Proceedings of the 37th Chinese Control Conference
July 25-27, 2018, Wuhan, China

A UAV-based Forest Fire Detection Algorithm Using Convolutional Neural Network

Yanhong Chen[1], Youmin Zhang[2]*, Jing Xin[1], Yingmin Yi[1], Ding Liu[1], Han Liu[1]

1. Xi'an University of Technology, Xi'an 710048, P. R. China
E-mail: 15891423152@sina.cn

2. Concordia University, Montreal, H3G 1M8, Canada
E-mail: youmin.zhang@concordia.ca

Abstract: With the new development in Unmanned Aerial Vehicle (UAV) in recent years, UAV equipped with color and infrared cameras becomes a new tool for carrying out the forest fire detection and fighting missions due to its advantages of low price, high maneuverability, and easy use. In order to detect a potential fire in its early stage, a UAV-based forest fire detection method using a convolutional neural network method is proposed in this paper. The effectiveness of the proposed fire detection algorithm is verified by using simulated flames in an indoor experimental testbed.

Key Words: Forest Fire Detection, Convolutional Neural Network, Unmanned Aerial Vehicle

1 Introduction

As precious natural resources, an important asset to human, and for maintaining good ecological balance, it is an arduous and pressing task to protect forest resources from being destroyed due to fires and artificial damages. Forest fires are dangerous disaster to human life, asset and environment and early detection of fires is very important to minimize the damages and loss.

Unmanned Aerial Vehicle (UAV) is a new type of aviation platform. In recent years, as its technology continues to mature, it has been applied in many fields, such as meteorological sounding, disaster monitoring, power line patrol and post-disaster rescue and so on. Especially the light weight and small size mini UAVs are characterized by low cost, easy operation and flexible maneuvering. They can adjust the work plan and onboard remote sensing equipment according to the real-time situations in the field, and are very suitable for the detection of forest fires. At present, forest fire research based on computer vision is paying more and more attention to combine with UAVs for its high flexiblility in carrying out forest fire detection and fighting mission [2]. Compared with infrared cameras, multispectral sensor, hyperspectral sensor and other heat-sensitive sensors, CCD camera has the advantages of low cost, easy operation and widely used in UAV-based forest fire detection.

There are many methods of forest fire detection, due to its multi-level network structure and thousands of network nodes, neural network can well approximate the complex functions and then characterize the distributed features of input data, providing a new method for forest fire detection. In recent years, Convolutional Neural Network (CNN), as a powerful automatic feature acquisition method, has been widely used in speech recognition, natural language processing, object detection and other fields. Especially for multi-dimensional input vector images, the model can be directly input to avoid the complexity of data recovery in features extraction and classification. Therefore, CNN has been adopted in this paper for learning the fire features.

This work was supported in part by National Natural Science Foundation of China under Grant 61573282, Natural Science Foundation of Shaanxi Province under Grants 2015JZ020, 2016JM6006, and 16JS069, Xi'an Science and Technology under Grant 2017080CG/RC043(XALG005), and Natural Sciences and Engineering Research Council of Canada.

2 Related Works

Forest fire detection methods are various and can be divided into traditional methods based on human monitoring and new fire detection systems based on computer vision. Among them, the traditional detection methods can be divided into four categories according to the order of development: surface patrolling, observatory detection, satellite remote sensing and aircraft patrolling. However, the video surveillance system based on computer vision is with the best prospect at present, and therefore this paper conducts a comprehensive study along this direction.

2.1 Traditional Fire Detection Methods

(1) In the early stage of forest fire detection, the main method used in China is manual patrolling. Although there are many shortcomings in this forest fire detection methods, such as small patrolling area, large error, insufficient coverage area, large cost burden, etc., many of the forest areas in China are still in the use of such a way because of its simple operation and low cost.

(2) Subsequent forest fire detection is observatory inspection. The inadequacies of this kind of forest fire detection method are also obvious: firstly, observation stands can not be established in the remote forest areas without living conditions; secondly, the observation effects of such forest fire detection are easily limited by topography.

(3) Forest fire detection based on satellite remote sensing then emerged. Its advantage is that it can detect a wider range of data faster. Nonetheless, there are also some disadvantages and shortcomings. The satellite platform itself has a fixed operation period, so the frequency of inspection is limited. If the optimal time is missed, the danger caused by the fire can not be reduced. The ground resolution is relatively low and needs to be verified on the ground, which spends a lot of manpower, material and financial resources.

(4) As a result of continuous improvement of science and technology, forest fire detection using aviation patrolling

used more and more due to its the advantages of its fast reaction speed, high operating efficiency, covering large patrolling area quickly. As one of the most effective forest fire monitoring methods , such a way of forest fire monitoring and fighting is already in foreign countries a major force in forest fire prevention. However, because of funding and other reasons, China's aviation forest protection is facing the problems of obsolete aircraft, aging equipment and insufficient flight capacity. The lack of an efficient, economical and easy-to-maintain flight platform has become one of the bottlenecks restricting the development of aviation forest protection.

2.2 Video Surveillance System Based on Computer Vision

Due to the rapid development of digital camera technology and video processing, video surveillance systems based on computer vision have replaced the traditional fire detection methods. The main driving force behind the study of Video Fire Detection (VFD) method is the popularization of video surveillance system and the maturity of machine vision technology. However, due to practical and commercial interests, there are few thematic literature on the related algorithms. All of the existing methods are based on the analysis and modeling of flame characteristics. Contextual support for flames, such as color model, region model, motion model, and time-frequency state model, are required for all levels of processing, from low level to high level.

(1) Pixel-based VFD method. Early VFD method is mainly based on the color and brightness of the flame, color image processing method based on flame color can significantly inhibit false detection caused by changes in brightness conditions (such as background lighting). Yamagishi et al. [7] used the flame color model of HSV space to weaken the influence of ambient light, wind motion, flame size and detection distance. The early forest fire detection system based on computer vision proposed by Li et al. [11]. The image processing technology carries on the image processing and the analysis unceasingly, according to the early forest fire image characteristic criterion finally determine if a fire has occurred.

(2) The VFD method based on the area of flame color movement. Hideaki et al. [7] developed a flame detection method to eliminate the influence of artificial light, wind and distance by calculating the spatiotemporal fluctuation data of the flame area extracted from the color information. Song et al. [14] considered that the flame is constantly in the image. Therefore, the study of their motion characteristics and edge characteristics, respectively, and the fusion of the two, to achieve a fast and accurate flame detection of different image sequences.

It can be seen that the artificial forest fire detection, such as ground patrolling and observation platform detection, is relatively dependent on the surrounding environment and lower in the accuracy and false positive rate. The use of aviation patrols and satellite remote sensing and other methods to achieve forest fire detection, depends on the equipment performance too much, and the cost is too high. Although the development of computer vision theory is not mature enough, computer video monitoring system not only conforms to the future directions of information industry, but also represents the future development trends of monitoring industry. It is highly valued to the research and development to academia, industry and management.

3 Using CNN for Forest Fire Detection

The traditional method of fire detection and the method used in this paper is shown in Fig. 1.

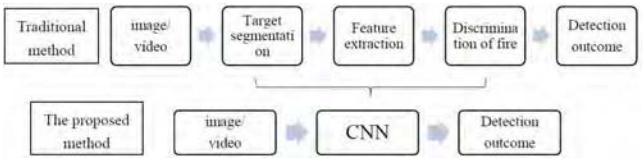

Fig. 1: Comparison of fire detection methods

In most traditional methods, in order to detect fires in image or video data, the target segmentation is firstly performed, typically separating the foreground objects (such as smoke and flames) from the background and then extracting the features of the object (in the detection of fire, color and motion features are usually extracted for flame detection, texture and wavelet features are usually for smoke detection), and then the features are classified by a series of classification methods, such as Support Vector Machine (SVM) algorithm, to obtain test results. It is feasible to detect fire by this method. However, after this series of processing, not only the detection speed will be limited, but also if the feature is improperly selected or the key features are not extracted, the detection efficiency will be greatly descend leave a space in between.Therefore, selecting and extracting the image features with better generalization ability can further improve the performance of the fire detection. So, we use CNN convolution network instead of traditional features extraction algorithm such as Local Binary Patterns(LBP), Scale-Invariant Feature Transform(SIFT) and so on to extract fire image features. Moreover, features learned by the CNN has better generalization ability.

The flow of the proposed fire detection method is shown in Fig. 2. The video images are usually noisy due to the influence of the detection environment and the uneven illumination during the acquisition. Before the detection, the preprocessing needs to be performed. Preprocessing includes image enhancement using histogram equalization, and image filtering using a non-linear median filter. The whole process of fire detection system can be divided into three steps. Firstly, according to the characteristic of fire, a CNN model is constructed for fire detection, including network structure, parameter setting and so on. Then, preprocessed images and its label information are adopted as samples data to train the constructed CNN. Finally, fire detection is performed using the well-trained CNN network in the current test data, the smoke detection is first performed, and if there is smoke in the test data, the alarm is directly triggered. If no smoke is detected, then the flame detection is performed, and once upon the detection of flame, the alarm is immediately triggered. If no smoke and no flame are detected, enter the next set of data for a new round of testing.

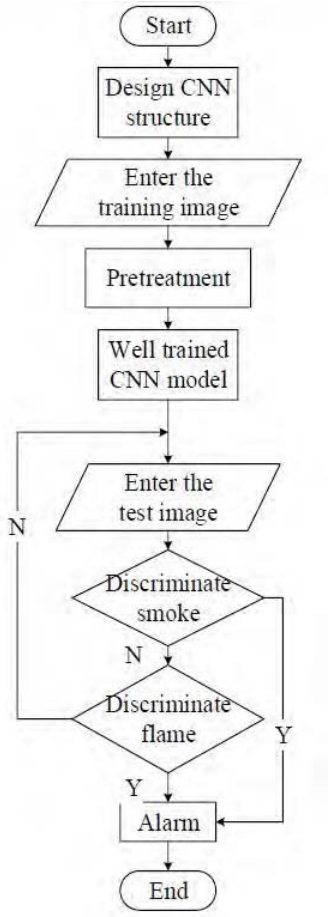

Fig. 2: System flow chart

3.1 Characteristics of CNN

Convolution neural network is a special deep neural network model. It combines the artificial neural network and deep learning technoledge. The convolutional neural network reduces the number of parameters by using three core ideas of local experience domain, weight sharing and pooling to improve the training performance, and can guarantee the robustness of the image to displacement, scaling and deformation.

The first major feature of CNN is the local perception filed. It is generally believed that human perception of the outside world is from the local to the global, and the spatial connection of the image is also the local pixel connection is more closely, while the pixel distance farther away is weaker. Therefore, each neuron does not actually need to perceive the global image, only needs to perceive the local, and then get the global information by integrating the local information at a higher level. The idea of network connectivity is also inspired by the visual system structure in biology. Neurons in the visual cortex receive information locally (i.e., these neurons only respond to stimuli in certain areas). Fig. 3 is a fully connected and partially connected schematic.

In the left diagram, each neuron is fully connected to all the pixels of the entire image, so each neuron corresponds to 1,000,000 parameters (connection weights), while in the right image, each neuron has 10 * 10 pixels connected locally to original image, only corresponding to 100 parameters, we can find that changing the way of connection be-

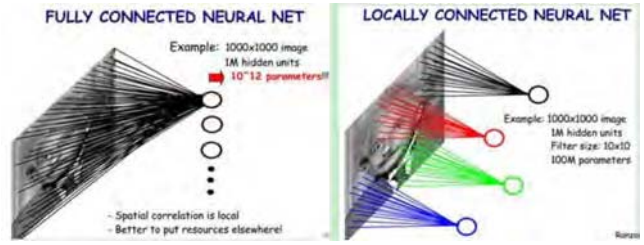

Fig. 3: Schematic diagram of full connections and local connections. The left is fully connected, the right is partially connected.

tween neurons in various layers, the number of parameters can be greatly reduced. But in this case, the parameters are still too much, then the second biggest advantage of CNN - weight sharing can also further reduce the parameters. The weight-sharing process can regard these 100 parameters (i.e., the convolution operation) as a way of extracting features that are position-independent. The implicit principle is that part of the image has the same statistical characteristics as the rest of the image. This also means that the features we learn in this section can also be used in another part, so we can use the same learning features for all the places in the image.

A natural idea for describing large images is to aggregate the features of different locations, for example, by calculating the average (or maximum) of a particular feature over a region of the image. These summary statistics not only have much lower dimensions (compared to using all the extracted features), but also improve the results (not easily over-fitted). This kind of aggregation operation is called pooling. According to the method of computing pooling, there are common mean pooling and maximum pooling, as shown in Fig. 4.

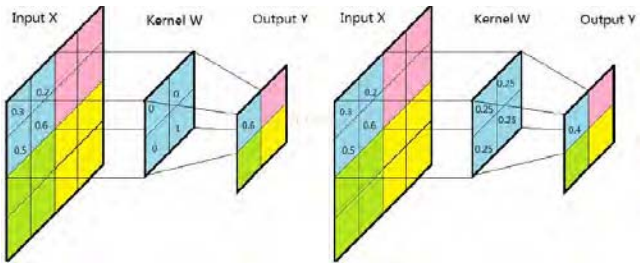

Fig. 4: Pooling process diagram. The left is max pooling, the right is mean pooling.

3.2 Structure of CNN

Convolution neural network is a multi-layer neural network, each layer includes a plurality of two-dimensional feature plan, each feature map includes a plurality of independent neurons. In most of the current research works, CNN is applied to many machine learning problems including face recognition, document analysis and language detection. Take the structure of the developed CNN fire detection experiment as an example, the structure and parameters are shown in Fig. 5.

As shown in the figure, this is a nine-layer convolutional neural network, the input image size is 64 * 64. After 20

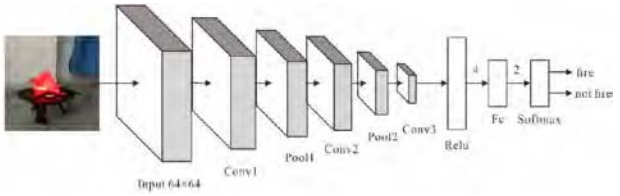

Fig. 5: Proposed CNN model of fire detection experiment.

convolutions with filters of 5 * 5 size, 20 feature maps are generated in the C1 layer, and then four pixels of each group in the feature map are summed, weighted, biased, then 20 feature maps are obtained in the S2 layer (in this system, each neuron of the feature map and the convolution layer of a 2 * 2 size of the domain connection, and using the max pooling). These maps pass through 50 filters of size 10 * 10 to get the C3 layer. This hierarchy also produces S4 as S2. Then through 500 filters of size 9 * 9 will get C5 layer. Finally, these pixel values are passed through Relu, rasterized and connected as a vector into the softmax classifier to get the output.

Convolutional-layer are typically feature extraction layers, also known as convolution layers. The input of each neuron is connected to the local receptivity of the previous layer and the local feature is extracted. Once the local feature is extracted, its position relationship with other features are also determined. The Pool-layer is characterized mapping layer, also known as the pooling layer. Each computing layer of the network consists of a plurality of feature maps, each of which maps to a plane whose weights on all planes are equal. The feature mapping structure uses the sigmoid function with small influence kernel as the activation function of the convolutional network, which makes the feature map invariant to displacement. In this system, each neuron of the feature map is connected to a 2 * 2 size area of the convolutional layer and uses the max pooling. If there are n feature maps, then the number of down-sampled feature maps is still n, but the output feature maps will be smaller.

Relu layer: Rectified Linear Units, combined with biological neurons, using any non-linear activation function, can make single-layer perceptrons have the ability to solve linearly inseparable. F7 layer: equivalent to the Relu layer expanded into a vector, fully connect with the Relu layer, contains 2,000 neurons. Output layer: Because we want to identify whether the image has a fire, it contains two neurons, F7 layer full connection.

3.3 Training of CNN

The purpose of the training is to get the weight of connection of each layer. It can be divided into three aspects: forward propagation, back propagation and weight updating.

(A) Forward Propagation Process

The idea of neural network forward propagation is relatively simple, as shown in Fig. 6, assuming that the previous node $i, j, k, ...$ are connected with the nodes $j, k, l, ...$ of this layer, then the value of this layer node are through the previous nodes and the corresponding connection weights w_{ij},w_{jk},w_{kl},... are weighted and summed. The final result is added with an offset item (the offset is omitted in the Fig. 6), and finally through a non-linear function, that is, activation functions, such as Rectified Linear Unit (ReLu), sigmoid and so on, the final result is the output of this layer of nodes. Through this method of continuous layers of operation, the output layer results will be eventually obtained.

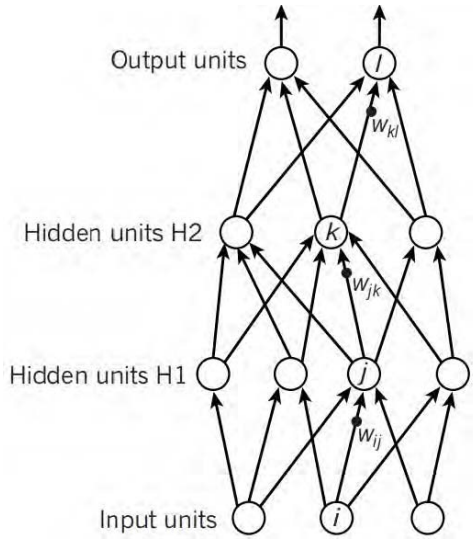

Fig. 6: Forward propagation process of convolutional neural network.

In Fig. 6, the i-th neuron of the input layer is represented as x_i and the neuron of the hidden layer H1 is denoted by j. Then, w_{ij} represents the connection weight from i to j. At this time, the output of neuron j can be represented by y_j.

$$\begin{cases} z_j = \sum_{i \in Input} w_{ij} x_i \\ y_i = f(z_j) \end{cases} \tag{1}$$

Similarly, we can calculate the output of H2 layer y_k and network output y_l:

$$\begin{cases} z_k = \sum_{j \in H1} w_{jk} y_j \\ y_k = f(z_k) \end{cases} \tag{2}$$

$$\begin{cases} z_l = \sum_{j \in H2} w_{kl} y_k \\ y_l = f(z_l) \end{cases} \tag{3}$$

(B) Back Propagation Process

(1) Finding the output layer error-sensitive items

The error sensitivity value of the last layer is equal to the output value f_x of the CNN network minus the sample tag value e_y, ie, f_x-e_y.

(2) When the next layer in the convolution layer is the pooling layer, find the error sensitivity term of the convolution layer.

Assuming that the l-th layer is a convolution layer, the $(l+1)$-th layer is the pooling layer, and the error-sensitive item of the pooling layer is ${\delta_j}^{l+1}$, and the error sensitivity term of the convolution layer is δ_j^l. The relationship between the two expressions are as follows:

$$\delta_j^l = upsample\left(\delta_j^{l+1}\right) \cdot \dot{h}\left(a_j^l\right) \tag{4}$$

where $\cdot$ denotes the dot product operation of the matrix (i.e., the product of the corresponding elements) and upsample () is the upsampling process. The general idea is that each

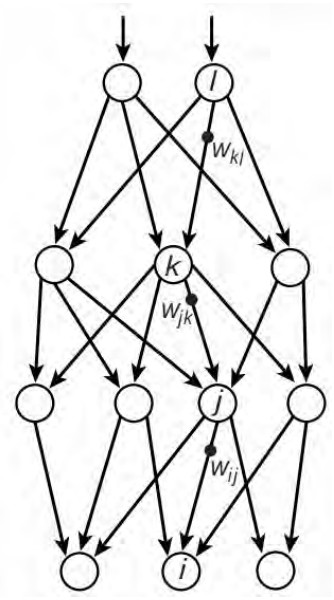

Fig. 7: Back propagation of convolutional neural network.

node in the pooling layer is composed of multiple nodes in the convolution layer Region. Therefore, the error-sensitive value of each node in the pooling layer is also jointly generated by the error-sensitive values of multiple nodes in the convolution layer, so long as the error-sensitive sum of all layers in the back-propagation is satisfied. The last notation $\dot{h}(a_j^l)$ denotes the derivative of the activation function at the j-th node of layer l-th (the input to the node).

(3) When the next layer of the pooling layer is convolution layer, find the error-sensitive item of the pooling layer.

Assuming that there are N channel maps in the $l-$th layer (Pooling layer), that is, there are N feature maps, the $(l+1)$-th layer is a convolution layer, and there are M features, and each channel in the l-th layer has its own error-sensitive value, which is calculated as the sum of the contributions of all the characteristic kernels of the $(l+1)$-th layer. Then the error sensitivity of the i-th feature in the $(l+1)$-th layer and the i-th channel in the l-th layer can be calculated as:

$$\delta_j^l = \sum\nolimits_{j=1}^{M} \delta_j^{l+1} * k_{ij} \tag{5}$$

Among them, * indicates the discrete convolution operation of the matrix, which is equivalent to rotating the convolution kernel and then performing related operations (up-down flip and left-right flip). In addition, the default pooling layer is a linear activation function, so there is no derivative of the corresponding node.

(C) Weight Update Process

Error sensitivity of the output layer was obtained from the front one, the error sensitivity of the convolution layer was deduced from the pooling layer, and the error sensitivity of the pooling layer was deduced from the convolution layer. Here, we mainly derive the weights in front of the convolution layer. Assuming now that we need the derivative of the weights and offsets between the i-th channel at the l-th layer and the j-th channel at the $(l+1)$-th layer, the formula is:

$$\begin{cases} \dfrac{\partial Loss}{\partial b_j} = \sum_{U,V} \left(\delta_j^{l+1}\right)_{U,V} \\ \dfrac{\partial Loss}{\partial k_{ij}} = x_i^l \odot \delta_j^{l+1} \end{cases} \tag{6}$$

where $\odot$ is the matrix-related operations, which can be implemented by the conv2() function, in the use of the function, the $(l+1)$-th layer of the j-th error-sensitive value to be reversed. $Loss$ is the Network loss function, which can be calculated by Eq. (7).

$$Loss = -\frac{1}{m}\sum_{i=1}^{m} y_l \lg f\left(x_i\right) + \lambda \sum_{k=1}^{L} \|w_k\|^2 \tag{7}$$

where L is the total layers of the CNN network, m is the number of samples (x_i,y_i), x_i is the input of the network, and $f(x_i)$ is the output of the network.

4 Experiment Results

At present, there are no common open fire image dataset. To verify the validity of the proposed fire detection method in this paper, a six-rotor drone (DJI S900) equipped Sony A7 camera is used to acquire forest fire images, as shown in Fig. 8. Due to the limited conditions, the flame image dataset here were generated using a flame simulator and corresponding label data is generated by manual approach. We conducted a CNN-based flame test on this dataset.

Fig. 8: Forest fire acquisition system.

Table 1: Experimental Data Set

Types	Train set	Test set	Verify set	Total
Not fire, front	75	25	5	105
Not fire, top	125	25	5	155
Fire, front	150	50	10	210
Fire, top	302	148	30	480
Total	652	248	50	950

There are positive samples and negative samples (be divided into four types of data) used in the experiment is shown in Table 1, which consists of 950 images. The first type is negative samples without fire, which are respectively from the front and the top image acquisition; the second type is a positive sample containing fires, also from the front and

the top of the image acquisition. All 950 images are divided into a training set, a test set, and a validation set, each containing a positive and a negative sample of varying numbers. After 5 iterations, a well-trained CNN model can be obtained. Related network parameters setting is shown in Fig. 9.

```
% Meta parameters
net.meta.inputSize = [64 64] ;
net.meta.trainOpts.learningRate = 0.0005 ;
net.meta.trainOpts.numEpochs = 5 ;
net.meta.trainOpts.batchSize = 30 ;
```

Fig. 9: CNN parameter settings.

We conduct several fire detection experiments on the test data. Fig. 9 shows two of the test results. It can be seen from Fig. 10 that proposed method is an effective fire detection method.

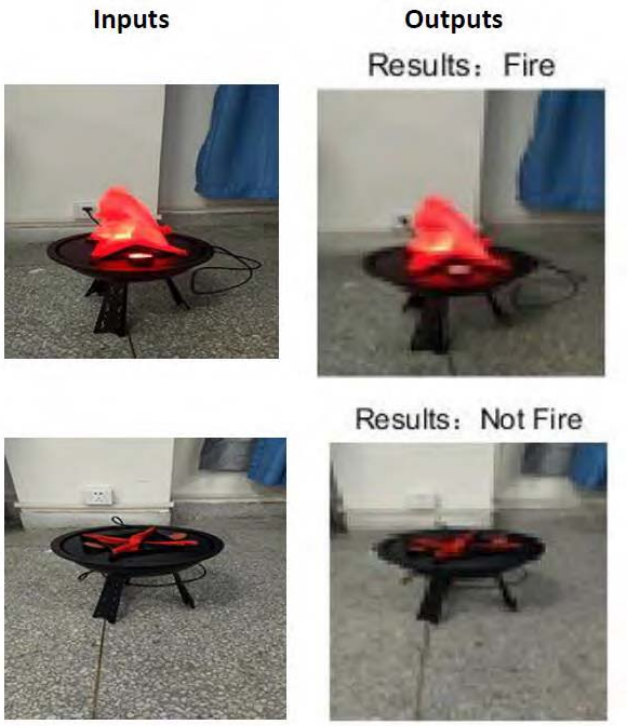

Fig. 10: Flame test results.

5 Conclusion

This paper presents a detection method for forest fire based on convolutional neural network (CNN). CNN reduces the number of parameters through three key ideas of local experience domain, weight sharing and pooling to improve training performance and ensure the robustness of the image to displacement, scaling and deformation. Especially for multidimensional input vector images, images can be imported directly into the model, which avoids the complexity of data reconstruction in feature extraction and classification. Compared with other detection methods, the algorithm complexity is reduced and the accuracy of detection is also improved. Meanwhile, the smoke appears before the flame, so smoke detection can fight for fire prevention and fighting for more time. If, for some reason, no smoke is detected, the high accuracy of flame detection reduces the risk of fire. Finally, the validity of the proposed method is verified by the fire detection experiments on the simulated flame.

References

[1] A. Ollero, J. R. Martnez-de-Dios, L. Merino. Unmanned aerial vehicles as tools for forest-fire fighting. in *International Conference on Forest Fire Research*, 2006.

[2] F. Yizhou, M. Hongbing. Video-based forest fire smoke recognition, *Journal of Tsinghua University*. 2015, 55(2): 243-250, 256.

[3] F. Tianju. Forest Fire Image Recognition Algorithm Based on Depth Learning and Its Implementation, *Beijing Forestry University*. 2016

[4] H. Yamagishi, J. Yamaguchi. Fire flame detection algorithm using a color camera, in *Proceedings of 1999 International Symposium on Micro Electro Mechanical System and Human Science*, 1999, 255-260.

[5] J.R.Martinez-de Dios, B.C.Arrue, A.Ollero, L.Merino, F. Gomez-Rodrguez. Computer vision techniques for forest fire perception, in *Image and Vision Computing*, 26(2008), 550-562.

[6] A. Krizhevsky, I. Sutskever, G. E. Hinton. Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, 2012: 1097-1105.

[7] M. Bisquert, E. Caselles, J. M. Snchez, V. Caselles. Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data, in *International Journal of Wildland Fire*, 2012, 21(8): 1025-1029.

[8] S. A. M. Saleh, S. A. Suandin, H. Ibrahim. Recent survey on crowd density estimation and counting for visual surveillance, *Engineering Applications of Artificial Intelligence*, 2015, 41: 103-114.

[9] Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes, in *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014: 1891-1898.

[10] H. Cruz, M. Eckert, J. Meneses and J. F. Martnez. Efficient forest fire detection index for application in unmanned aerial systems (UASs), *Sensors*, 2016, 16(6): 893.