

An improved YOLOv3-tiny method for fire detection in the construction industry

Jichao Li^{*1,a}, Shengyu Guo^{2,b}, Liulin Kong^{3,c}, Siqu Tan^{4,d}, Yican Yuan^{5,e}

¹ School of Economics and Management China University of Geosciences, CUG Wuhan, China

² School of Economics and Management China University of Geosciences, CUG Wuhan, China

³ School of Engineering China University of Geosciences, CUG Wuhan, China

⁴ School of Economics and Management China University of Geosciences, CUG Wuhan, China

⁵ School of Economics and Management China University of Geosciences, CUG Wuhan, China

Abstract. To prevent fire accidents on construction site and improve the accuracy of fire detection, an improved YOLOv3-tiny method (I-YOLOv3-tiny) is proposed in this paper. Although the YOLOv3-tiny has a fast detection speed and low equipment requirement, the accuracy is relatively low on fire detection. The improvement of the I-YOLOv3-tiny method is followed by three steps. Firstly, the feature extraction of fire images is enhanced by optimizing the network structure. Secondly, a multi-scale fusion is used to improve the detection effect of fire targets. Finally, the anchor boxes that are suitable for fire data sets are selected by k-means clustering. The results show that I-YOLOv3-tiny has an increased percentage of 4 on the mAP, the Recall rate has an increased percentage of 4, and AVG IOU has an increased percentage of 6. The proposed model meets the real-time performance of fire detection. This study is of theoretical and practical significance on fire safety management and accident prevention in the construction industry.

1 Introduction

Construction sites are extreme hazardous due to its dynamic, temporary and decentralized nature [1]. Among the numerous risks and accidents that construction workers face, the fire is one of the most dangerous hazards and the leading fatal cause. In 2019, more than 200 thousand fires broke out in China, leading 1335 fatalities and 837 injuries. Among them, 2,926 fires occurred in construction sites, an increase of 5.8 percent year-on-year [2]. For instance, on November 6, 2020, a fire accident occurred at a construction site in Xiaogan city. An electric welding worker violated regulations and caused a spark to ignite auxiliary building materials. Therefore, it is extremely necessary to carry out a real-time detection and early warning of fire accidents.

Fire detection equipment mainly determines whether the fire occurs according to light, temperature, smoke and gas [3]. The prediction accuracy is relatively low and it is not suitable for long-distance monitoring in a large field of vision. Compared with the traditional fire detection equipment, fire detection based on computer vision and convolutional neural network algorithms does not require fire information such as temperature. Convolutional neural networks can automatically extract target features and conduct classification and recognition to avoid the steps of manual feature extractions. The networks have been widely used in forest fire [4], tunnel fire [5] and other types of fire protection.

The YOLOv3-tiny model is a simplified version of YOLOv3 [6], which does not occupy a large amount of memory and is suitable for deployment on low-end machines or mobiles to achieve target detection tasks. However, this model cannot fully extract the features of fire images because of the shallow number of network layers. In addition, the detection effect on small fire targets is poor. Therefore, an improved YOLOv3-tiny (I-YOLOv3-tiny) method is proposed based on the YOLOv3-tiny model for fire detection on construction sites. The proposed method improves the accuracy of fire detection by three steps as: (i) optimizing the network structure; (ii) multi-scale predicting; (iii) k-means clustering.

The rest of this paper is organized as follows. Section 2 briefly introduces the YOLOv3-tiny model. Section 3 describes the I-YOLOv3-tiny method. Section 4 illustrates the experimental process to verify the feasibility and effectiveness of the proposed method. Section 5 concludes and discusses the limitations of the study and potential future work.

2 YOLOv3-tiny model

The YOLOv3-Tiny model has 13 convolution layers and 6 max-pooling layers. The feature map is subsampled by the maximum pooling method, and the stride of the first five max-pooling layers is 2, and the stride of the last max-pooling layer is 1. Because the shallow network is easier to realize positioning and the deep network is

^{a*} Corresponding author: Jichao.L19@cug.edu.cn

^be-mail: guoshy@cug.edu.cn ^ce-mail: kongliulin@cug.edu.cn ^de-mail: tansiqi@cug.edu.cn

^ee-mail: 20181002062@cug.edu.cn

easier to obtain the rich semantic information of the target, the method of feature fusion is adopted to carry out multi-scale prediction. The input image size of the model is 416×416, and the output size is 13×13 feature map after 5 max-pooling layers with stride of 2, which is suitable for detecting large-sized targets. Additionally, the feature map of 13×13 is upsampled and fused with the convolutional layer of 26×26 to achieve the second-scale prediction, which is suitable for the detection of medium-sized targets.

3 The I-YOLOv3-tiny algorithm model

3.1 Optimization of the network structure

Max-pooling can reduce the image size and extract the key information. However, the pooling window only retains the maximum characteristic information of fires, which results in a partial loss of fire characteristic information. Particularly, the complex background environment of the construction site interferes with fire detection. Therefore, the max-pooling layer is replaced by a convolutional layer with a size of 3×3 and a stride size of 2. Compared with pooling, convolution retains more fire feature information and enhances the feature extraction capability of the network. In addition, the 11th max-pooling layer and the 12th convolutional layer are removed from the YOLOv3-tiny network, and three convolutional set are added to improve the detection performance of the model. The I-YOLOv3-tiny network structure is shown in Fig. 1.

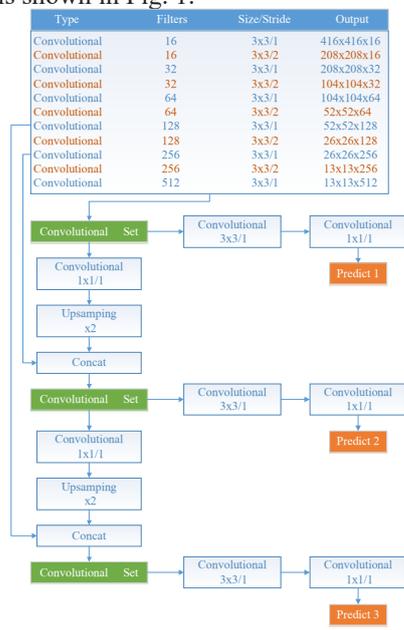


Figure 1. The network structure of I-YOLOv3-tiny algorithm

3.2 Multi-scale prediction based on feature fusion

In the convolutional neural network, the shallow layer contains less feature semantic information, and it is easier to locate the target position. In the deep network, the feature semantic information is rich, while the target

position information is rough. To make full use of the shallow feature information and improve the detection effect of small targets, a scale prediction is added on the basis of the original model prediction by referring to the idea of multi-scale module of feature pyramid networks [7]. Since it is necessary to ensure that the output size of the feature map is consistent in the multi-scale feature fusion, the deep convolutional layer is upsampled twice and concatenated with the shallow convolutional layer to form three scale predictions of 13×13, 26×26, 52×52, which are used to detect large, medium and small fire targets respectively.

3.3 K-means cluster analysis

The series of YOLO algorithms draw on the idea and mechanism of anchor boxes in Faster R-CNN [8], and determine the number and size of anchor by clustering data sets. The appropriate anchor can improve the accuracy and speed of detection tasks. However, the YOLOv3-tiny algorithm obtains 6 anchors based on COCO data clustering, averaging to 2 different scale predictions. The original data set is rich in target categories, and the obtained anchor boxes have generalization. The characteristics make this algorithm not suitable for the fire detection scenes on site since accurate target box information is difficult to be obtained. Therefore, k-means cluster analysis is performed on the fire data sets to re-determine the number and the corresponding size of anchor box. The distance function in clustering is:

$$d(\text{box}, \text{centroid}) = 1 - \text{IOU}(\text{box}, \text{centroid})$$

Where, the centroid represents the center of the cluster, the box represents the sample, and $\text{IOU}(\text{box}, \text{centroid})$ is the intersection over union of the cluster center box and the cluster box.

For this study, $k = 1 \sim 12$ are selected to obtain the relationship between the number of clustering centers k and average intersection over union (AVG IOU), as shown in Fig.2. With the increase of k value, AVG IOU increases gradually. When $k > 9$, the increase of the curve gradually becomes flat. 9 clustering centers are selected by balancing the cost of accuracy and speed. The corresponding 9 anchors are (21,32), (48,63), (63,119), (84,218), (119,76), (132,151), (157,272), (243,167), (289,317) respectively, which are evenly distributed to 3 scale predictions.

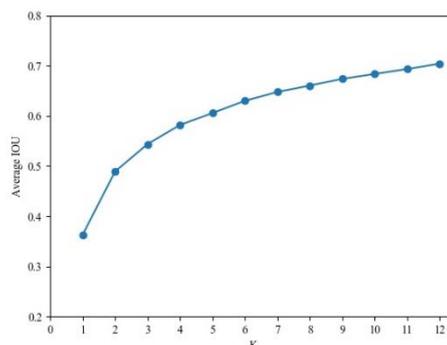


Figure 2. K-means clustering results

4 Experiments and results

4.1 Experimental environment and data sets

The experimental platform is shown in Table 1. The data set consists of fire videos and images from construction sites and other fire scenes such as forest fires on the Internet. To realize the training of the I-YOLOv3-tiny algorithm model, fire images are extracted from the videos according to the interval. The data set is expanded by means of mirroring and a total of 8050 images are obtained, in which 7245 images are randomly selected to form the training set and the remaining 805 images to form the test set. Then, the LabelImg box software is used to annotate the data set to generate a .xml file with the corresponding category label fire.

Table 1 The experimental platform

Name	Configuration
Operating system	Ubuntu 18.04
Central processing unit (CPU)	Intel Core i7-9750
Random access memory (RAM)	8GB
Graphics Processing Unit (GPU)	GTX1650
GPU Acceleration library	CUDA10.2, CUDNN7.6.5
Deep learning framework	Darknet
Programming language	C, Python3.7

4.2 Experimental training process

The YOLOv3-tiny algorithm and the I-YOLOv3-tiny algorithm are used to train on the fire data set, respectively. In view of the limitations of the hardware and software of the experimental platform, the network training parameters are set: batch is 64, subdivisions is 8, the initial learning rate of the model is 0.001, the momentum coefficient is 0.9, the maximum iterations number is 50020, and the learning rate is multiplied by 0.1 after 40000 and 45000 iterations. During the training, the input size of the model is changed every 10 iterations to grantee the model has a well detection effect for images of different sizes. The average loss values of the two networks after training are shown in Fig.3. With the increase of the number of iterations, the average loss value of the I-YOLOv3-tiny algorithm is significantly lower than that of the YOLOv3-tiny algorithm, which indicates that the I-YOLOv3-tiny algorithm has a better learning ability.

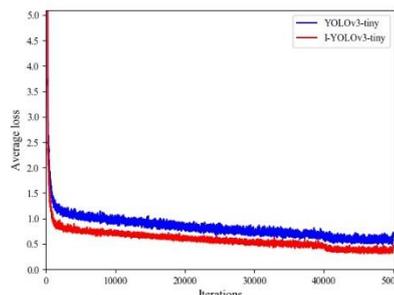


Figure 3. Average loss curve

4.3 Experimental results

To test the performance of the I-YOLOv3-tiny algorithm model, this paper uses the average precision rate (mAP), the Recall rate, and the AVG IOU as the evaluation indexes. The Recall represents the proportion of the detected correct samples in the samples to be tested. The higher Recall is, the less the missed detection is. The test results are shown in Table 2. Compared with YOLOv3-tiny, I-YOLOv3-tiny has an increased percentage of 4 on the mAP, which indicates that the detection accuracy has been improved. In addition, the Recall rate has an increased percentage of 4, which indicates that the improved model has a low missed detection. Furthermore, AVG IOU has an increased percentage of 6, which indicates that the improved model generates a higher overlap rate of prediction box and ground truth box, and the fire position is more accurate.

Table 2 The experimental test results

Model	mAP	Recall	AVG IOU
YOLOv3-tiny	0.88	0.83	0.67
I-YOLOv3-tiny	0.92	0.87	0.73

To test the detection speed of the I-YOLOv3-tiny algorithm model, a fire scene on-site is simulated, and the video stream resolution is 1280×720. The results show that the detection speed of the I-YOLOv3-tiny model is an average of 63.8 frames per second, which meets the real-time performance. Additionally, two groups of detection images in the video are selected for display, as shown in Fig.4, where (a) represents the original image, (b) represents the YOLOv3-tiny detection, and (c) represents the I-YOLOv3-tiny detection. It can be found that I-YOLOv3-tiny can locate the fire position well and detect small fire to prove this model has better performance.



(a) Original image

(b) YOLOv3-tiny detection

(c) I-YOLOv3-tiny detection

Figure 4. Test results

5 Conclusion

Construction is one of the most dangerous industries. The I-YOLOv3-tiny method for fire detection is proposed to improve construction safety and prevent accidents. The accuracy of fire detection at construction sites is improved through network structure optimization, multi-scale feature fusion and k-means dimensional clustering. The proposed method is suitable for deployment to low-end machines or mobile devices. From a theoretical aspect, this method enriches convolution network studies and provides a solution on real-time fire detection at construction sites. From a practical aspect, this method has a well robustness and could be potentially applied on video monitoring on sites to reduce fire accidents.

However, there are still some shortcomings, such as a small number of fire data sets, and the detection speed requires to be further improved. In the future, high-quality sample data sets need to be added, and the detection speed should be improved based on the accuracy rate. In addition, this method is considered to be integrated into cameras on mobile robots for fire-safety inspections on sites.

Acknowledgment

This work was supported by the National Natural Science Foundation of China [grant number 71801197]; the Fundamental Research Funds for National University, China University of Geosciences (Wuhan) [grant number 1910491C06].

References

1. H. Li, M. Lu, S.-C. Hsu, M. Gray, and T. Huang, "Proactive behavior-based safety management for

construction safety improvement," *Safety science*, vol. 75, pp. 107-117, 2015.

2. Fire and Rescue Department Ministry of Emergency Management. (2019). National fire reports in 2019. Available: <https://www.119.gov.cn/article/3xBEjR54K>
3. G. Lin, Y. Zhang, G. Xu, and Q. Zhang, "Smoke detection on video sequences using 3D convolutional neural networks," *Fire Technology*, vol. 55, pp. 1827-1847, 2019.
4. Z. Jiao, Y. Zhang, J. Xin, L. Mu, Y. Yi, H. Liu, et al., "A deep learning based forest fire detection approach using UAV and YOLOv3," in 2019 1st International Conference on Industrial Artificial Intelligence (IAI), 2019, pp. 1-5.
5. K. B. Lee and H. S. Shin, "An application of a deep learning algorithm for automatic detection of unexpected accidents under bad CCTV monitoring conditions in tunnels," in 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), 2019, pp. 7-11.
6. J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.
7. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2117-2125.
8. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, pp. 1137-1149, 2016.