Video Smoke Detection using Deep Domain Adaptation Enhanced with Synthetic Smoke Images

Gao Xu, Qixing Zhang, Gaohua Lin, Jinjun Wang, Yongming Zhang
State Key Laboratory of Fire Science, University of Science and Technology of China, Hefei, Anhui 230026, PR China

Abstract

In this paper, a deep domain adaptation based method for video smoke detection is proposed to extract a powerful feature representation of smoke. Due to the smoke image samples limited in scale and diversity for deep CNN training, we systematically produced adequate synthetic smoke images with a wide variation in the smoke shape, background and lighting conditions. Considering that the appearance gap (dataset bias) between synthetic and real smoke images degrades significantly the performance of the trained model on the test set composed fully of real images, we build deep architectures based on domain adaptation to confuse the distributions of features extracted from synthetic and real smoke images.

This approach expands the domain-invariant feature space for smoke image samples. With their approximate feature distribution off non-smoke images, the recognition rate of the trained model is improved significantly compared to the model trained directly on mixed dataset of synthetic and real images. Experimentally, several deep architectures with different design choices are applied to the smoke detector. The ultimate framework can get a satisfactory result on the test set. We believe that our method own strong robustness and may offer a new way for the video smoke detection.

Keywords: Synthetic smoke image, deep architecture, domain adaptation, feature distribution

Introduction

Video smoke detection, as a promising fire detection method especially in open or large spaces and outdoor environments, has been researched more than ten years [1]. Compared to the traditional video smoke detection methods based on the pattern recognition technology to extract shallow features manually, the deep architecture obtains the more essential features independently.

Typically, the team of Microsoft Research developed a residual network [2] and achieves 3.57 % error rate on the ILSVRC 2015 classification task, which is lower than the 5.1 % error rate of the human eyes.

Due to the smoke images samples limited in scale and diversity for CNN training, we use synthetic smoke images to extend the training set. For our task, the practical benefit of synthetic smoke images is to increase recognition power of the model trained on them, rather than their visual effects. As there is certain appearance gap between synthetic and real smoke images, the difference of their statistical distributions can degrade the performance of the trained model on test set composed of real images. To tackle this problem, we apply the domain adaptation (DA) method to build the deep architecture, which is the standard approach to alleviate dataset bias caused by a difference in the statistical distributions between training and test data. This method confuses the distributions of features extracted from synthetic and real smoke images, and expand the domain-invariant feature space of smoke images off non-smoke images.

Smoke image dataset for CNN training

We build a synthesis pipeline (see Fig. 1) to produce synthetic smoke images of high diversity. The production process is automated in Blender-python. Compared to the pipeline for rendering the 3D rigid models in [3], visual simulation of smoke is more complex as the representation of synthetic smoke image is determined by numerical simulation and media rendering for smoke. Especially, due to the fuzzy transparency of early smoke, it is necessary to render smoke with background image instead of composition of them.

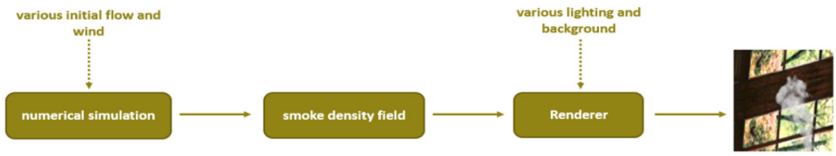


Fig. 1. Synthesis pipeline overview. To increase the diversity, the initial flow, wind, lighting and background are set randomly.

In detail, 5K real smoke images were selected every 5 or 25 frames from the captured videos, while 30 K synthetic smoke images were produced.

Another work is to extract the smoke region, as our recognition object is separated rectangular region of smoke. We cropped out the entire rectangular region of smoke. Meanwhile, the same number of non-smoke images were collected. In addition, a test set including 1000 images is created to evaluate the performance of the trained model.

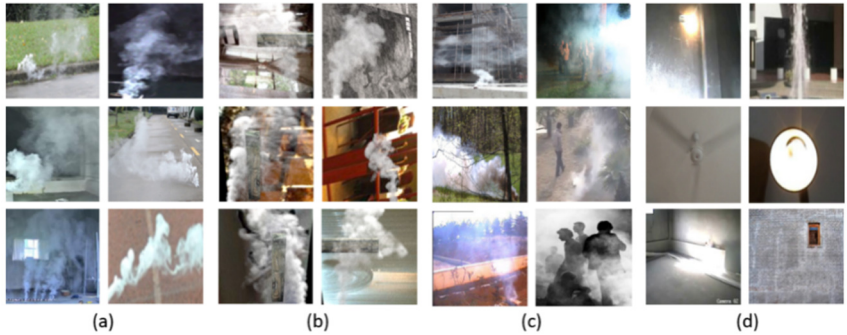


Fig. 2. Image samples: (a) real smoke images; (b) synthetic smoke images; (c) test smoke images; (d) test non-smoke images.

Network and layer function

As shown in Fig. 3, the whole dataset is divided into two datasets. The source dataset contains synthetic smoke images and real non-smoke images, and the target dataset contains real smoke images and real non-smoke images. In this case, we denote with multi-label (y_i^s, y_i^d) for each sample x_i of these two datasets, in which y_i^s indicates whether x_i is a smoke image or not, y_i^d indicates whether x_i is a real or synthetic image.

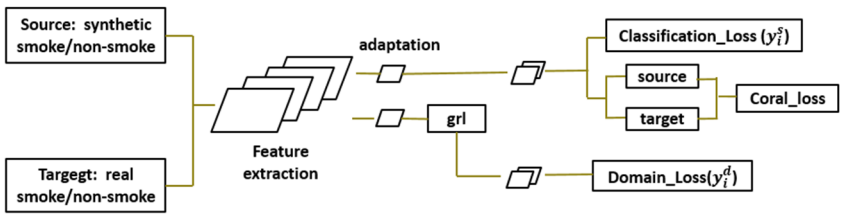


Fig. 3. This CNN architecture contains shared feature extraction layers, feature adaptation layers, and three loss function layers.

The softmax loss and hingle loss are defined as the classification loss and domain loss. At the training time, our architecture minimize the classification error L_s to obtain discriminative feature representation. Meanwhile, domain loss L_d is maximized to confuse (close) the distributions between synthetic and real smoke images through gradient reversal layer (GRL) [4]. The GRL plays the all-pass function during forward and multiplies the gradient by φ during backpropagation. We add the adaptation layer proposed in [5], which is used to prevent overfitting to the particular nuances of the distribution of synthetic images in source dataset.

Experiments show that it is difficult to obtain a satisfactory equilibrium between the two objectives (see Table 2). This supervised training only aligns the means but not the every local distribution. Therefore, the correlation of their features cannot be guaranteed. To confuse fully the feature distributions between them, we added a CORAL [6] loss layer to align the second-order statics of the source and target feature distributions for correlation alignment to make them closer.

The joint loss function of our architecture is as follow,

$$L = \alpha_{label} * L_s + \beta_{domain} * \varphi * L_d + \gamma_{coral} L_{coral} \quad \text{(Eq. 1)}$$

Experiments

In this section, the effectiveness of synthetic smoke images to training is confirmed and relevant evaluation of the domain adaptation based deep architecture are performed.

It is essential to clarify the effect of synthetic smoke images to the detectors. We train the model of AlexNet on different datasets. In order to qualitatively and quantitatively evaluate the performance of model at the whole test set, the correct detection(CD), the error detection(ED) and missed detection(MD) are measured in Table 1.

Table 1. The performance of model of AlexNet trained on different datasets.

Training set (contains non-smoke images)	CD	ED	MD
Real smoke images	0.6690	0.0526	0.6420
Synthetic smoke images	0.5700	0.2160	0.8060
Mixed dataset of real and synthetic smoke images	0.7380	0.0162	0.5160

Due to the limitation of real smoke images in scale and diversity, the model trained on them miss nearly 64% of the test smoke images which are quite different from the training smoke images. The performance of the model trained on the synthetic smoke images is more terrible, due to the synthetic smoke images are certain different on appearance from real images. Meanwhile, we train the model on the mixed dataset which consists of almost the same number of real and synthetic smoke images. The results show that the predictive accuracy is slightly improved.

Next, we use the domain adaptation based deep architecture to tackle this problem. Several deep architectures based on domain adaptation are trained on them. The predicted results of architectures with different design choices are reported in Table 2.

Table 2. The performance of different deep architectures based on domain adaptation.

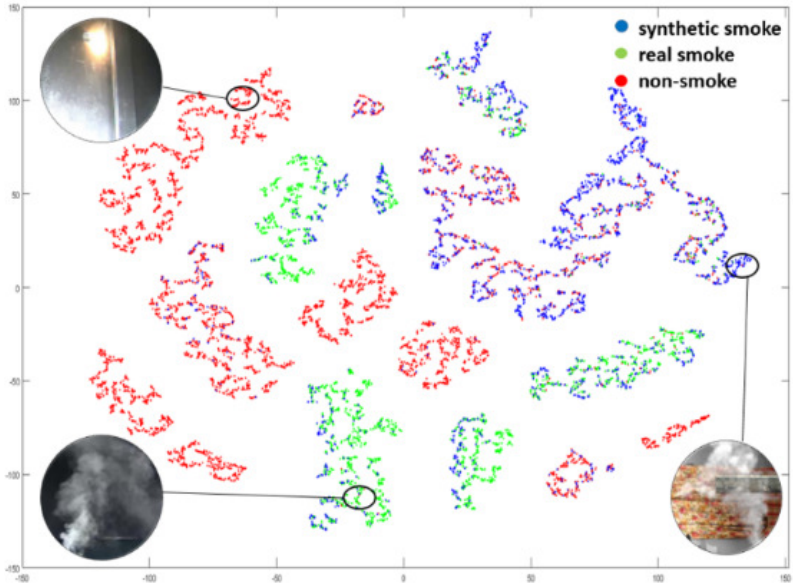
Architecture (with feature extraction layers)	CD	ED	MD
Ours with GRL	0.8170	0.1768	0.1920
Ours with GRL + adaptation layer for L_d	0.8080	0.2079	0.1640
Ours with GRL + adaptation layer for L_s and L_d	0.8520	0.1633	0.1240
Ours with GRL + CORAL with adaptation layer	0.9470	0.0447	0.0620

It can be seen that the predicted accuracy of the models of these deep architectures based on domain adaptation are improved significantly than that of the general architecture trained on the mixed dataset.

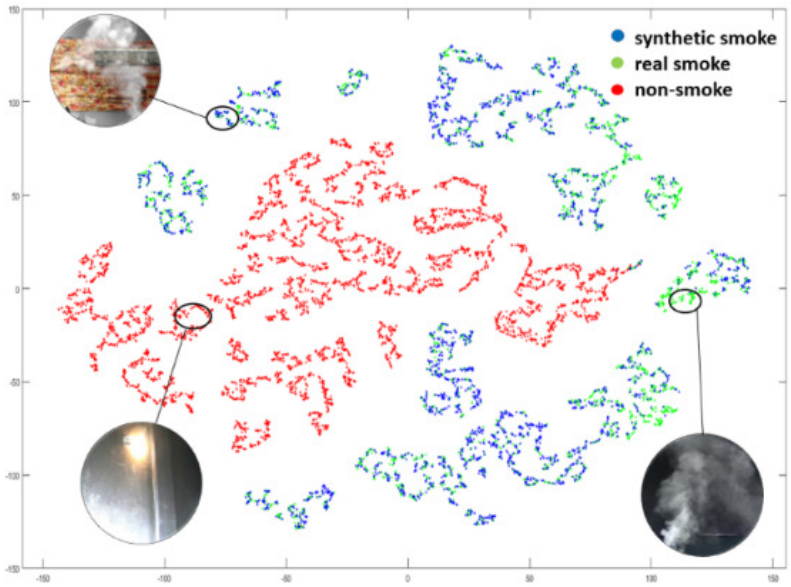
These architectures all use the gradient reversal layer (GRL) to connect to domain loss layer, for confusing the features of synthetic and real smoke images. By comparison, the adaptation layer is actually useful as it avoids the overfitting in synthetic smoke images, when added on the bottom of the layers connected to the L_s and L_d loss layers.

In our experiment, the last architecture achieved the best performance. As we describe before, the CORAL loss layer make the distributions of the two datasets closer, especially for the smoke image samples in the two datasets as the non-smoke image samples are basically the same. Of course, better performance may be obtained to adjust the architecture and training.

In order to represent the distributions of features extracted by adaptation architecture more intuitive, t-SNE visualizations of feature distributions are showed in Fig. 4. These point cloud show the effect of network on the distribution of deep feature representations of synthetic smoke real smoke and non-smoke images. Obviously, as shown in Fig. 4(i), the features of synthetic smoke images are easily confused with non-smoke images rather than real smoke images. By comparison, the feature distributions of real and synthetic images in Fig. 4(ii) are aligned and confused fully, separated from non-smoke images. This case has a fine performance as it expand the domain-invariant feature space for smoke images as expected.



(i) Results obtained with the model of Alexnet trained on mixed dataset.



(ii) Results obtained with the model of adaptation architecture trained on the source and target datasets.

Fig. 4. The visualizations for feature distributions of synthetic smoke (blue), real smoke (green) and non-smoke (red) image samples.

Conclusion

We propose a deep domain adaptation based approach for video smoke detection. We systematically produced adequate synthetic smoke images rich in variation. In order to prevent the degradation in performance of trained model caused by the appearance gap between synthetic and real smoke images in training set, we apply the domain adaptation based deep architecture to the classification task. Experiments confirmed the effectiveness of synthetic smoke images to training and investigate the effects of different design choices of the deep architecture on the predicted results. The ultimate framework can get a satisfactory result on the test set.

Acknowledgements

This work was supported by the National Key Research and Development Plan under Grant No. 2016YFC0800100, the National Natural Science Foundation of China under Grant No. 41675024, and the Fundamental Research Funds for the Central Universities under Grant No. WK2320000033 and No. WK6030000029. The authors gratefully acknowledge all of these support.

References

[1] Çetin, A.E., K. Dimitropoulos, B. Gouverneur, N. Grammalidis, O. Günay, Y.H. Habiboğlu, B.U. Töreyin, and S. Verstockt, *Video fire detection–Review*. Digital Signal Processing, 2013. 23(6): p. 1827-1843.

[2] He, K., X. Zhang, S. Ren, and J. Sun. *Deep residual learning for image recognition*. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[3] Su, H., C.R. Qi, Y. Li, and L.J. Guibas. *Render for cnn: Viewpoint estimation in images using cnns trained with rendered 3d model views*. in Proceedings of the IEEE International Conference on Computer Vision. 2015.

[4] Ganin, Y. and V. Lempitsky, *Unsupervised domain adaptation by backpropagation*. arXiv preprint arXiv:1409.7495, 2014.

[5] Tzeng, E., J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, *Deep domain confusion: Maximizing for domain invariance*. arXiv preprint arXiv:1412.3474, 2014.

[6] Sun, B. and K. Saenko. *Deep coral: Correlation alignment for deep domain adaptation*. in Computer Vision–ECCV 2016 Workshops. 2016. Springer.