Joint Network Smoke Recognition based on Channel Attention Mechanism

Shanju Jin¹, Tongzhou Zhao², Xiaoyun An³

College of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, Hubei, 430205, China
*21807010013@stu.wit.edu.cn

Abstract. Aiming at the problems that traditional fire smoke recognition methods in a low recognition accuracy, a fusion network based on VGG16 is proposed, which use channel attention mechanism and contain Dense Blocks network to extract smoke features. To avoid the loss of smoke features, channel attention mechanism in backbone network is automatically to learn the importance of feature in this network. The experiment results show that the accuracy of this network is 3.0% higher than VGG16 neural network, and which is effective and feasible in smoke recognition tasks.

1. Introduction

Smoke recognition has become an important technology of early fire warning. Traditional smoke recognition methods usually use the color feature, edge feature, shape feature, motion and dynamic texture features of smoke to recognize smoke.

Lin[1] et al proposed a method that use the motion characteristics of smoke to extract the foreground, extract the optical flow and edge features of smoke in the moving area to recognize the smoke. Zhao[2] et al obtained the floatability by analyzing the central motion characteristics, and used the local binary motion model to obtain dynamic texture to recognize smoke. Zhou[3] et al proposed a method to use the local extreme value region segmentation method to recognize the fire smoke in a long distance, but this method in a high false alarm rate, especially the thick smoke. Li[4] et al proposed a more discriminative feature operator by using the texture and edge features of smoke. The above smoke recognition methods are based on the color, texture, shape and other features of the smoke image. However, these features are easily affected by the complex external environment, and the smoke scene is often fuzzy and the background is changeable which makes the feature extraction particularly difficult. The emergence of deep convolution neural network solves this problem well.

Compared with the traditional methods, the convolution neural network has a significantly improvement in the accuracy of image recognition. On the basis of smoke texture features, Chen[5] et al obtained a kind of based on the cascaded convolutional neural network by fusing the static and dynamic texture information of the smoke. The smoke texture recognition framework improves the accuracy of smoke recognition, but processing the two types of texture information separately will easily increase the complexity of the algorithm Although the above models improve the accuracy rate to a certain extent, most of these networks have many problems, such as large number of parameters, slow recognition rate and high false alarm rate, which are difficult to meet the actual needs. Xie[6] et al constructed a spatiotemporal dual path 3D residual convolution neural network by weighted fusion of attention mechanism and temporal and spatial characteristics of smoke, and realized an end-to-end...
smoke recognition algorithm. However, the fusion of the two mechanisms leads to the increase of network layers and the long training time. According to the above problems, a smoke recognition network is proposed in this paper, which based on VGG16[7] network by introducing Dense Block[8] model and adding channel attention mechanism in backbone network, which significantly improves the recognition accuracy and saves time.

2. Related work

2.1 Convolution Neural Network
There are many works on smoke recognition using deep learning. Xu[9] constructed a depth domain adaptive network which trained on the dataset of synthetic smoke and real smoke. The result show that the smoke feature space of the domain invariant domain is extended and the error detection rate is reduced, but the smoke detection performance of the training model in the real scene will be affected by the dataset of synthetic smoke. Wang[10] fused VGG16 with ResNet50 and improved the accuracy of smoke recognition in small sample data, but the complex of the model led to longer training time. Therefore, a joint network smoke recognition based on channel attention mechanism has high smoke recognition rate and real-time performance.

2.2. VGG16 network
In this paper, a Dense Block network is added to the VGG16 network, and the channel attention mechanism is introduced to improve the network recognition rate and reduce the missed recognition rate. Compared with other networks using larger convolution kernels for feature extraction, VGG16 uses $3 \times 3$ small convolution kernels repeatedly to replace the 16 layer network model with large convolution kernels, which is helpful for local feature extraction and can extract more abstract high-order features. Figure 1 shows VGG16 network diagram.

![Figure 1. Figure with VGG16 network.](image)

The propagation process of VGG16 includes convolution operation, activation function (Relu), pooling, full connect, softmax classification. The convolution layer and pooling layer of vgg16 network adopt the same kernel function, and the convolution block structure is formed by stacking the convolution layer and pooling layer, which has the advantages of simple structure and easy to form deep network structure.

3. Our Approach

3.1. Dense Block network
Although the structure of VGG16 is simple, the increase of network layer will lead to more network parameters, which is easy to cause resource waste and affect network performance. Therefore, this paper takes VGG16 as the basic network and introduces dense connection network block. By modifying the network, the transmission rate of information and gradient is enhanced, the network identification ability is improved, and the parameter quantity is reduced to a certain extent. Dense block connection network (Dense Block) mainly includes Bottleneck Layer and Transition Layer. The function of Dense Block is to control the close connection between layers to realize feature reuse, enhance image detail extraction.
and improve transmission efficiency of convolution neural network, and avoid resource waste; the role of the Transition Layer is to prevent overflow by controlling the number of channels in the network. Figure 2 shows the Dense Block network diagram.

![Dense Block network diagram](image)

The basic structure of Dense Block connection network includes Batch Normalization(BN), Rectified Linear Unit(ReLU), convolution operation with convolution kernel of 1×1 and 3×3. The input of the current layer comes from the output of all layers in front of it, which makes the transfer of features and gradient more efficient. The Dense Block network adopts the structure of linear combination functions BN + ReLU + 1×1Conv and BN + ReLU + 3×3Conv, that is, the 3×3 convolution of each Dense Block contains 1×1 Bottleneck_Layer convolution operation. In this way, the amount of calculation is reduced by reducing the number of characteristic graphs, and the characteristics of each channel can be fused. Moreover, 1×1 convolution operation is added in the adjacent Dense Block to compress the relevant parameters.

3.2. Channel attention mechanism

Attention mechanism in computer vision enables neural networks to learn the important features of the target by ignoring irrelevant features. In order to improve the recognition accuracy of the network and enhance the generalization ability of the network, we introduce an end-to-end training channel attention mechanism (channel_att) in Bottleneck_Layer, which can obtain more features of the image through the weight information acting on the corresponding feature position or dimension. Because the small changes of the output characteristics of the activation layer in the network will have a great impact on the final fusion results, we need better regular features to prevent over fitting. Therefore, normalization is added to the channel_att to regularize the feature map, so as to obtain a stable image representation. Figure 3 shows the Improved channel_att.

![Improved channel_att diagram](image)

In channel_att, for an input two-dimensional image, the output feature through Bottleneck_Layer activation layer is represented as channel_att input. L2 normalization is used in this paper, it can be realized by the operation in neural network, the operation equation as shown.

\[
\hat{x}_i = \frac{x_i}{\max(\|x\|, \theta)}
\]  

(1)

where, \(x_i\) is the input feature, \(\hat{x}\) is the input feature after regularization, \(\|x\|\) is expressed as.
\[ \| s \|_2 = \sqrt{\sum_{i=1}^{m} |x_i|^2} \]  
\hspace{10cm} (2)

where, \( m \) is the number of samples for L2 normalization. In the channel_att, the L2 normalized layer is connected to the input of the global average pooling layer. The input is the output of the activation function of Bottleneck_Layer, and the output is to obtain the normalized image features.

The normalized feature map is processed into global average pooling operation, and each channel will get its own global information. The equation is shown in eq (3).

\[ z_d = F_d(u_d) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} u_d(i, j) \]  
\hspace{10cm} (3)

where, \( u_d \) is the feature graph of size \( WH \times D \), and the output result represents a one-dimensional array of length \( D \), \( z_d = [z_1, z_2, z_3, ..., z_d] \), The corresponding feature maps of \( D \) channels obtained by global average pooling are connected with each channel adaptively through convolution and activation function, The equation is shown in eq (4).

\[ s = F(z, w) = \delta(g(z, w)) = \delta(\theta(\varphi(w(z)))) \]  
\hspace{10cm} (4)

where, \( w_1 \) and \( w_2 \) are two different fully connected layers, \( \varphi \) is the relu activation function and \( \delta \) is sigmoid activation function. The equation is shown in eq (5).

\[ s = F(u_d, s_d) = u_d \times s_d \]  
\hspace{10cm} (5)

\( s_d \) is the feature weight corresponding to the D channel, and \( x_d \) is the feature map corresponding to the D channel after activation. With the introduction of channel_att, the Dense block network can extract more abundant attention information, which can effectively extract the features of the network and enhance the robustness of the network.

3.3. Joint network based on channel attention mechanism

By adding Dense Blocks network and channel attention mechanism to the structure of VGG16 network, a joint network smoke recognition method based on channel attention mechanism is constructed. Dense Blocks network are added to make feature reuse; channel attention mechanism improves the accuracy of network recognition. The network structure is shown in Figure 4.

![Figure 4. Joint network based on channel attention mechanism](image)

The network model constructed in this paper consists of 7 convolution layers, 2 dense network blocks, 4 pooling layers and 2 full connection layers. The input image size is 224 × 224 × 3, and the number of convolution kernels in the first layer is 3×3 and the number is 24; The convolution kernels of layer 2 Dense Blocks are 1 × 1 + 3 × 3, and the number of convolution kernels is 96;The size of convolution kernel in the third Transition_Layer is 1 × 1 and the number is 48;The size of convolution kernels in layers 4-6 is 3×3 and the number is 96; The size of convolution kernel in layer 7 is 1×1 + 3×3 and the
number is 240; The size of convolution kernels in layers 8-10 is $3 \times 3$ and the number is 480; The 11-12 layers is full connection layers, and the corresponding size is 2.

4. Experimental results and analysis
The running environment of this experiment is ubuntu16.04, the CPU is i7-7600k and GPU is NVIDIA GTX 1070ti, using Python 3.6 language, the deep learning framework is TensorFlow.

4.1. Dataset
As the fewer of smoke datasets, some data are added to this public dataset. This public dataset consists of 2800 smoke images and 2900 non smoke images. In addition, the sum of images and smoke video frames obtained from online search make up 16568 smoke maps and 6315 non smoke images. The dataset contains data of different shapes, concentrations and sizes of smoke areas in different scenes. The details of these data are shown in Table 1 and some data graphs are shown in Figure 5.

Table1. Dataset details table.

|                      | Number of smoke images | Number of non-smoke images |
|----------------------|------------------------|---------------------------|
| Train                | 14990                  | 5217                      |
| Test                 | 1578                   | 1098                      |

![Figure5. Partial data.](image)

4.2. Related evaluation parameters
In order to measure the effectiveness and efficiency of the smoke recognition algorithm in this paper, we use the recognition rate (RR), accuracy rate (AR), false alarm rate (FAR) and precision (Pre) as the evaluation indexes of this experiment[6], the definition is as follows.

$$RR = \frac{T_p}{T_p + \eta_p} \times 100\% \tag{6}$$

$$AR = \frac{T_p + T_\mu}{Q_p + Q_n} \times 100\% \tag{7}$$

$$FAR = \frac{T_\mu}{T_p + \mu_n} \times 100\% \tag{8}$$

$Q_p, Q_n$ and represent the number of smoke and non_smoke image samples, $T_p$ is the number of smoke image samples, $\eta_p$ is the number of mistakenly detected non smoke images in smoke image samples, $T_\mu$ is the number of correctly detected non smoke images in non_smoke image samples, and $\mu_n$ is the number of mistakenly detected smoke images in non_smoke image samples.

4.3. Experiment
The comparative experiments in this paper are VGG16 and DenseNet, The method in this paper is based on VGG16 and Dense Block integrated network and joint network based on channel attention mechanism. In order to facilitate the experiment, the following are called VGG16, DenseNet, Inception, DENVGG and DENVGG_ATT, The experimental results are shown in Table 2.
Table 2. Comparative experiment table.

| Model Name | RR(%) | AR(%) | FAR(%) | Test_time(ms) |
|------------|-------|-------|--------|---------------|
| VGG16      | 97.33 | 96.15 | 3.8    | 0.056         |
| DenseNet   | 97.73 | 97.80 | 2.4    | 0.071         |
| Inception_v3 | 99.56 | 99.55 | 0.3    | 0.037         |
| DENVGG     | 99.81 | 99.85 | 0.06   | 0.085         |
| DENVGG_ATT | 100.0 | 100.0 | 0      | 0.045         |

As can be seen from table 2, compared with the general recognition network, the joint network method based on channel attention mechanism has higher recognition rate and accuracy rate, and lower false alarm rate. Compared with VGG16, DenseNet and Inception_v3, the recognition rate is increased by 2.48%, 2.08% and 0.25%, the accuracy rate and of DENVGG Net are improved by 3.7%, 2.05% and 0.3% respectively, and the false alarm rate is reduced by 3.74%, 2.34% and 0.24% respectively. It shows that the fusion network with Dense Blocks network improves the smoke recognition ability compared with the original network. Compared with DENVGG Net, the accuracy rate of DENVGG_ATT Net is improved by 0.04%, the recognition rate is increased by 0.37%, the false alarm rate is reduced by 0.06%, and the time is reduced by 0.045 ms. Experimental results show that the proposed joint network smoke recognition method based on channel attention mechanism effectively improves the performance of smoke recognition, can reliably identify smoke in various scenarios, and has better application value.

5. Conclusion
Aiming at the problems of low accuracy and low recognition accuracy of smoke recognition network, a joint network smoke recognition method based on channel attention mechanism is proposed in this paper. By adding Dense Blocks connection network in the convolution layer of VGG16, the reuse of features is improved and the information flow in the network is enhanced. Channel attention mechanism is introduced to enhance the generalization ability of the network and reduce the computational complexity. Compared with VGG16, DenseNet and Inception in the same dataset, the results show that the DENVGG_ATT method can improve the recognition accuracy and recognition rate to 100% and reduce the false alarm rate to 0, and reduce the training time to 0.04ms, which verifies the practicability of the method in this paper. Because the smoke recognition databases published on the Internet are not unified, with small numbers and low scene complexity, it is difficult to compare the algorithms. The following work will further improve the network and expand the training dataset to enhance the performance of the network.

Acknowledgments
This work was supported by the Wuhan Institute of Technology Software engineering laboratory and the eleventh Graduate Innovation Fund of Wuhan Institute of Technology (CX2019254, CX2019250).

References
[1] Lin, C.Z., Zhang, W., Wang, X. (2018) Smoke Detection Algorithm based on Dense Optical Flow and Edge Features.J. Computer engineering and Science., 040 (007):1213-1220.
[2] Zhao, Y.Q., Zhou, Z., Xu, M.M. (2015) Forest Fire Smoke Video Detection Using Spatiotemporal and Dynamic Texture Features.J. Journal of Electrical and Computer Engineering, 2015:1-7.
[3] Zhou, Z. Q., Shi, Y.S., Gao, Z.F, et al. (2016) Wildfire Smoke Detection based on Local Extremal Region Segmentation and Surveillance.J. Fire Safety Journal., 85:50-58.
[4] Yin, M.X., Wang, L., Sun, L.Y. (2019) Adaptive Smoke Detection Algorithm based on Multi-Feature Fusion.J. Building Science, 35(9): 26-31.
[5] Chen, J.Z., Wang, Z.J., Chen, H.H, et al. (2016) Dynamic Smoke Detection Using Cascaded Convolutional Neural Network for Surveillance Videos.J. Journal of University of Electronic Science and Technology of China, 46(6):992–996.
[6] Xie, H., Chen, Y.J., Yuan, X.F., Chen, H.F., Wang, L.C. (2020) Video Smoke Detection Using Spatiotemporal Dual Path 3D Residual Convolutional Network. J. Computer Engineering and Applications, 56(18):143-149.

[7] Lin, G., Zhang, Y., Xu, G, et al. (2019) Smoke Detection on Video Sequences Using 3D Convolutional Neural Networks. J. Fire Technology, 1827–1847.

[8] Huang, G., Liu, Z., Maaten, L.V. D, et al. (2017) Densely Connected Convolutional Networks. J. Computer Vision and Pattern Recognition, arXiv:1608.06993.

[9] Xu, G., Zhang, Y., Zhang, Q, et al. (2017) Deep Domain Adaptation based Video Smoke Detection Using Synthetic Smoke Images. J. Fire Safety Journal, 93(9):53-59.

[10] Wang, W.P., Mao, W.P., He, J.L. (2017) Smoke Recognition based on Deep Transfer Learning. J. Journal of Computer Applications, 37(11):3176-3181.