

Fire alarm based on image semantic description

Xuan Zhao^a, Dengyin Zhang^{b,*}, Huanhuan Bao^c

Institute of Internet of Things, Nanjing University of Posts and Telecommunications University, Xinmofan Road, Nanjing, China

^a 2458900120@qq.com, ^b zhangdy@njupt.edu.cn, ^c 335785739@qq.com

*Corresponding Author

Keywords: CNN, Object detection, Semantic description, Fire alarm

Abstract: The complexity of outdoor smoke scenes often leads to low accuracy of outdoor fire alarms. We propose a fire alarm method of automatic generation of image semantic description to improve the accuracy of early warning in outdoor scenes. By establishing a semantic context model based on Bayesian network, the convolutional neural network is used to extract feature values from the labelled samples and train them to realize object detection. Then we train the context semantic model of the target object in the segmentation region to realize the semantic description of the scene to distinguish whether the smoke is caused by fire. The experimental results on the image datasets of five types of smoke scenes show that the method we propose in this paper has higher accuracy and lower recall rate than the existing methods.

1. Introduction

The outdoor fire has the characteristics of wide spread and large space. At present, the fire alarm method of traditional sensors based on temperature and respirable particulate matter detection cannot meet the requirements of outdoor fire early detection. The current widely distributed digital cameras and computer vision technology enable real-time transmission of video images, real-time monitoring and instant outdoor fire warning.

So far, relevant researchers have taken advantage of the characteristics of far-reaching and easy detected smoke in fire disasters, and many image smoke detection methods have been proposed for fire alarm. However, existing smoke detection methods cannot distinguish fire smoke scenes from non-fire smoke scenes (such as fireworks and firecrackers, wild barbecue scenes, etc.), which results in false alarms. To this end, under the premise of ensuring the reliability of the fire system, improving the accuracy of fire early alarm has become one of the key issues that need to be solved urgently. Recently, different natural language description methods for automatically generating images have been used in image processing fields such as automatic image annotation, image retrieval^[1], image database management, and mining image important information^[2-3]. Compared with single object detection, the image semantic description combines the positional relationship and attribute association of the context scene with objects, which can realize the deep semantic mining of the images.

In this paper, statistical methods, intelligent computing, machine learning and other technologies are used to mine the objective laws hidden in smoke images, and the intelligent recognition of smoke scenes is integrated into fire warning. Combined with the method of image semantic description, and via the definition of the logical relationship between the local semantic objects and fire behavior, it is automatically determined whether the smoke is caused by fire. On this basis, the convolutional neural network is used to learn the spatial context of the image components in the smoke scenes, and the local features of the semantic objects in smoke scenes are obtained. At the same time, the spatial relationship between the local objects and smoke is used to correct the judgment results. The semantic description of smoke scenes and the fire early alarm model are obtained. Using spatial context feature^[4] fusion makes subsequent scene representation and decision strategies more accurate and efficient.

2. Local context semantic model for target objects

The implementation of target recognition is the basis of object-oriented semantic description^[5]. Traditional semantic model usually realizes the semantic expression of the scenes based on the frequency of the target objects in the image, which ignores the influence of the surrounding environment on the semantic expression of the target objects. On the basis of the object detection^[6] results, the paper uses the semantic analysis method to model the local context of the target objects^[7], and train the particular space that affects local semantic objects in different scenarios, and apply it to the subsequent work of the scene expression in order to reduce the interference and influence of the surrounding environment on semantic expression.

2.1. Local context model based on Bayesian network

Bayesian network is a probabilistic network in machine learning, and is also a Probabilistic Graphical Model^[8-9] (PGM). The known conditions and results are mapped by prior knowledge using conditional probability to predict the probability of the results. In order to obtain the most accurate probability conditions, it is necessary to use a large number of manually labelled training sets to train the model prediction likelihood that best matches the training data distribution result in the test processes. The local context model based on Bayesian network in this paper is used to establish the contextual relationship model of the target objects in local scenes, and prepare for the subsequent semantic expression work. Among them, the target objects, the scenes within the limited scope of the target object, and the relationship between the two are used as three nodes to establish a probability map model based on Bayesian network.

In smoke scenes, various objects that may cause suspected fire smoke, such as fireworks, chimneys, bonfires, etc., can become targets. On the other hand, the features of scenes within the local context of the target object can also affect the fire alarm judgements. For example, if the firecrackers are fired with its surrounding area not empty or not any supervision, the system will trigger the fire alarm just in case. To this end, we used the scene feature sampling process shown in Figure. 1.

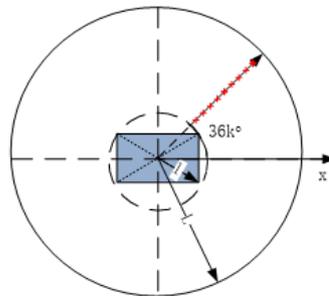


Figure. 1 Model of sampling features of the target objects scope

In figure. 1, the blue rectangle is represented as a detected target object box, and its one half length of its diagonal is denoted as l . Assume that L equals ten times the l , L will the radius of sampling range of the target object's contextual features and the centre of the detected target object box is the centre of the sampling round. Take the centre of the detected target object box as the polar coordinates' original point, and $L-l$ is the divided into 10 parts in the same direction. So, in the $36k^\circ$ direction, the feature point sampling coordinates are $(0.9l+0.1L, 36k^\circ)$, $(0.8l+0.2L, 36k^\circ)$... $(L, 36k^\circ)$, where $k = 1, 2, 3 \dots 10$. In this way, the spatial positional relationship between the surroundings and the target objects, which implies whether there is a fire hazard or the like, is obtained.

2.2. Semantic context learning based on smoke scenes

Firstly, based on the survey data of a large number of acquired data sets and smoke scenes, the exhaustive method is used to define the correlation sets R of relationships between the partial target objects and the context scenes. In addition, considering the positional relationships between the contextual objects and the target objects in searching areas, in order to reduce the complexity of

semantic expressions, we will divide the polar coordinates into $0^\circ\sim 90^\circ$, $90^\circ\sim 180^\circ$, $180^\circ\sim 270^\circ$, $270^\circ\sim 360^\circ$ at its centre, corresponding to the expressions of "top right", "upper left", "lower left", and "lower right" in the correlation set. Taking Figure. 2 as an example, its corresponding expression in association sets is that "the car on the upper left" and "the pedestrian on the upper right".



Figure. 2 Example of an expression in correlation sets

To different target objects, the importance of object information to fire alarm judgment is different. In order to distinguish the significant degree of related information, we add weights to reflect the relationships between different target objects and corresponding contextual objects. When the influence is 0, the weight is 0, and when there is influence, the weight is 1. The artificially set weights are shown in Table 1, and will be continuously supplemented and improved during the experiment.

Table 1 Target objects and contextual objects' weights.

Contextual objects \ Target objects	Vehicle	Human	Bicycle	Buildings	Trees	Lawn
Fireworks	1	1	1	1	1	1	
Chimney	0	0	0	1	0	0	
Vehicle	1	0	0	0	0	0	
Censer	0	1	0	1	0	0	
.....							

Furthermore, multiple contextual objects may appear in the same scenario, and the corresponding parameters need to be weighted and summed. The weight is obtained through the neural network training. The specific neural network structure is shown in Figure. 3, wherein the input is the target object's category, the contextual objects' categories and positional relationship between them included in the picture containing smoke scene.

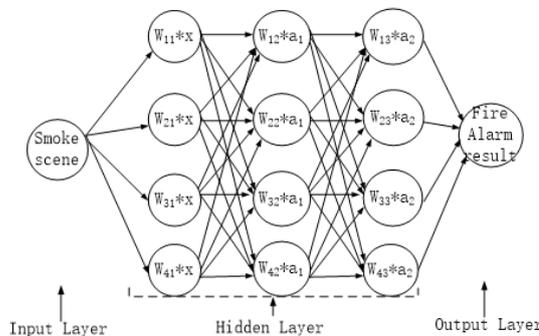


Figure 3 Training fire alarm results neural network.

3. Experimental results

In order to verify the feasibility of the proposed scheme, we collect a data set composed of fire smoke videos and pictures captured by online crawling, live shooting, and searching from databases. The videos processing method is to capture one image every ten frames. The data set includes 1,520

fireworks and firecracker scenes, 250 exhaust emission scenes for motor vehicles, 500 outdoor barbecue scenes, 1080 smoke emission scenes from chimneys, and 250 incense scenes in temples. A partial fireworks display in the database is shown in Figure. 4. Since the number of images in each part of the scene is different, we divide each scene's data set into a training set and a test set in a ratio of 8:2.



Figure. 4 Sample database.

In order to check the effect of semantic expression on fire warning in different scenarios, we defined fire different scenes of firecrackers, chimneys, campfire, motor vehicle exhausts, and incense burners as targets of Firework, Chimney, Campfire, Vehicles emissions, and Censer. And measure the early fire alarm effect by the accuracy rate p and the recall rate r . Their equation is as shown in formulas (1) and (2).

$$p = TP / (TP + FP) \quad (1)$$

$$r = TP / (TP + FN) \quad (2)$$

TP is the number of correct alarms output by the system, FP is the number of wrong alarms that the system outputs, and FN is the number of times the system should alarm but no alarm is generated. At the same time, in order to further prove the robustness and reliability of the method, we will use the fire alarm method depend on video image smoke detection^[10] to apply to the test set of this experiment. For comparison, the results are shown in Table 2, wherein this paper's method is marked as method1 and the smoke detection fire alarm method is marked as method2.

Table 2 Comparison of fire alarm accuracy and recall rate.

Test sets	Pictures No.	Method1 accuracy	Method2 accuracy	Method1 recall	Method2 recall
Fireworks	304	90.30%	50.00%	80.30%	97.80%
Chimney	216	98.10%	3.50%	Na	Na
Campfire	100	95.70%	47.60%	87.10%	98.20%
Vehicle	50	96.20%	5.70%	85.60%	87.10%
Censer	50	94.70%	7.90%	85.70%	98.50%
Average		94.10%	29.10%	82.70%	96.89%

As is shown in Table 2, method1 has the highest accuracy rate for the fire warning of Chimney, because the contextual objects of the chimney emission smoke scene is less likely to cause interference after image segmentation. Similarly, the accuracy rate of fire alarm corresponding to Firework is the lowest, because the fireworks and firecrackers are affected by the contextual objects which are difficult to accurately detect. It's important to note that Na appears in the table for the number of samples of labelled alarm is 0 in the smoke emission scene of the chimney, that is the denominator of p and r is 0. Secondly, when the comparison method is applied to the data set of this paper, the accuracy is lower while the recall rate is higher.

On the other hand, from the complexity of the algorithm and take the ratio of the total processing time to the total number of images as the measured value. The fire warning method proposed in this paper deals with the average duration of each picture is 0.5 seconds, and the average value of the comparison method is 0.57 seconds. In general, adding the logical judgment of contextual objects in smoke scenes effectively improves the accuracy of the fire alarm and ensures the detection efficiency.

4. Conclusion

In view of the problem that the image smoke detection method fail to confirm whether the smoke is caused by fire, this paper proposes an object-oriented semantic description of the fire warning method, which intelligently assists the monitoring personnel to determine the fire alarm result in a certain range around the smoke while monitoring whether the smoke is generated, and generate basic semantics to facilitate the reading and invoking of videos. Compared with the method based on video image smoke detection, the system's scene adaptability and early warning reliability are improved, which expands the application range of image smoke detection system, and can be popularized on the basis of urban outdoor electronic eye, improving the resource utilization.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61571241), the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (No. 15KJA510002 and 17KJB510043), Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX18_0273), the Ministry of Education-China Mobile Research Foundation, China (No. MCM20170205), the Six talent peaks project in Jiangsu Province (DZXX-008), the Research Foundation for Advanced Talents (no. NY217146).

References

- [1] JIWON C and CHANGICK K. Object-aware image thumbnailing using image classification and enhanced detection of ROI[J]. *Multimedia Tools and Applications*, 2016, 75(23): 16191-16207. doi: 10.1007/s11042-015-2926-5.
- [2] LIU Chang, ARMIN S and MARK L. Automatic generation of composite image descriptions[C]. 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, 2017: 2612-2618. doi: 10.1109/FSKD.2017.8393188.
- [3] KHITEM A, MOHAMED F, and IMED R F. Fuzzy hypergraph of concepts for semantic annotation of remotely sensed images[C]. 2017 International Conference on Advanced Technologies for Signal and Image Processing, Morocco, 2017: 1-8.
- [4] NIMA T, SURYAKANTH R. G, and LIANG Jianming. Automated polyp detection in colonoscopy videos using shape and context information[J]. *IEEE Transactions on Medical Imaging*, 2016, 35: 630-644. doi: 10.1109/TMI.2015.2487997.
- [5] FU Kun, JIN Junqi and CUI Runpeng. Aligning where to see and what to tell: image captioning with region-based attention and scene-specific contexts[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12). doi: 10.1109/TPAMI.2016.2642953.
- [6] AMIN Z, GENE C, and DINEI F. Joint denoising/compression of image contours via shape prior and context tree[J]. *IEEE Transactions on Image Processing*, 2018, 27: 3332-3344.
- [7] WEN Chenglin, ZHOU Guangfu, GAO Jingli, et al. Object recognition based on improved context model[J]. *Chinese Journal of Electronics*, 2018, 27(03): 573-581. doi:10.1049/cje.2018.03.014.
- [8] ZHANG Yao, CAI Hua, CUI Kai, et al. Research on spatio-temporal context learning based on Harris-Surf[C]. 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, 2017: 1-6. doi: 10.1109/CISP-BMEI.2017.8301915. doi:10.1109/CISP-BMEI.2017.8301915.
- [9] ZHANG Wei and WANG Shulin. Inference of cancer progression with probabilistic graphical model from cross-sectional mutation data[J]. *IEEE Access*, 2018, 6: 22889-22898.
- [10] EMMY P C, VINSLEY S, and SURESH S. Multi feature analysis of smoke in YUV color space for early forest fire detection[J]. *Fire Technology*, 2016, 52(5): 135-159.