Simple and Efficient Smoke Segmentation Based on Fully Convolutional Network

Mao-shen LIU1,2, Xiao-tian XIE1,2, Ke GU1,2 and Jun-fei QIAO1,2,*

1Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
2Beijing Key Laboratory of Computational Intelligence and Intelligent System
*Corresponding author

Keywords: Fully convolutional network, Smoke segmentation, Flare stack smoke.

Abstract. In this paper, a shallow fully convolutional network for image smoke segmentation is designed to solve the real-time monitoring of smoke emitted by the flare stack. This algorithm can quickly and effectively distinguish the smoke area in the image, which can determine the actions of the flare stack control system to improve combustion efficiency. The main difficulty in segmenting the smoke in the flare stack image is the variegated texture and shape of the smoke and the varying brightness, color and other disturbances of the background. According to the above problems, we only use one layer of convolution to extract large numbers of low level features such as texture and color, and further utilize two special convolutional layers, separable convolution and 1×1 convolution, to map the final segmentation result. Through experiments on different data sets, our algorithm has the best accuracy and efficiency.

Introduction

The flare stack is an important facility in the petrochemical plant to ensure the safe production of the plant and reduce environmental pollution [1]. Real-time and accurate monitoring of the flared flare stack can effectively provide reliable feedback information of the flare stack control system. Efficient combustion of the flare stack is not only an important technical benchmark in the industry, but also a requirement for energy conservation and environmental protection. Therefore, monitoring the combustion state of the flare stack is of great significance.

Identifying smoke by pixel is often referred to as smoke segmentation. The difficulty of smoke segmentation compared to other targets is that the smoke shape, texture and color are variable and easily confused with the background. The traditional method of smoke segmentation is designed for the characteristics of smoke under certain circumstances, such as shape [2], color [3] and motion [4], so that the accuracy is greatly affected by environmental changes. In the past few years, deep learning methods have achieved excellent results in the field of computer vision. Almost the traditional algorithm has been withdrawn from the historical arena. Image semantic segmentation based on convolutional neural networks is a hotspot research. The fully convolutional network (FCN) was first applied in semantic segmentation tasks in [5], and the accuracy of semantic segmentation has been greatly improved. Subsequently, a semantic segmentation network based on encoder-decoder network structure uses a symmetric network structure to learn detailed segmentation results [6]. At present, most of the segmentation network feature extraction parts are based on VGG16 [7] or ResNet [8] structures, such as [9] and [10], which have unprecedented advantages in multi-category semantic segmentation tasks.

The above image semantic segmentation network is used for multi-object segmentation, which makes the algorithm structure complex. However, the problem of image smoke segmentation belongs to the segmentation of specific targets. Only the two types of pixels of the smoke region and the non-smoke region need to be distinguished. Therefore, using the complex structure described above makes the network too redundant, which is not conducive to rapid detection in the industry. Through our analysis of large amounts of data, the main difficulty in the segmentation of smoke in the flare stack image is the varied texture and shape characteristics of the smoke and the varying brightness, color and other disturbances of the background sky. These features are low-level features that can be
extracted by shallow convolution. Therefore, we design an extremely shallow network with only three layers of convolution. A large number of low-level feature maps are extracted through the first layer, which is equivalent to mapping each pixel point of the original image to a high-dimensional space. Further, the classification result of the input image pixel points is obtained by two-layer pointwise convolution, which is equivalent to the two-layer full connection between the first layer feature map channels. According to the image having local structural information, the first pointwise convolution layer is replaced by separable convolution [11]. Compared with other deep learning methods, the proposed algorithm has advantages in speed and accuracy in the smoky image segmentation problem of the flare stack. The rest of the article mainly introduces the preparation of training data, model design and comparative experiments, and finally summarizes.

Proposed Algorithm

As mentioned above, the main difficulty in the segmentation of smoke in the flare stack image is the variegated texture and shape features of the smoke and the varying brightness, color and other disturbances of the background sky. These low-level features are confused with each other, making the segmentation of smoke different from the traditional semantic segmentation task. To this end, we have designed a special network to solve the problem of fast smoke segmentation, referred to as ‘FSSN’. The details of the preparation of the data set and the algorithm design principles and the training are described in detail below.

Data Preparation

The image sample we prepared is the real picture obtained by the flare stack test field. The sample is mainly divided into two parts under the influence of two major factors, sunny and cloudy. In order to facilitate the manual segmentation of the smoke area to make benchmark labels, the smoke image only intercepts the area containing the smoke. Due to the lack of smoke data, the image deformation [12] is used to expand the number of smoke images, as shown in A of Fig. 1. The image segmentation benchmark is produced by splitting the smoke images by manual adjustment of the threshold value, and the result is that the smokeless area are zero and the smoke area are one, as shown in B of Fig. 1. Further, the benchmark of the smokeless image is an all-zero matrix, as shown in C of Fig. 1. In order to uniform images sizes during training, we cropped all images into 128*128 sizes. Finally, by the method of image rotation, the number of smoke images is increased to balance the smoked and smokeless pixel regions. In the end, the training image was about 36000, the validation image was about 3400, and the test image was about 3400.

Network Structure and Principle

Our network is very shallow compared to existing image segmentation networks and works well in our missions. The network successfully uses only three layers to segment smoke areas from different sky backgrounds. It is mainly due to the specially designed structure for objects and background features in this task. The proposed FSSN is described in detail below.

Task Analysis. The main difficulty of this task is the complex change of the sky backgrounds. The traditional image segmentation algorithms need to manually design the feature extraction method based on human experience, which is not suitable for tasks with complex background changes and
high accuracy requirements. Therefore, we can learn a large number of feature extractors to complete
the feature extraction task in a data-driven manner. The convolution operation can effectively extract
features from RGB images and complete simple feature extraction such as texture and edges. It is also
possible to synthesize various grayscale images for segment smoke under different backgrounds by
combining RGB three channels. As shown in Fig. 2, the first two lines show the 128 feature maps
extracted from the RGB image by the first layer of convolution, each providing different important
information for image segmentation. It is difficult to manually design 256 feature extraction methods
to complete the image segmentation task.

**Network Structure.** All hyperparameters in the network are obtained by trial-and-error. The
network structure is shown in the Fig. 2. The first layer consists of 128 5*5 size convolution kernels.
Since the image has local characteristics, the category of each pixel is more easily distinguished by
local pixels. Therefore, the second layer uses a separate convolution which consists of a depthwise
convolution and a pointwise convolution. The depthwise convolution, kernel size is 5*5, obtains local
information of the feature maps of the pixel to be classified. And the pointwise convolution operation
and the third layer form a channel-wise fully-connected classifier. As shown in Fig. 2, the F_i
expresses 128 kinds of image information. Then, by inputting the local information of each pixel
point of the 128 feature maps, depthwise local pixels, into the channel-wise classifier, then the
classification of each pixel of the input image is obtained. The activation function of the first two
layers is Relu and Softmax is used in the last layer to output the class probability of each pixel. During
training, the last layer uses the SpatialDropout [13] to randomly discard 20% of the feature maps to
improve the generalization of the network and enhance the robustness of the network.

![Network Structure Diagram](image)

**Experiments**

In order to verify the algorithm proposed in this paper, we designed a smoke segmentation
experiment for the image of the flare stack. The test data is divided into two data sets: patch images
and entire images. The patch images set contains 3600 128*128 images retained during the process of
making the training data set, and the entire images set contains the actual scenario of the burning
flare. Due to computer performance limitations, the entire image is downsampled to a size of 480*480
for testing. The algorithms involved in the comparative experiments in this paper are FCN8, FCN32,
and SegNet algorithms. The experimental environment is the Windows 10 operating system running
on a PC with Inter(R) Xeon(R) CPU 2.00GHz and an Nvidia GeForce GTX 1080.

![Experiments Diagram](image)
Training Details. In this task, the segmentation task is considered to be a binary classification task for each pixel, so the loss function of the network is selected as categorical crossentropy. The network uses the adadalt optimization method [14], where the initial learning rate is set to 1, and the adadelta decay factor is set to 0.95. The input of the network needs to be normalized by subtracting the mean value and dividing the variance for each pixel. It is necessary to disturb the training sample sequence and set the batch to 4. Finally, the number of training epochs of the network is set to 50 in total.

| Methods | pixel acc. [%] | mean acc. [%] | mean IU [%] | f. w. IU [%] | mean time [ms] | parameters |
|---------|----------------|--------------|-------------|-------------|----------------|-------------|
| FCN-32s | 96.95          | 92.90        | 90.45       | 95.03       | 159            | 134.28 Million |
| FCN-8s  | 97.58          | 94.68        | 93.05       | 96.44       | 176            | 134.27 Million |
| SegNet  | 96.68          | 93.72        | 90.98       | 94.59       | 147            | 5.46 Million   |
| FSSN    | 98.07          | 97.89        | 96.28       | 97.20       | 43             | 0.046 Million   |

Comparative Analysis. The indicator results on the patch images set are given in the Table 1. It is obvious that our algorithm performs best on the smoke image segmentation test set compared to other algorithms. In the results, the performance of the FCN-8s algorithm is optimal in the comparison algorithm, where the mean IU is 93.05%. Our proposed FSSN yields a high result of 96.28% on mean IU. At the same time, our algorithm averages 43ms per patch, which is only 24.4% of FCN-8s. Our network parameters are almost ignorable compared to other methods. Accuracy, efficiency and network complexity indicate that our proposed algorithm is more suitable for practical projects. Fig. 3 shows the results of the entire image in the flare stack. The FCN32 effectively identified some of the smoke areas in the image, while the smoke in (2) and (3) was not recognized. FCN8 has improved performance compared to FCN32, but it also ignores some areas of smoke. The SegNet network recognizes a richer smoke area, but it does not effectively resist the interference of complex backgrounds, such as (3) results. The results of our FSSN shows us a better effect. In the (1) and (2), it is effective to segment all the smoke areas visible to the naked eye on the sunny. At the same time, the effect is also evident in the (3) (4) cloudy image. Moreover, the anti-interference of the algorithm is very strong, as in (3), the smoke area is accurately identified in a complex background.

![Figure 3. Segmentation results of entire images. In the segmentation result, white is smoke and black is background.](image-url)
Summary
This paper designs a fully convolutional neural network for smoke segmentation of flare stack images. Since smoke is mainly present in images with low-level features, the proposed is designed to extract a large number of low-level features to classify each pixel. A classifier for each pixel is formed by a layer of separate convolution and a pointwise convolution. The separation convolution can extract the local information of the pixel to be identified, so that the network is more robust to background interference. Compared with the classical image semantic segmentation network, the proposed algorithm has an extraordinary effect on accuracy and efficiency. It effectively illustrates the effectiveness of the proposed network in the flare stack image smoke segmentation. It is further explained that the smoke segmentation does not require a deep network to extract the semantic features, and only the low level features can effectively complete the smoke segmentation task.

References
[1] S. David, A proposed comprehensive model for elevated flare flames and plumes, Flaregas Corporation, AIChE 40th Loss Prevention Symposium, Apr. 2006.
[2] F. Alexander, C. H. Danilo, and J. Kanghyun, Fast smoke detection for video surveillance using CUDA, IEEE Transactions on Industrial Informatics, Volume 14, Issue 2, pp. 725-733, 2018.
[3] A. G. Sagar, N. Shashank, and N. Preeti, Smoke Detection in Digital Frames. International Research Journal of Engineering and Technology, Volume 5, Issue 4, pp. 3843-3846, 2018.
[4] J. Yang, L. Gaohua, W. Jinjun, F. Jun, and Z. Yongming, Early video smoke segmentation algorithm based on saliency detection and gaussian mixture model, Computer Engineering, Volume 42, Issue 2, pp. 206-209, 2016.
[5] E. Shelhamer, J. Long, T. Darrell, Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39(4):640-651, 2014.
[6] B. Vijay, K. Alex, and C. Roberto, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE transactions on pattern analysis and machine intelligence, Volume 39, Issue 12, pp. 2481-2495, 2017.
[7] S. Karen, and A. Zisserman, Very deep convolutional networks for large-scale image recognition, International Conference on Learning Representation, 2015.
[8] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.
[9] C. Liangchlieh, P. George, K. Iasonas, P. Kevin, and L. Y. Alan, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE transactions on pattern analysis and machine intelligence, Volume 40, Issue 4, pp. 834-848, 2018.
[10] G. Lin, A. Milan, C. Shen, and I. D Reid, Refinenet: Multi-path refinement networks for high-resolution semantic segmentation, CVPR, pp. 5168-5177, 2017.
[11] M. Wang, B. Liu, and H. Foroosh, Design of Efficient Convolutional Layers using Single Intra-channel Convolution, Topological Subdivisioning and Spatial "Bottleneck" Structure, arXiv preprint arXiv:1608.04337, 2016.
[12] S. Schaefer, T. McPhail, and J. Warren. Image deformation using moving least squares. ACM Trans. Graph. 25, 3, 533-540, Jul. 2006.
[13] J. Tompson, R. Goroshin, A. J., Y. LeCun, C. Bregler, Efficient object localization using convolutional networks. CVPR. 2015.
[14] M D. Zeiler, ADADELT A: an adaptive learning rate method. arXiv:1212.5701, 2012.