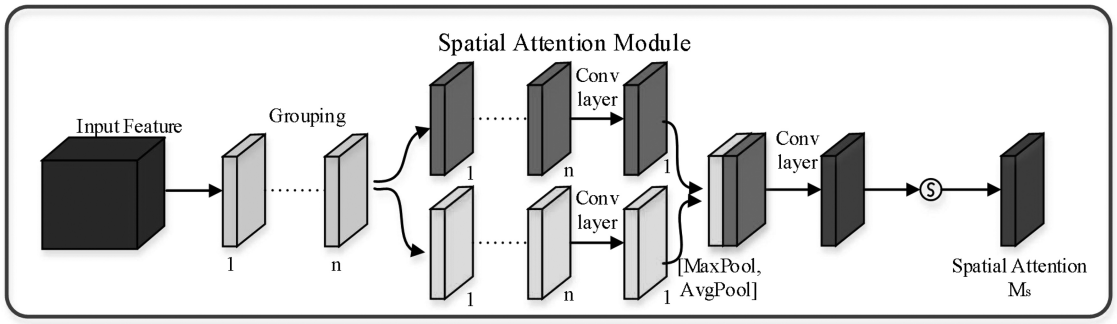


Figure 2. Spatial attention module.

Figure 3. The pink box is the box recognized after training.

convolution kernel. n_h and n_w are the number of blocks in the vertical and horizontal directions, respectively, and n_h and n_w is set to 9. $f(n_h, n_w)$ represent the pixel value in (n_h, n_w) in the output feature map. $b(n_h, n_w)$ is the block of size (h, w) in the n_hth row and n_wth column. $B(n_h, n_w)$ is the feature map block of size (h, w). In order to find the maximum or mean value of the feature block, it needs to be convolved with the convolution kernel of h_w.

In the spatial attention module (shown in Fig. 2), firstly, the input feature is divided into n parts, and then calculate the mean value and maximum value are on the channel dimension. Secondly, the feature map is convolved with a 1×1 conv. Then the obtained feature map is concatenated on the dimension channel to generate the 2-dimension feature map. Finally, the spatial attention weights are generated after the convolution and sigmoid functions. The formula for calculating the mean and maximum values of the dimensions is as follows:

$$f_k(i,j) = \max_{\dim}(F_k) \quad (5)$$

$$f_k(i,j) = \text{mean}_{\dim}(F_k) \quad (6)$$

where the $f_k(i, j)$ is the pixel value of the k-th dimension, i-th row, and j-th column. F_k is divided into k feature maps.

3.1.2 *Improved NMS Algorithm*

Considering that the shape and boundary of smoke are not fixed and the concentration size is not consistent in various places within it, there may be mis-localization, which leads to a large box with a small box or two inter small boxes within the smoke area. The IOU value of these boxes and the area of the labeled boxes is less than the threshold, so these boxes are considered False Positives (FP). The NMS algorithm is improved to eliminate these FP boxes.

The sparse and variable attribute of smoke leads to the appearance of multiple boxes for the same object during detection. As shown in Fig. 3, multiple boxes with different confidence appear for the same smoke, and the redundant boxes cannot be eliminated by the existing NMS algorithm. Based on this, an improved NMS method is proposed in this paper, and its main flow chart is shown in Fig. 4.

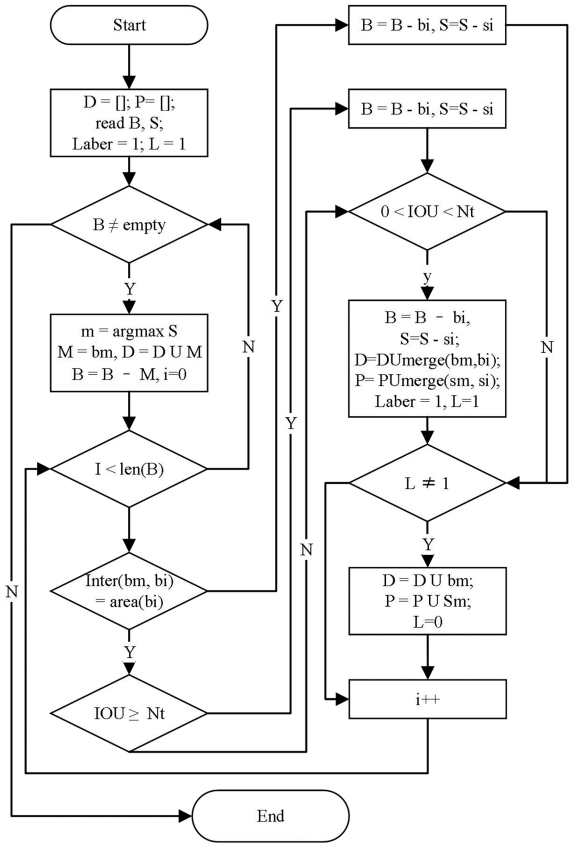


Figure 4. Improved NMS flow chart.

where B and S represent the set of detection boxes and corresponding scores. D and P represent the set of boxes and scores. The *Laber* and L are flag bits. The Inter(b_m, b_i) represents the area of the intersection of b_m and b_i. The Merge(b_m, b_i) represents the fusion of b_m and b_i box. The Merge(s_m, s_i) represents the score after the fusion of b_m and b_i box. The improved NMS algorithm adds two parts: Inter(b_m, b_i) = area (b_i) and $IOU < N_t$. The first part is to eliminate the small box in the big box; the second part is to fuse the intersecting boxes.

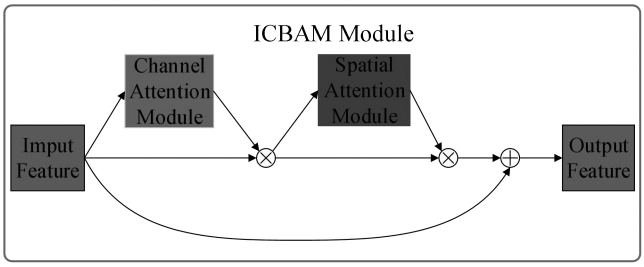


Figure 5. ICBAM module.

To eliminate the small boxes in the large box, we need to determine if the little box is inside the large box. Therefore, we need to determine whether the area of the intersection of box BM and box BI is equal to the area of small box BI. If they are equal, box BI is inside box BM. We eliminate box BI; Otherwise, it stays the same. For boxes less than the IOU threshold, the intersecting boxes are fused to eliminate the boxes with low scores and resize them. Based on the algorithm proposed by Zhou [18], this paper presents an improved NMS algorithm. First, the minimum value of the upper-left coordinate of B_m and B_i box and the lower right coordinate's maximum value are combined to form the new box $B_{\max}$. Secondly, the two boxes $B_{\max}$ and B_m are weighted combined.

$$X_1 = \min(x_{11}, x_{21}); Y_1 = \min(y_{11}, y_{21}) \quad (7)$$

$$X_2 = \min(x_{21}, x_{22}); Y_2 = \min(y_{21}, y_{22}) \quad (8)$$

$$B_{\text{merge}} = \frac{S_m * B_m + S_i * B_{max}}{S_m + S_i} \quad (9)$$

$$S_{\text{merge}} = S_m \quad (10)$$

In the above equation, (x_{11}, y_{11}, x_{12}, y_{12}) is the coordinate of B_m, (x_{21}, y_{21}, x_{22}, y_{22}) is the coordinate of B_i, (X_1, Y_1, X_2, Y_2) is the coordinate of $B_{\max}$, B_m and B_i represent two adjacent the box, S_m and S_i represent the score of the box. B_{merge} and S_{merge} represent weighted merged boxes and scores.

3.1.3 *Fusion Network*

The connexion of channel attention and spatial attention modules in the network is shown in Fig. 5.

In the Improved Convolutional Block Attention Module (ICBAM) module, the channel attention and spatial attention modules are connected in the network. Figure 6 is the structure diagram of the improved faster R-CNN network structure with an increased attention mechanism proposed in this paper.

The network structure before Block 5 is same to VGG16. The ICBAM module is located between the last two layers of Block 5. The improved feature extraction network is shown in the dotted box. Through experiments, we found that directly connecting the ICBAM module would lead to non-convergence of the network. Therefore, the BN layer was added to solve this problem on the second convolution and third layers of the Block 5.

3.2 Fusion Algorithm of Visible Light and Infrared Thermal Imaging

3.2.1 *Infrared Thermal Imaging Algorithm*

The infrared detection algorithm is mainly used to detect objects containing heat sources, and exclude white clouds and other white objects without heat sources. In this paper, we propose a processing algorithm combining gamma transformation [19], Gaussian filter and grayscale morphological operation. Firstly, power law transformation is used to suppress pixel values at normal temperature, and gaussian filtering is used to eliminate interference pixels. Secondly, image filtering is used to remove outliers, fill holes and smooth edges by the morphological operation. Finally, image segmentation uses maximum inter-class variance [20], fills the holes of the image through morphological dilation operations, and then extracts the image's features to obtain the object frame coordinates. The flow of the algorithm is shown in Fig. 7.

3.2.2 *Fusion Algorithm of Visible Light and Infrared Thermal Imaging*

The comparison algorithm of box coordinates for visible image detection of smoke and those for infrared image detection of smoke is shown in Fig. 8. Firstly, we obtain the coordinates of the visible light and infrared thermal imaging object detection frame. Secondly, box coordinates are expanded for visible image detection of smoke. Finally, the expanded visible smoke detection box coordinates are compared with those of the infrared smoke box to obtain the final smoke detection frame's coordinates.

The flowchart of the box expansion algorithm and the box comparison algorithm are shown in the figure below (shown in Fig. 9).

The coordinates of the frame are different on the infrared and visible light images with different resolutions. Therefore, the frame mapping coordinate algorithm is used to solve this problem. By comparing a large number of visible light and infrared images, we get a 50 pixels difference in the Y-axis direction of the visible light and infrared images. Therefore, the mapping algorithm is an increase of 50 pixels values in the Y axis of the infrared coordinate.

The formula of the box expansion algorithm were given as:

$$\begin{aligned} w &= x_2 - x_1 \\ h &= y_2 - y_1 \\ x_{\text{center}} &= x_1 + w/2 \\ y_{\text{center}} &= y_1 + h/2 \end{aligned} \quad (11)$$

$$\begin{aligned} xx_1 &= x_{\text{center}} - \alpha^* w/2 \\ yy_1 &= y_{\text{center}} - \alpha^* h/2 \\ xx_2 &= x_{\text{center}} + \alpha^* w/2 \end{aligned} \quad (12)$$

Equation 11 is an algorithm for transforming coordinate values (x_1, y_1, x_2, y_2) into centre coordinate values (w, h, x_{center} , y_{center}). Equation 12 is an algorithm for enlarging the box, where α is the coefficient for enlarging the width and height, which is two in this paper.

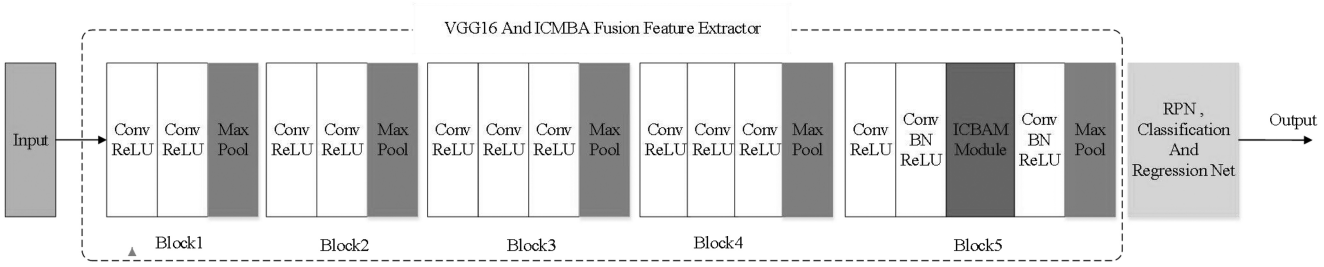


Figure 6. Improved faster R-CNN network structure.

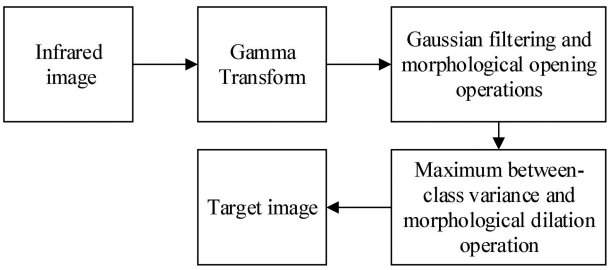


Figure 7. Infrared image processing algorithm flow.

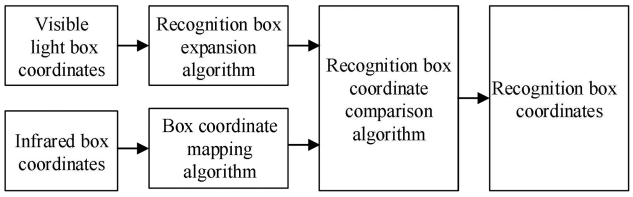


Figure 8. Algorithm flow.

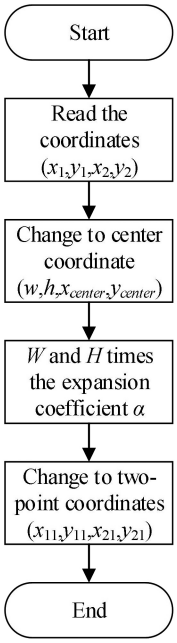


Figure 9. Expanded algorithm flow chart.

The comparison algorithm of the two boxes were given as:

$$\eta = area/area\ 2 \tag{13}$$

$$\text{box} = \begin{cases} (x_1, y_1, x_2, y_2) & \eta \geq T \\ 0 & \eta < T \end{cases} \tag{14}$$

In the above formula, the area represents the area where the visible light and the infrared object detection box intersect; the *area 2* is the area of the infrared thermal imaging object detection box. η is the ratio; the T is the threshold. When η is greater than or equal to T, the visible box coordinates are obtained; otherwise, there are no box coordinates. The value of T in this paper is 0.5.

3.2.3 *Smoke Detection System of Visible and Infrared Thermal Imaging*

The flow chart of visible light and infrared image detection system is shown in Fig. 11.

Firstly, visible light images are obtained by dual spectral imaging. Secondly, the image is classified and recognized by the network model. Finally, read the infrared image and judge the heat source. If there is a heat source, a comparison algorithm is implemented to eliminate the non-smoke box to obtain the final detection box and improve the detection accuracy.

4. Experiments

4.1 Experiment Description

4.1.1 *Experimental Dataset*

To better validate the effectiveness of the proposed method, we constructed a more comprehensive dataset which is about 2000 smoke images. And the images come from the Internet, the laboratory of Bilken University in Turkey [21], the laboratory of Qiming University in Korea [22] and self-shot. Further more, 80% of the dataset is divided into a trainset and 20% into a test set (shown in Fig. 12).

4.1.2 *Video Test Dataset*

Since this paper considers combining the information of visible images and infrared images to improve the detection accuracy, but there does not exist a dataset containing both types of these images. Therefore, in this paper, the video dataset is obtained by visible light cameras and infrared thermal imaging cameras, which contains a total of six

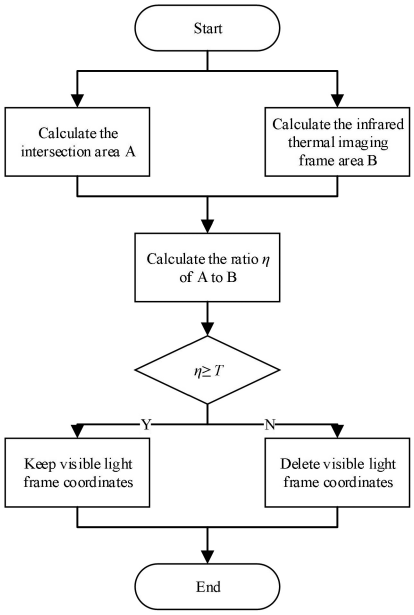


Figure 10. Flow chart of box comparison algorithm.

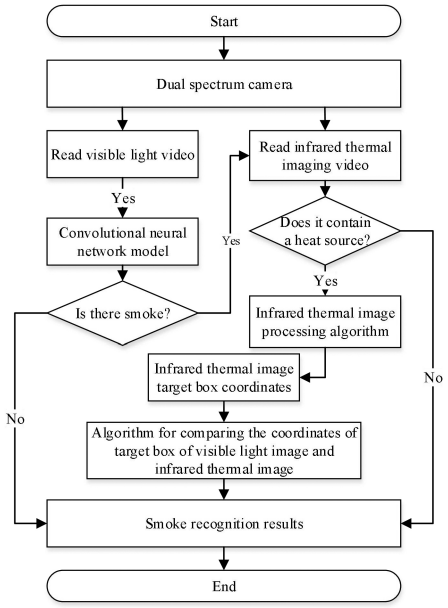


Figure 11. Flow chart of the detection system.

Figure 12. Smoke dataset.

videos, as shown in Fig. 13, and the details of the videos are shown in Table 1.

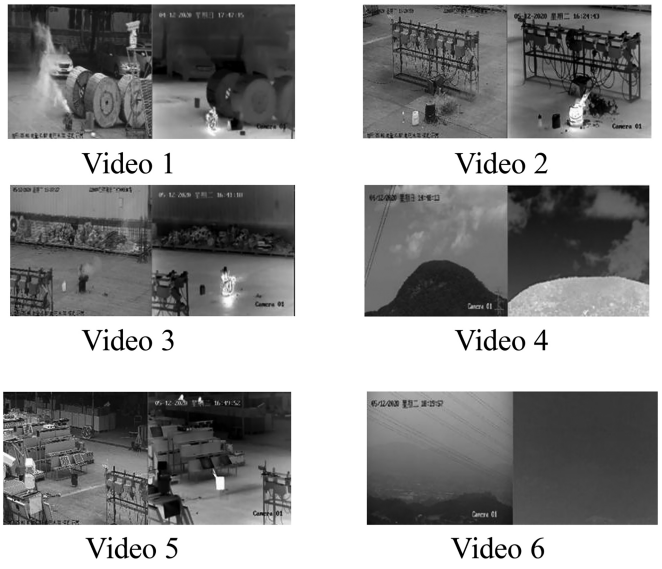


Figure 13. Test video.

4.1.3 *Evaluation Indicators*
Accuracy (ACC), Precision, Recall and Average accuracy are used to compare different smoke detection algorithms' indicators. In Table 2, the TP stands for true positive rate, the FP for false-positive rate, the FN for false-negative rate, TN for true negative rate shown in Table 2:

4.1.4 *Experimental Environment*
The experimental environment of this paper is the Ubuntu16.04 system, Python 2.7, Caffe. In the hardware device section, the graphics card is NVIDIA GeForce GTX1060Ti, and the CPU is Intel Core i5-8400. The emulator is written in Python and uses the neural network library Caffe and GPU.

4.2 Model Training

The following explains the training method of fusion deep network. Because of the limited number of data sets, it is difficult to train the fusion network model through these images. This paper adopts a transfer learning method to solve the problem of small sample training. Figure 14 shows the accuracy change curve of the fusion depth network and VGG16 network during the training. The training-related hyperparameters are shown in Table 3. Figure 14 shows that the AP value of the fusion depth network model is superior to that of a single network. AP values also tend to stabilize at 60,000 iterations. The performance of the fusion network proposed in this paper is better than the VGG16 network.

4.3 Performance Comparison

4.3.1 *Performance Comparison of Visible Light Smoke Detection Network Mode*
To verify the beneficial effects of the network, we compared smoke detection performance based on the different network model.

Table 1
Details of Test Video

Video Serial Number	Type	Frames	Description of the Video
1	Smoke video	776	Fast moving thick smoke
2	Smoke video	554	Fast moving thin smoke
3	Smoke video	329	Smoke with white background
4	Smokeless video	906	Fast moving white clouds
5	Smokeless video	551	Video with heat source
6	Smokeless video	682	Video with fog

Table 2
Evaluation Criteria

Name	Calculation Formula
Accuracy(ACC)	ACC=(TP+TN)/(TP+FP+TN+FN)
Precision	P =TP/(TP +FP)
Recal	R =TP/(TP +FN)
Average accuracy	AP = $\int_0^1 P(R)d_R$

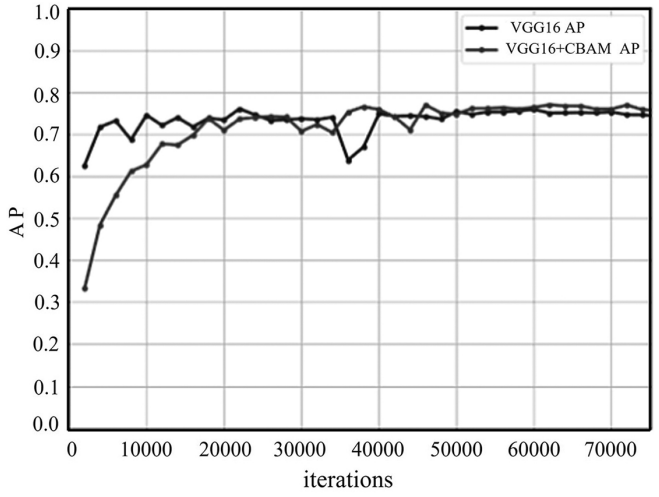


Figure 14. Change curve of network AP value.

Table 3
Training Hyperparameters

Network Model	Base_lr	Gamma	Iter_Size
VGG16	0.001	0.1	2
VGG16_ICBAM	0.0001	0.1	2

In our work, the confidence thresh is set to 0.3, and IOU value is 0.5. We compare the recall, accuracy, and AP of the VGG16, VGG16_ICBAM, and VGG16_ICBAM_INMS networks on the test set. It can be seen from Table 4 that our method is higher than VGG16 and VGG16_ICBAM.

Table 4
Comparison of Different Network Performances

Network Model	Recall	Precision	AP
VGG16	82.03%	74.32%	76.06%
VGG16_ICBAM	85.94%	39.93%	76.11%
VGG16_ICBAM_INMS	82.05%	65.20%	77.64%

Table 5
Comparison of Different Methods

Method	Our Method	SSD	YOLOV3
AP	77.64%	77.29%	76.09%

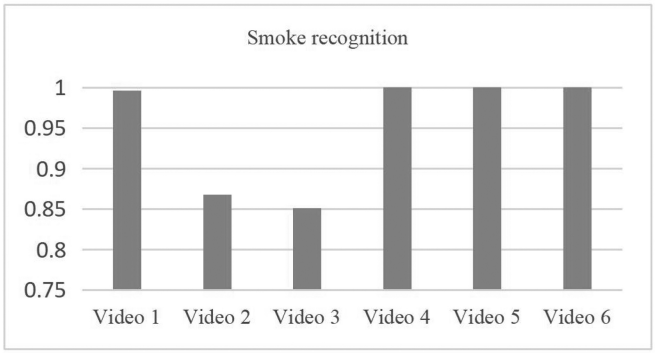
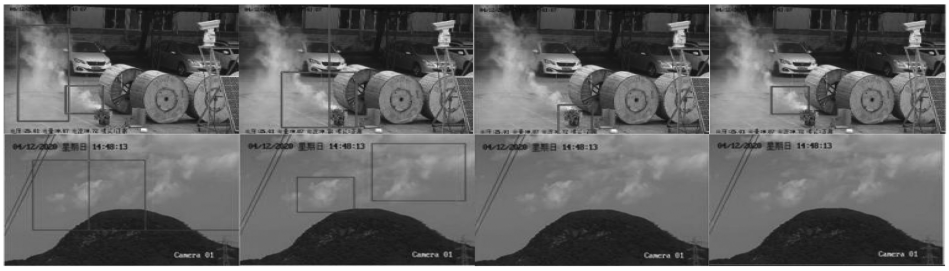


Figure 15. Accuracy rate of test video (unit %).

We compare the method in this paper with AP values of detection methods such as YOLO and SSD. As shown in Table 5, it can be seen our method is better than the other object detection.

4.3.2 *Performance Test of Smoke Detection System*
To verify the detection effect in real scenarios, we use the video for testing. During the testing process, the non-smoke video frames are removed from the video in order to simplify the calculation. Further, this paper uses accuracy as the standard for smoke detection. The accuracy of video detection is shown in Fig. 15.

In Fig. 16, seeing that the moving white clouds have been misidentified, and only the smoke images containing



(a) (b) (c) (d)

Figure 16. the column (a) is the result of visible light detection; the column (b) is the result of box detection after visible light expansion; the column (c) is the detection result of infrared thermal imaging; the column (d) is the combination of infrared thermal imaging and visible light.

heat sources are identified. The algorithm in this paper achieves the goal of reducing ignorance.

5. Conclusion

In recent years, using computer vision to detect smoke in video is the mainstream research direction in fire detection. Considering the characteristics of smoke objects with variable shape and irregular motion, traditional methods have poor performance, especially high false positive rate. Deep learning-based methods have improved smoke detection performance to some extent. However, just based on depth features from the feature extraction leads to difficulties in effectively distinguishing the objects which are similar to smoke, such as clouds and fogs. Meanwhile, the traditional method based on single visible image only considers static and dynamic features, but does not consider the heat source features of smoke, while infrared images can characterise the heat source features well. Therefore, combining visible and infrared images, this paper proposes a dual-spectral video smoke detection framework based on the improved Faster R-CNN. Firstly, we propose the fusion network of the VGG16 network and ICBAM and the improved NMS algorithm.; secondly, differing from smoke detection methods that rely only on static features, motion features and features of smoke heat source information are added as a way to improve the smoke detection accuracy. The experimental results prove the effectiveness of our method.

In the subsequent work, we will investigate pure deep learning-based smoke video detection methods, build new spatio-temporal Transformer models to explore long-range dependencies between pixels on space, and, at the same time, model spatio-temporal relations to improve detection performance.

Acknowledgement

This work was supported in part by the Sichuan Provincial Science and Technology Plan Project (Grant No. 2022NZZJ0036, 2021YFQ0070), in part by the Spring Plan of Ministry of Education of China (No. Z2017083), in part by the Open Research Fund of Key Laboratory of Integration and application of solar energy technology (Grant No. 2017-TYN-Y-02), in part by the Key scientific research fund of Xihua University (Grant Nos. Z1420210 and Z1620211), in part by the Open Research Fund of Health Management Development Center (Xihua University, No. s2jj2017-023, s2jj2017-039).

References

[1] J. Shi, F. Yuan, and X. Xia, Video smoke detection: a literature survey, *Journal of Image and Graphics*, *23*(3), 2018.

[2] C. Emmy Prema, S. Vinsley, and S. Suresh, Multi feature analysis of smoke in YUV color space for early forest fire detection, *Fire Technology*, *52*(5), 2016, 1319-1342.

[3] T. Chen, Y. Yin, S. Huang, and Y. Ye, The smoke detection for early fire-alarming system base on video processing, *International Conference on Intelligent Information Hiding and Multimedia, (*Pasadena, CA, USA, 2006), 427-430.

[4] L. Dong, and J. YU, Smoke detection method in video based on image separation, *Computer engineering*, *41*(9), 2015, 251-254.

[5] K. Nikhil, Convolutional Neural Networks*, Deep Learning with Python: A Hands-on Introduction*, 2017, 63-78.

[6] Z. Yin, B. Wan, F. Yuan, X. Xia, and J. Shi, A deep normalization and convolutional neural network for image smoke detection, *IEEE Access*, *5*(5), 2017, 18429-18438.

[7] Y. Hu, and X. Lu, Real-time video fire smoke detection by utilizing spatial-temporal ConvNet features, *Multimedia Tools and Applications*, *77*(22), 2018, 29283-29301.

[8] G. Xu, Y. Zhang, Q. Zhang, G. Lin, Z. Wang, Y. Jia, J. Wang, Video smoke detection based on deep saliency network, *Fire Safety Journal*, *105*(105), 2019, 277-285.

[9] F. Yuan, L. Zhang, X. Xia, B. Wan, Q. Huang, and X. Li, Deep smoke segmentation, *Neurocomputing*, *357*(357), 2019,248-260.

[10] Q. Zhang, G. Lin, Y. Zhang, G. Xu, and J. Wang, Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images, *Procedia Engineering*, *211*(211), 2018, 441-446.

[11] G. Lin, Y. Zhang, G. Xu, and Q. Zhang, Smoke detection on video sequences using 3D convolutional neural networks, *Fire Technology*, *55*(5), 2019, 1827-1847.

[12] C.M. Li, X.Y. Qu, Y. Yang, H.M. Gao, Y.C. Wang, *et al.*, High-Resolution remote sensing image segmentation method based on SReLU. *International Journal of Robotics and Automation*, *34*(3), 2019, 225-234.

[13] C.M. Li, H.M. Gao, Y. Yang, X.Y. Qu, and W.J. Yuan. Segmentation method of High-Resolution remote sensing image for fast object detection, *International Journal of Robotics and Automation*, *34*(3), 2019, 216-224.

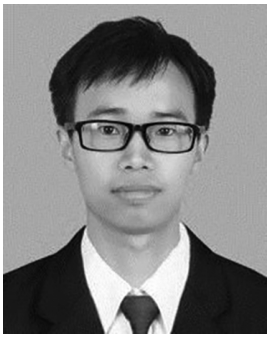
[14] P. Liu, C.L. Song, J.M. Li, X.Y. Simon, C. Xingyu, L. Chuanfu, F. Qiang, Detection of transmission line against external force damage based on improved yolov3, *International Journal of Robotics and Automation*, *35*(6), 2020, 460-468.

[15] T.D. Dung, G. Capi. Application of neural networks for robot 3D mapping and annotation using depth image camera, *International Journal of Robotics and Automation*, *37*(6), 2022, 529-536.

[16] S. Woo, J. Park, and J.Y. Lee, CBAM: Convolutional Block Attention Module, *european conference on computer vision*, 2018, 3-19.
[17] J. Hu, L. Shen, and S. Albanie, Squeeze-and-Excitation networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(8), 2020, 2011-2023.
[18] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, EAST: An Efficient and Accurate Scene Text Detector, *computer vision and pattern detection*, 2017, 2642-2651.
[19] R.C. Gonzalez, and R.E. Woods, *Digital image processing*, 3th ed, (Chain: Ruan Qiuqi, Electronic Industry Press, 2011), 66-435.
[20] X. Qin, C. Yuan, Y. Deng, Y. Shi, and J. Yuan, An improved Ostu image segmentation method, *Journal of Shanxi University*, *36*(4), 2013, 530-534.
[21] http://signal.ee.bilkent.edu.tr/VisiFire/Demo/.
[22] https://cvpr.kmu.ac.kr/.

Biographies

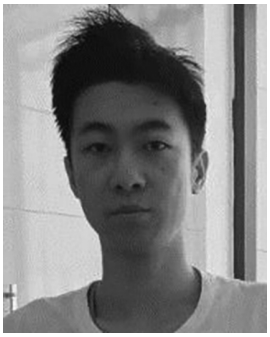
Changlin Song, born in 1973, is currently a lecturer at Xihua University, Chengdu, P.R. China. He received his Ph.D. degree from Southwest Jiaotong University, China, in 2004. His research interests include electric drive, signal detection and processing, and computer and communication technology.



Chuanfu Liu born in 1992. He received his master degree at Xihua University, Chengdu, P.R. China. His research interests include robotics and machine vision.

Tingping Feng is currently pursuing his master degree in Mechanical Control Engineering at Xihua University, Chengdu, P.R. China. His research interests lay in robot path planning and bio-inspired algorithms.

Junmin Li, was born in 1975. He received the Ph.D. degree from Sichuan University, China, in 2014. He is currently a Professor with Xihua University, Chengdu, P.R. China, since 2009. His research interests include robotics and machine vision.



Yanlin Pan received his bachelor's degree at Xihua University, Chengdu, P.R. His research interests include electronic sensors.

Simon X. Yang is currently a professor at the University of Guelph, Canada. He received his Ph.D. degree from the University of Alberta, Canada. His research interests include intelligent systems, robotics, sensors, communications and networking, control systems, and vision and signal processing.