

Personnel identification and distribution density analysis of subway station based on convolution neural network

Lei Wang^{1,2}, Kun Zhang¹, Jian Kang², Meng Peng¹, and Jiayu Xu^{1,*}

¹Tsinghua University, School of Environment, Beijing, China

²Research Institute of Chemical Defence, State Key Laboratory of NBC Protection for Civilian, Beijing, China

Abstract. In this paper, a method based on convolution neural network and multi-camera fusion is proposed to improve the recognition accuracy of crowd and then the personnel distribution of subway station platform is analyzed. In this method, tensorflow is used as the deep learning training framework and the yolov4 neural network algorithm is used to identify the subway station platform area using three videos synchronously. Through affine transformation and time average statistics, the passenger density of each sub-area is calculated and the distribution of personnel density in the whole area is analyzed. The results show that the number of people recognized by multiple cameras is 58% higher than that by single camera. The new recognition method has high recognition rate for the actual scene with large crowd and more obstacles. Finally some areas with high risk of personnel aggregation have been found, which should be the focus of safety monitoring.

1 Introduction

In recent years, the development and utilization demand for underground space has grown increasingly due to the acceleration of urbanization. Due to narrow space and high personnel density, public safety events may occur easily in subway stations and the like [1-2]. It is possible to determine high-risk areas and deploy monitoring or emergency devices in advance to provide support for evacuation or rescue and greatly reduce accident loss if the number and distribution of personnel can be obtained accurately and personnel density rules can be studied.

At the present stage, studies about personnel distribution density mainly focus on investigation and regional measurement. In 2016, Zhou Jingjing et al firstly proposed the warning mechanism for public places with high personnel density and calculated personnel density in major areas of Ditan Park through regional measurement and positioning data of mobile phones [3]. Liu Yong et al obtained the personnel density in commercial complex with the black-box theory model based on investigation data [4]. The current studies about personnel in subway stations mainly focus on analyzing passenger flow volume during

* Corresponding author: jiayu_tsinghua@163.com

operation and population characteristics during an emergency evacuation. In 2020, Ju Yi et al created the simulation model for evacuation of high-density crowds, conducted computer simulation, studied evacuation rules of high-density crowds, analyzed specific behaviors of crowds after emergency and optimized the evacuation path[5]. In 2021, Feng Qilin et al studied diffusion rules of pollutants through computer simulation, collected a number of personnel from different evacuation paths based on simulation results, and then used the diffusion concentration of pollutants for describing the harm of poisonous gas on personnel[6]. For the personnel evacuation process, the current studies mostly focus on emergency processes after optimization events. Besides, it is possible to predict high-risk areas and deploy monitoring in advance based on the description of personnel distribution before the event. However, there is not any report now about personnel distribution under normal working conditions of subway stations.

With its rapid development, deep learning technology has obtained huge progress in target detection, recognition, and tracking, and has been used in daily life [7-9]. The current image recognition technology mainly includes target detection algorithm based on area suggestion and area classification (e.g. R-CNN[10], Fast R-CNN[11], Mask R-CNN[12]) and target detection method based on regression (e.g. YOLO[13], SSD[14]). In 2013, Yang Zhaojun et al introduced the artificial intelligence (AI) method which selected camera video as a processing object to study the human head extraction and tracking technology [15]. In 2019, Chen Xiao introduced the deep learning algorithm related technologies, conducted human detection and calculated a final number of people with the target detection method, proposed the dilated residual model of multi-scale detection for the large difference in personnel scale and personnel shielding, created extraction network for scale-specific characteristics, and improved accuracy rate of detection [16]. In the same year, Shen Shoujuan improved the accuracy rate of student detection and counting in the classroom with an improved YOLOv3 algorithm [17].

The current image recognition and detection methods mainly focus on improving the accuracy rate of identifying a single image. However, these methods have not been realized in subway stations because they cannot meet actual demands under complex scenarios and the shielding of obstacles. This paper identifies the personnel object for multiple on-site monitoring images with the machine vision and target detection algorithm YOLOv4, obtains the number and position of passengers, and proposes the statistical method of personnel distribution density, then compares and analyzes personnel distribution at a different operating time in a subway station of Beijing with the proposed method.

2 Model training methods

2.1 Creation of dataset

This paper selects a station platform of Line 15 in Beijing as the object. Due to the simple scenario and high density of people, this paper creates a dataset with screenshots of on-site monitoring video for training to improve the recognition rate. Finally, this paper selects video files on Monday from cameras at 10 different positions, then randomly selects 200 screenshots at a different time with OpenCV, and finally keeps 885 screenshots as samples of typical operating conditions of monitoring video in the subway station after manual filtration for incomplete and repeated pedestrian information. The dataset includes most of the scenarios and pedestrian conditions and therefore can meet the training demand of the model. All samples are annotated manually with LabelImg, where at least one label and at most 66 labels are provided. Each image is provided with 19.56 labels on average and there are 17310 labels in total. All labels are "person".

1.2 Model training

The model training is combined with pre-training and dataset training. The VOC2007 dataset for pre-training includes 20 categories and 9963 annotated images, with a total of 24,640 annotated objects. In the process of training, the pre-training model can accelerate the convergence rate of gradient descent [18-19], reduce vanishing gradient or gradient explosion problems, and break symmetry to ensure that various hidden units can learn different information. The deep learning framework is tensorflow1.15, the OS system is ubuntu18.04, and the CUDA version is v10.0. For each iterative training, the sample size is set to 10, the maximum number of iterations is set to 100, and “early_stopping” is set to early termination after continuous iterations for 6 times due to lack of significant improvement. Finally, the program is terminated in advance at the 78th Epoch and “loss” is kept at about 130.

2 Personnel identification and calculation of distribution density

2.1 Area setting and personnel identification

After the test, the trained YOLOv4 model can effectively identify figures in the image with a high accuracy rate but cannot identify those figures at the edge of the image or the back of the column. There is a certain gap between the obtained total number of people and actual value. As shown in Fig. 1, to meet the actual demands of the subway station platform, this paper invokes three groups of camera images at the same time which cover the north, central and south platform space with the boundary of two rows of columns to effectively avoid shelter of the column.

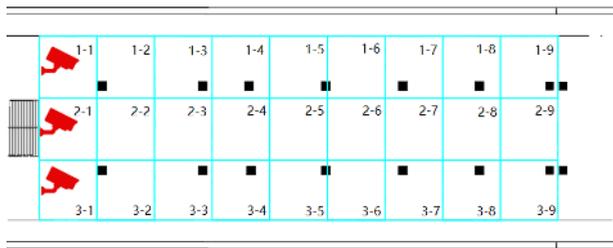


Fig. 1. The scheme of cameras and sub-aeras of subway station platform.

2.2 Distribution density model

The proposed personnel distribution density[21] is a time-homogenous concept, i.e., the specific value between the total number of people who appeared in a certain time τ in area i and the total number of people in the whole computational domain, with expression as below:

$$D_i = \frac{d_i}{\sum_{i=1}^m d_i} \quad (1)$$

$$d_i = \sum_{j=1}^x P_{i,j} \quad (2)$$

where, m is the total number of sub-grids divided on the ground of computational domain, of which the quantity determines the degree of evaluation. d_i is the total number of people who appeared at a certain time τ in area i ; $P_{i,j}$ is the number of people who appeared at the

j th frame in area i ; x is the number of video frames in a certain time τ . According to the definition, $\sum D_i = 1$ is true.

When τ and m are determined, the center of figure rectangle recognized with YOLOv4 can be projected to the sub-grid on the ground with PNPoly algorithm to count people $P_{i,j}$ in the i th sub-grid at the j th frame and further calculate personnel distribution density D_i in area i with Formula (1) & (2)[20]. As shown in Fig. 1, this paper divides the subway station platform into 27 sub-areas, studies personnel distribution within 10 min during morning and evening peak (i.e. $m=27$ & $\tau=10$ min), and extracts a 600-frame image with equivalent interval as the recognition object as per the sampling rate 1 fps.

3 Results and discussion

Fig.2 shows the identification result of the monitoring video. The train interval of the target line during morning and evening peak is 5 min, and the τ value covers arrivals and departures of two trains on both sides so that there are two peaks in personnel counting. Fig. 2(a) shows passenger flow on the north side of the platform. The train on the north side stops steadily at the time 20s and the number of passengers increases from 10 to 45, and then the passenger crowds flow to the center of the platform. From the exit of the escalator, the number of passengers slowly decreases to 10 and below at over 140s. It is the waiting time at 150-420 s. This is the last stop so that there is a slight change in the number of waiting passengers on the north side. The recognition results can quantitatively describe actual conditions. As the train arrives, the number of recognized passengers tends to increase fast and then decrease slowly within over 120s. Fig. 2(b) shows the change in the number of passengers in the center of the platform. From the time 90s, the number of passengers in the north side starts to increase. The 2nd peak occurs at 150s due to the arrival of the train. Fig. 2(c) shows the waiting area on the south side of the platform. The results indicate that there is a small number of passengers getting off in the direction, but the number of waiting passengers tends to increase slowly.

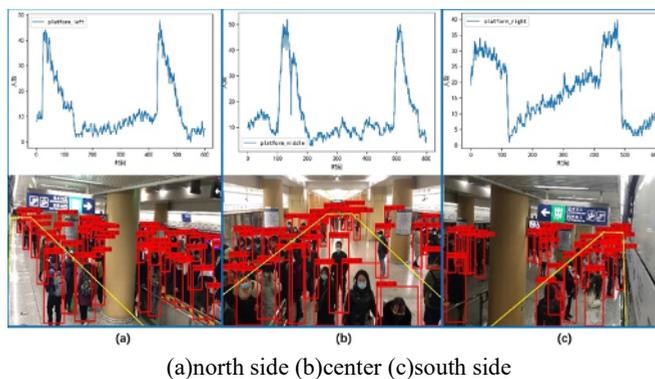


Fig. 2. The personnel identification of subway station platform.

The single-video recognition method analyzes and counts passengers in the whole area with a camera in the center of the platform. The multi-video fusion recognition technology analyzes the north-side, central and south-side area with 3 monitoring videos and obtains the total number of passengers in the whole area after calculation. The results indicate that the above two methods have a consistent accuracy rate, because cameras cover the central area, and the space is wide and not shielded. However, the accuracy rate is decreased greatly when the single camera in the center is used for the south and north side due to angle, figure and column shielding. Under such a scenario, the accuracy rate of multi-video

fusion recognition can increase 36% and 55.6%, respectively. To sum up, the multi-video recognition technology divides the subway station platform into three areas and analyzes them independently with different videos to avoid the influence of shielding and depth of field. Besides, the accuracy rate of recognition increases from 52.5% for the single video to 83.3%, which is closer to actual conditions.

After statistics, the interval of the subway line is 5-8 min. To explore personnel distribution density under different τ value, τ is taken as 1 min, 2 min, 5 min, 7 min, 10 min, 15 min, 20 min and 25 min, respectively to compare the influence of different time span τ on personnel distribution density. Fig. 3 shows the results. If $\tau = 1$ min is true, Area 3-9 is the high-risk area in the southeast corner of the subway station platform. With the increase of τ , the density commonly increases for the south-side area and continuously decreases for the north-side area. This indicates that more passengers wait for the train on the south side and stay for a longer time at the platform; there are more outbound passengers in the north side, and they usually stay here for a short time, so that the relative density is low. If $\tau = 7$ min is true, the overall density distribution of the platform has been formed basically; the personnel density in the south side is high, and the relative density for Area (3-1) in the west side and near the stairway is maximum. If $\tau \geq 10$ min is true, overall outbound and inbound conditions of trains on both sides are covered in the time span, and time-homogeneous distribution features of personnel tend to be consistent so that density value for each area is relatively stable.

Fig. 4 shows the personnel distribution density of the subway station platform during morning and evening peaks (7:00-9:00). The results indicate that passengers mainly wait for the train on the south side of the platform and stay for a long time; the personnel density here is 2-10 times of the north side; the personnel density in Area 1-1 and Area 2-1 near the stairway reaches 0.08 and above. During morning and evening peaks, there are more outbound passengers on the north side with a high flow rate, and passengers usually stay here for a short time. All passengers have to pass through the central stairway exit 1-2, where the flow rate and the number of passengers are large. Therefore, the calculated density is high, i.e. The public seats are set in the Area D = 0.059. 9-2 and passengers can have a rest here so that the density value is relatively high.

Under the same conditions, the personnel density distributions during morning and evening peak, in the morning and in the evening are collected, and then density values in the south side area 1-1, south side area 2-1, central stairway area 1-2, north side area 1-3, north side area 3-3, and north side area 8-3 are compared, respectively. Fig. 5 shows the results. In the full-day span, the personnel density for Area 1-2 & 2-1 is maximum; the personnel density for Area 1-2 keeps at 0.05-0.06, followed by Area 1-3; and the personnel density for Area 3-3 is minimum. Therefore, personnel distribution density in the subway station is regular. In the full-day scale, the ranking of density value for each position is basically stable, so that the index can be used for effectively distinguishing high-risk and low-risk areas. The subway station is taken as an example. The personnel distribution density indicates that Areas 1-1 & 2-1 on the north side of the platform and near the stairway are high-risk areas where crowding, stampede or large-area poisoning may occur in case of any unexpected safety accident as the high-density node. Therefore, it is necessary to strengthen passenger guidance and monitoring in the above areas, set emergency evacuation paths based on characteristics of the platform, and assign more safety supervisors to minimize the risk and harm.

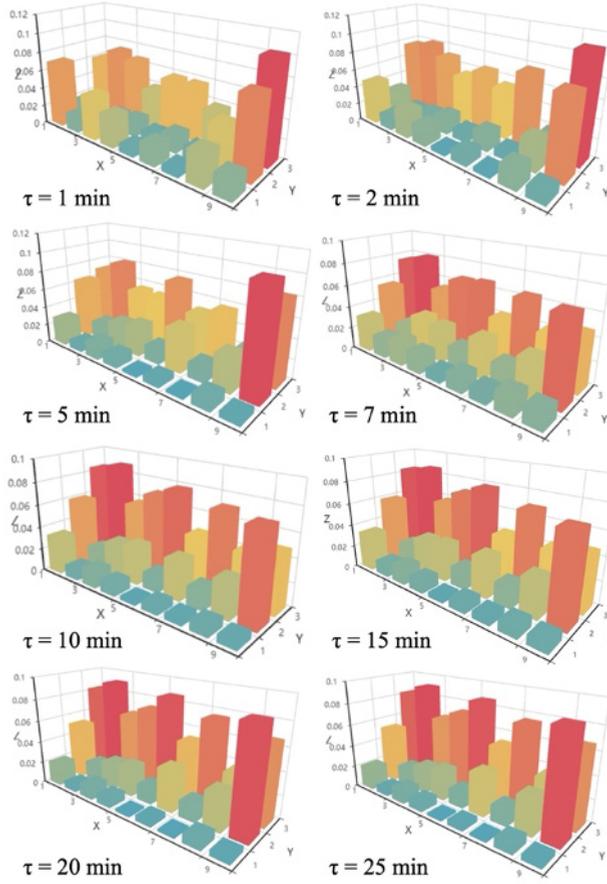


Fig. 3. Personnel distribution density of subway station platform under different τ .

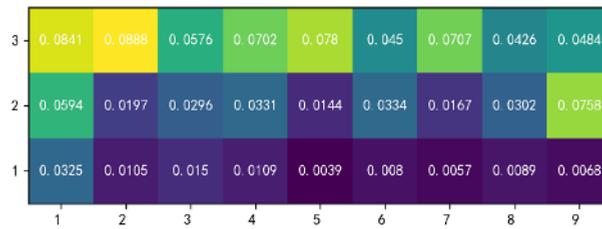


Fig. 4. Personnel distribution density of subway station platform during morning peak ($\tau = 10$ min).

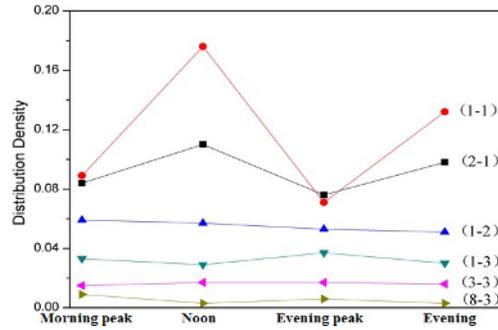


Fig. 5. Personnel density of typical sub-aeras in different times($\tau = 10$ min).

4 Conclusions

The YOLOv4-based multi-video synchronization and identification method shows a high accuracy rate for large space or shielded personnel recognition and statistics. Compared to a single video, the method can identify and collect personnel count in subway stations.

The proposed personnel distribution density is a time-homogenous evaluation index. If τ is small, the information will be incomplete, and the result will be one-sided. If $\tau > 7$ min is true, all information about arrivals and departures of trains on both sides and related passengers will be covered in the time span. If τ continues to increase, the density distribution tends to be stable.

The analysis results for a typical time period of a day indicate that the high-density area during morning and evening peak is located at the side where passengers stay for a longer time; the number of passengers is small during the non-peak period, and passengers usually stay at the area. As a result, the density value will be further increased. Therefore, the monitoring and evacuation solution shall focus more on the area where passengers stay for a longer time at the subway station platform. It is possible to fast and accurately determine the area with high personnel distribution density in the subway station platform to provide data support for early prevention and control.

References

1. Jonathan S., Markus B., Nils D., Propagation of Tracer Gas in a Subway Station Controlled by Natural Ventilation[J]. Journal of Heat Island Institute International, 2014:9-2
2. Feng qilin, Yang Yibin, Deng Ye, Cai Hao, Jiang Mingrui, Lu Jingyu, Zhang Canxin, Zhang Boyuan. Multi-robot Source Positioning Experiment in the indoor Time-Varying Flow Field And Numerical Verification [J/OL]. Environmental Engineer:1-12[2021-02-25].
3. Zhou Jingjing, Liu Dongdong. Research on Dense Personnel Risk Monitoring And Early Warning Mechanism – taking Beijing Ditan Temple Fair in 2016 as an example. China Civil Engineering Society, 2016:9.
4. Liu Yong, Xu Zhisheng, Zhao Wangda. Personnel Density Survey And Analysis for Commercial Complex [J]. Journal of Safety Science and Technology, 2015,11(05):175-180.

5. Ju Yi. Simulation Research on High-density Personnel Evacuation Based on Intelligent Agent Model [D]. China University of Mining and Technology,2020.
6. Feng Qilin, Yang Yibin, Deng Ye, Cai Hao, Jiang Mingrui, Lu Jingyu, Zhang Canxin, Zhang Boyuan. Multi-robot Source Positioning Experiment in the Indoor Time-Varying Flow Field And Numerical Verification [J]. Environmental Engineer, 2021,39(02):73-81.
7. Yang Jian, Chen Shi'e. Application of Smoke Detection in Family Fire Prevention Based on Image Processing Technology [J]. Computer Knowledge and Technology,2021,17 (06):189-191.
8. Chen Tao, Lu Hong. Research on Pedestrian Detection Technology Based on Deep Learning [J]. Software Guide,2021,20(01):76-80.
9. Wang Zilong. The Key Technology of Abnormal Behavior Detection In Family Video Monitoring Environment [D]. Nanjing University of Posts and Telecommunications,2020.
10. Zhang Juan, Mao Xiaobo, Chen Tiejun. A Research Overview of Moving Target Tracking Algorithm [J]. Application Research of Computers,2009,26(12):4407-4410.
11. Ross Girshick. Fast R-CNN,2015, arXiv:1504.08083
12. Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick. Mask R-CNN,2018, arXiv:1703.06870
13. Redmon J , Divvala S , Girshick R , et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Computer Vision & Pattern Recognition. IEEE, 2016.
14. Liu W , Anguelov D , Erhan D , et al. SSD: Single Shot MultiBox Detector[J]. 2015.
15. Yang Zhaojun. Research on improved algorithm based on people counting and tracking in video and its realization [D]. Nanjing University of Posts and Telecommunications,2013.
16. Chen Xiao. Research on people counting algorithm in video based on target detection [D]. Harbin Institute of Technology,2019.
17. Xiao Wei, Lu Jinkang, Li Boshen, Wu Qishuo, Bai Yingdong, Pan Chao. Optimization of real-time people counting detection based on Faster RCNN [J]. Journal of Changchun University,2020,41(04):369-374.
18. Shen Shoujuan, Zheng Guanghao, Peng Yixuan, Wang Zhanqing. Detection of students in classroom and counting method based on YOLOv3 algorithm [J]. Software Guide,2020,19(09):78-83.
19. Li Jinming, Qu Yi, Pei Yuhao, Yi Zejiang. Slightly-adjusted pedestrian re-identification based on pre-trained convolutional neural network model [J]. Computer Engineering and Applications,2018,54(20):219-222+229.
20. Wang Manli. Research on image recognition algorithm based on convolutional neural network (CNN) [D]. Beijing University of Civil Engineering and Architecture,2020.
21. Jiang Yan, Li Xuesong, Fu Xiandi. Fault representation of ultra-high water-content old oil field and potential of high-efficient digging for surplus oil [J]. Petroleum Geology & Oilfield Development in Daqing,2019,38(05):246-253.