Video Smoke Detection Using Deep Neural Networks Enhanced with Synthetic Smoke Plume Generation

Qixing Zhang*, Gao Xu, Gaohua Lin, Kewei Wang, Yongming Zhang*

State Key Laboratory of Fire Science

University of Science and Technology of China, Heifei, China, qixing@ustc.edu.cn zhangym@ustc.edu.cn

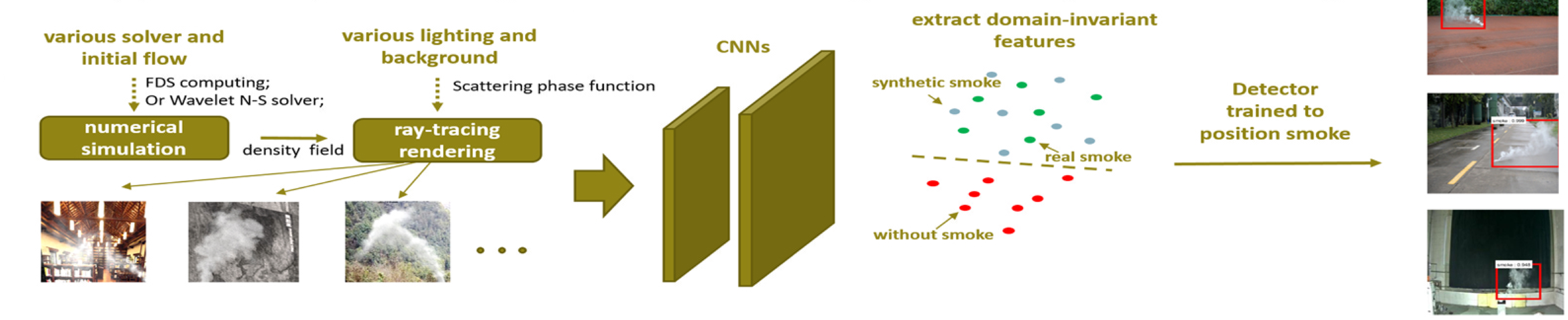
Goal: More Accurate Video Smoke Detection (VSD)

- VSD is a promising fire detection method especially in open or lager spaces and outdoor environments
- Current VSD technology based on traditional pattern recognition methods can not make this promise a reality in engineering
- Great success in the field of computer vision has been achieved by deep learning

Challenge:

- Large data set is expected to improve the deep learning performance
- Lack of fire smoke video samples due to the occasionality of fire accidences
- Difficult and expensive to collect samples by designed experiments
- Wide variation in the smoke shape, background and lighting conditions is expected

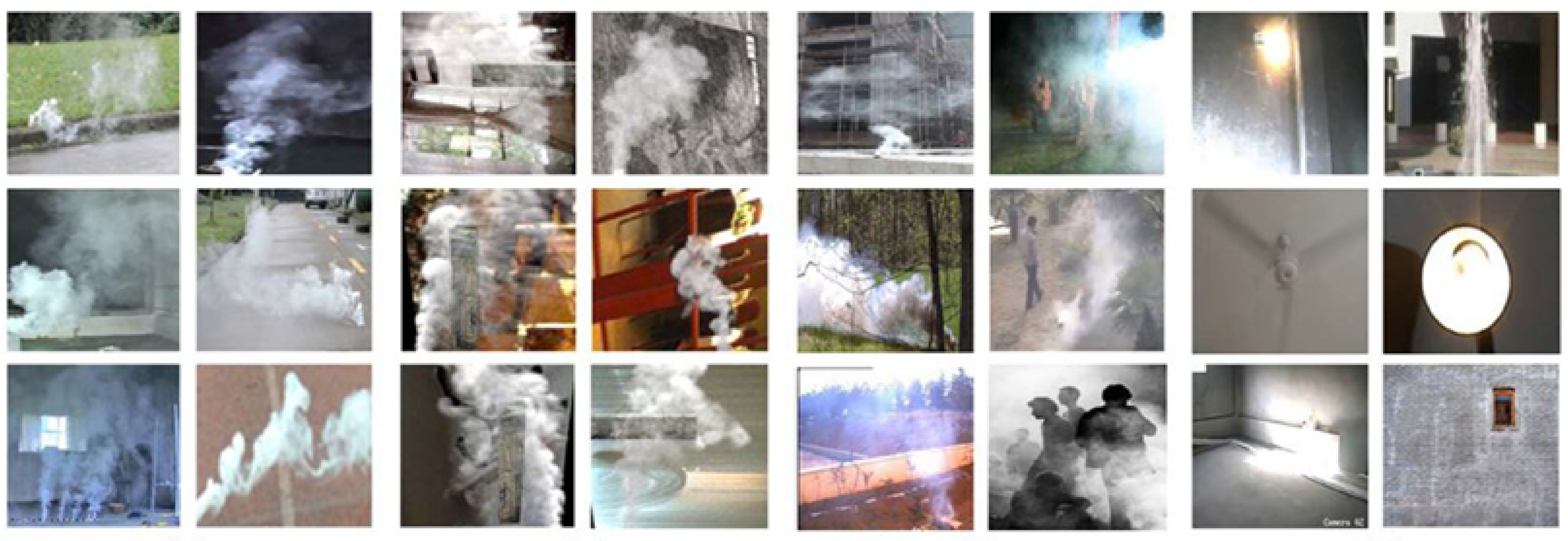
Our Approach: Synthesizing smoke video images for deep CNN training



We systematically produced adequate synthetic smoke images with variation in the smoke shape, background and lighting conditions. A deep architectures based on domain adaptation is build to confuse the distributions of features extracted from synthetic and real smoke images.

Dataset:

5K real smoke images, 30K synthetic smoke images of high diversity and a test set including 1000 images are created.

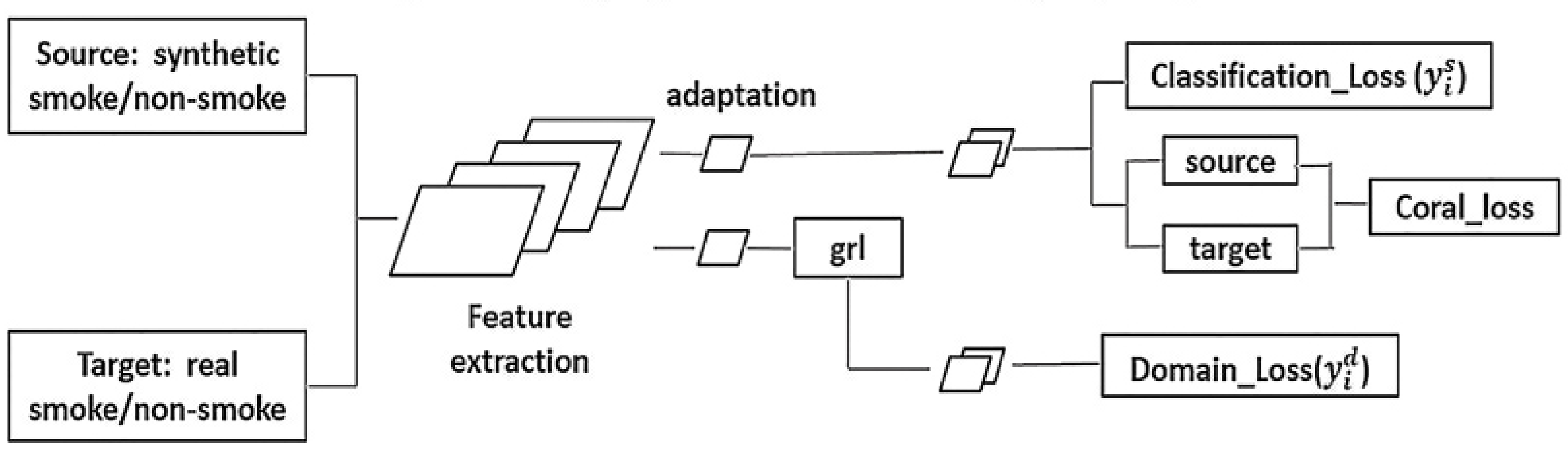


(a) (b) (c) (d)

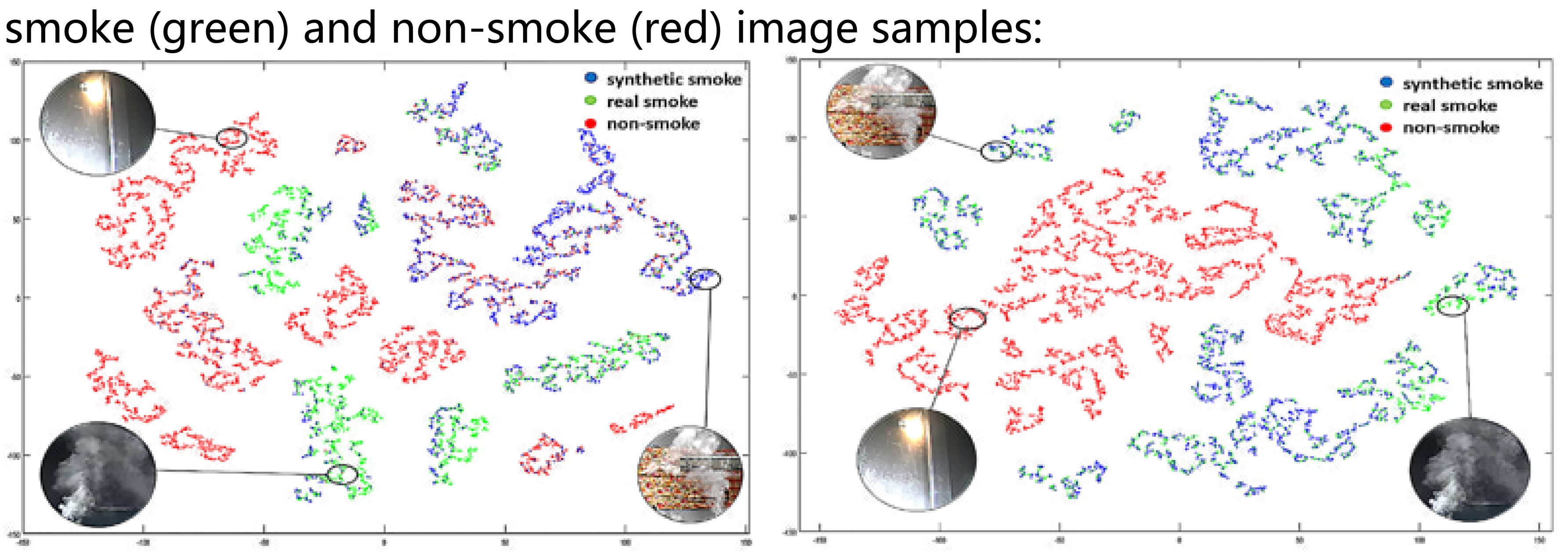
Image samples: (a) real smoke images; (b) synthetic smoke images; (c) test smoke images; (d) test non-smoke images.

Deep domain adaptation:

The whole dataset is divided into source dataset and target dataset. At the training time, our architecture needs to achieve two objectives: minimize the classification error L_s to obtain discriminative feature representation and maximize the domain loss L_d to confuse (close) the distributions between synthetic and real smoke images through gradient reversal layer(GRL).



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



(i) (ii)

(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.

Evaluation:

Firstly, we used the network structure of AlexNet and train the model on different datasets to clarify the effect of synthetic smoke images to the detectors. The performance of model of AlexNet trained on different datasets:

Training set(contains non-smoke images)	accuracy	false alarm	false negative
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

Secondly, several deep architectures based on domain adaptation are trained on the whole dataset(the target dataset contains 5k real smoke images, and the source dataset contains synthetic smoke images from 5k to 30k). The performance of different deep architectures based on domain adaptation :

Architecture (the layers added to feature extraction layers)	accuracy	false alarm	false negative
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. In our experiment, the last architecture achieved the best performance. The CORAL loss layer make the distributions of the two datasets closer.

The ultimate framework can get a satisfactory result on the test set. Inspired by the preliminary success, we believe that our approach is a start in the direction of utilizing deep neural networks enhanced with synthetic smoke images for video smoke detection. As the appearance of synthetic smoke images directly affects the performance of trained model, in the future work, we will investigate the effect of variations caused by synthesis on recognition power of model and our method maybe benefit from it.

References：

[1] Esteva A, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. Nature, 2017, 542(7639):115.

[2] Verstockt, S., et al. Future Directions for Video Fire Detection. Fire Safety Science, 2011. p. 529-542.

[3] Movshovitz-Attias, Y., et al. How useful is photo-realistic rendering for visual learning? arXiv preprint arXiv:1603.08152, 2016.