Journal of Artificial Intelligence Practice (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2022.050201
ISSN 2371-8412 Vol. 5 Num. 2

Rail Transit Passenger Counter based on TOF

Hongqing Feng[a,*], Zundong Zhang[b]

North China University of Technology, Beijing, China
[a]f18236901857@163.com, [b]zdzhang@ncut.edu.cn
**Corresponding author*

Keywords: TOF Camera, Subway Passenger Counter, SORT, MobileNet-SSD

Abstract: In some developed countries, passenger counter has become the standard product of rail transit industry. This product has irreplaceable significance for the statistics of rail transit passenger flow, and clear passenger flow data is the data basis for rail transit intelligent operation and efficient management. Therefore, based on the hardware foundation of TOF fisheye camera, we built an environment simulating subway door in the laboratory, trained mobilenet SSD target detection model with convolutional neural network framework Caffe, and realized reasoning with sort multi-target tracking algorithm. Finally, through experimental test, the recognition accuracy is as high as 98%, which can be directly put into market after transplantation.

1. Introduction

1.1. Background and Necessity

With the development of society, the construction and use of subway has been popularized in many provinces and cities in China. The passenger flow is counted through the entrance gate of subway station, and the data practicability is not high. By mastering the real-time passenger flow data, obtaining the passenger flow information of subway channel and platform in time and accurately, and flexibly adjusting the passenger transport organization of the transfer station, it can provide the basis for scientific decision-making for passenger flow dredging and flow restriction at the entrances and exits of the transfer station and the transfer channel, and scientifically formulating the emergency evacuation plan. Therefore, it is necessary to build an automatic rail transit passenger flow monitoring and counting system that can meet the dispatching operation. This paper is based on this problem.

1.2. Status of Passenger Counter

Bus card counting: through the system design of the card and data interface, we can obtain the data of passengers' boarding and alighting time and corresponding stations. Because the collection process of bus card information is little affected by human factors, the collected data is true and reliable, and can accurately reflect the temporal and spatial distribution characteristics of urban bus travel. However, due to the general practice of swiping the card when getting on the bus and not swiping the card twice when getting off the bus, the number of people getting off the bus, the distribution of starting and ending points and other information cannot be analyzed and speculated according to the information obtained from the card information; With the rapid development of economy, there are more and more floating population in the city, which leads to many people can not use the card, which makes the accuracy of collecting passenger count with the card very poor [1].

Automatic passenger counting: automatic passenger counting is an effective method to automatically collect the time and place of passengers getting on and off, combined with vehicle automatic positioning, wireless information transmission and other technologies. It is a necessary part of the intelligent public transport system and needs to be used in coordination with other systems, with high requirements for cost and operation environment [2].

With the development of image-based passenger counting technology, a series of new methods based on video passenger counting technology are proposed to solve the problem of more abundant passenger information [3]. The working principle of this technology is to obtain the passenger flow video image through the camera installed above the entrance and exit and key areas, cooperate with the computer to recognize the moving target, and count the passenger flow accordingly.

2. Development of Rail Transit Passenger Counter based on TOF

2.1. Overview#shown Figure 1

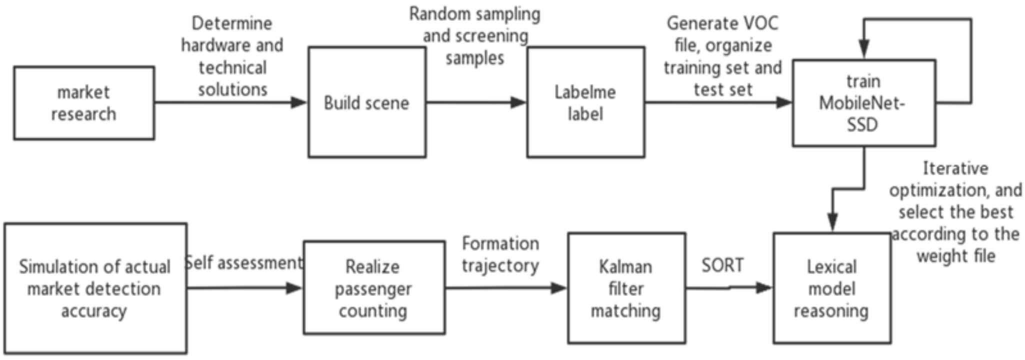

Figure 1: Development process

2.2. Market Research

2.2.1. Research Status of Foreign Passenger Flow Forecasting

Based on the utility maximization theory, the discrete choice model was first proposed by domencich in 1975. The core of the model is the prebit model family and the logit model family, in which the multivariate logit model is still widely used [4]. Eric J Miller introduced the development of foreign rail transit passenger flow forecasting models in detail, and clarified that the existing passenger flow forecasting models basically include three core parts: overall demand model, mode division model and distribution model [5]. At present, nestedlogit model is used in most cases. The logistet model is used to predict the accessibility of public transport in both the US and the US.

2.2.2. Research Status of Domestic Passenger Flow Forecasting

In 2015, Yan Yan and Ni Shaoquan analyzed the principle and formula of nestedlogit model, took the travel behavior of passengers in urban rail transit network as the research object, divided the passenger selection process into two levels: transfer times selection and transfer path selection, and established nestedlogit model for passenger flow distribution of urban rail transit [6].

In 2016, Wang Bo and Huang Jianling carried out cluster analysis on stations in combination with the characteristics of historical passenger flow, predicted the inbound and outbound volume, OD distribution and passenger flow distribution of new stations and existing stations under the condition of new line network, and carried out passenger flow prediction and line operation evaluation by taking four new lines opened by Beijing Metro at the end of 2014 as examples to study the impact of new lines on existing lines [7].

As far as Beijing alone is concerned, the operation scale of rail transit has reached 14 lines, 336 kilometers and 146 stations, including 19 transfer stations. By 2015, the number of transfer stations in Beijing will reach 52, and the daily passenger volume will exceed 8 million. At present, the maximum passenger flow of Beijing rail transit is about 220000 per station per day, and the passenger flow of a single transfer channel reaches 12000 per hour during the peak period. However, due to the large construction time span of Beijing rail transit lines, the design of many transfer stations is limited by existing stations, with long transfer channels, insufficient transfer capacity and prominent potential safety hazards. It is difficult to judge whether measures such as current restriction, evacuation and train abandonment should be taken according to the passenger flow status of transfer channels, entrances and exits [8].

2.3. Hardware Selection

After investigation, it can be found that the accuracy of using binocular passenger counter or video detection to determine the passenger flow is very unsatisfactory. The colorful background environment, the dense crowd sheltering each other, and the uneven height of passengers are all inevitable influencing factors. Therefore, we associate it with the somatosensory game in the Xbox game console. Under the complex background environment, the game console can still accurately identify the player's location, operation actions, and even small movements back and forth. After a lot of information inquiry, we learned about Kinect sensor and the light coding technology behind it, and further locked in TOF (time of flight), a more mature technology with similar principle. Finally, we purchased the TOF depth detection chip of melexis brand for development. As shown in Figure 2.

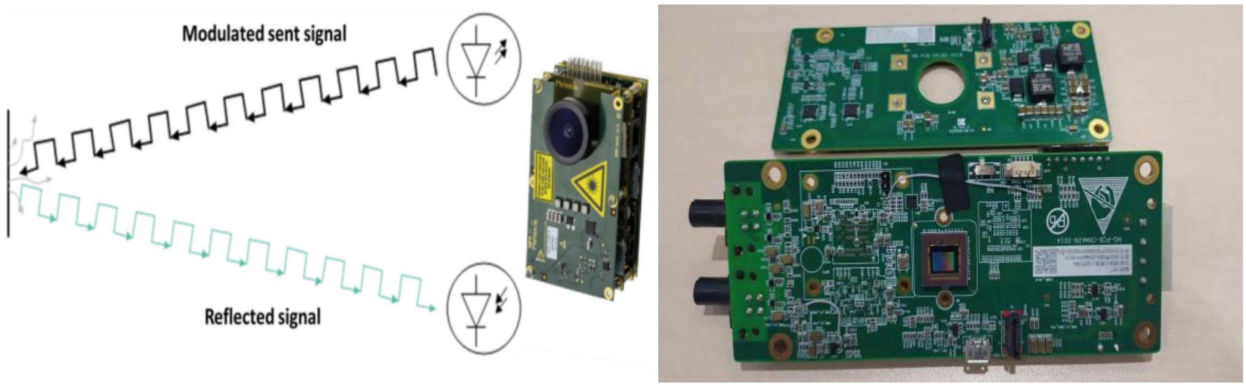

Figure 2: TOF Technology and depth detection chip

2.4. Scene Construction and Sampling

We built a door frame with a height of 1.9m and a width of 1.5m using a simple triangular support to simulate the door of the subway (as shown in Figure 3). The frame is placed in the laboratory corridor to collect data all day during the day.

Figure 3: Scene construction drawing

Above the frame is a TOF fisheye depth camera, which is connected to the nearby industrial computer through the data cable. The industrial computer obtains the real-time picture and stores the sample picture at the same time. Here, considering the problem of sampling quality, we set the pixel points (i.e. someone passes) in the video when they are more than 1.2m away from the ground and the sample area is greater than 40 * 40. When a valid sample is detected, the industrial computer restores the depth map according to the pixel depth value and brightness value returned by the camera and stores it in the folder.

2.4.1. TOF and binocular vision fusion Algorithm

(1) Algorithm idea

Through the analysis of the above scenes, combined with the technical advantages of TOF depth camera and binocular stereo vision, three problems are mainly solved: the first is the measurement defects of TOF depth camera in low reflectivity area and binocular stereo matching in low texture or repeated texture area; The second is to solve the measurement defect of occlusion area in binocular stereo matching; The third is to solve the measurement defect of TOF depth camera under strong ambient light, so that the three-dimensional depth camera can adapt to the depth information measurement of different scenes [9].

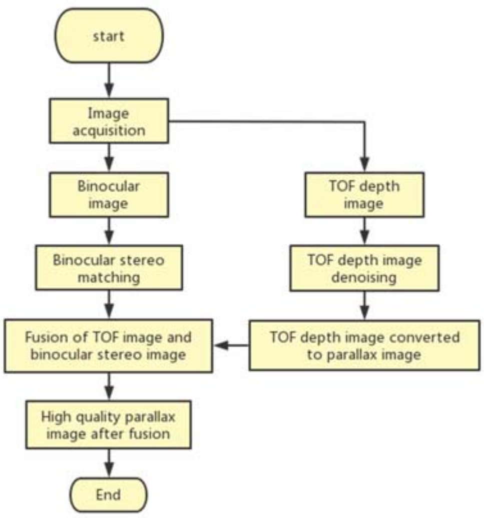

Figure 4: TOF and binocular stereo image fusion

As shown in Fig. 4, the denoised TOF depth image is mapped to the reference viewpoint of the left eye image to realize the fusion of TOF and binocular parallax image. It is necessary to denoise the TOF depth image to improve the quality of TOF depth image. After the noise reduction of the first part of the depth image, the TOF depth image is mapped to the left eye image in the binocular, and then the parallax range of binocular stereo matching is constrained to determine the best matching point to obtain the fused high-quality binocular parallax image [10].

2.4.2. TOF Depth Image Denoising Algorithm Processing

The exposure time and temperature drift of the TOF depth image will bring system noise. When the TOF depth camera is working, if the exposure time is not set properly, the obtained light intensity amplitude is too strong or weak, which will reduce the signal-to-noise ratio. Therefore, it is necessary to adaptively adjust the 3D exposure time of the TOF depth camera in the scene, and judge whether there is a low exposure or overexposed area through the light intensity amplitude value received by the TOF depth camera, If the average light intensity amplitude obtained is too high, the exposure time of TOF camera will be reduced dynamically; If the average light intensity amplitude value obtained is too low, the exposure time of TOF camera will be increased dynamically to make the average light intensity amplitude value obtained fall within a normal working range. According to the reference manual of TOF sensor provided by EPC company, the light intensity amplitude value falls within the range of 100lsb to 2000lsb, the measured depth image quality is good, the signal-to-noise ratio is large, and the effect of noise reduction of the system is played, That is, the algorithm processing flow of adaptive exposure adjustment is shown in Fig. 5.

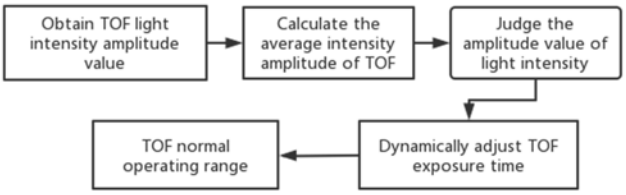

Figure 5: Flow chart of adaptive exposure adjustment algorithm

After adaptive exposure, the quality of TOF depth image is improved, and the problem of system measurement error caused by overexposure or low exposure in a large area of the image is effectively solved. However, there will still be overexposure of individual pixels in the object in the continuous area. The median value of 3x3 pixels in its neighborhood can be used to replace the overexposure or low exposure pixel depth value; Although the image obtained by the TOF array sensor is a three-dimensional depth image, the image noise is also similar to the traditional two-dimensional image sensor, which will have pepper and salt noise and speckle noise. Therefore, the median filter processing can effectively deal with the measurement error caused by the depth value of overexposed or low exposed pixels and the bad pixels of the TOF array sensor [11].

2.5. Marking and Training

After a week of data collection and sorting, we selected 5000 pseudo color images as qualified sampling data, and used labelme software to mark the images. When marking, frame the overhead part of each "passenger" in the image as the identification target, and generate the corresponding XML file and JPG file, as shown in Figure 6.

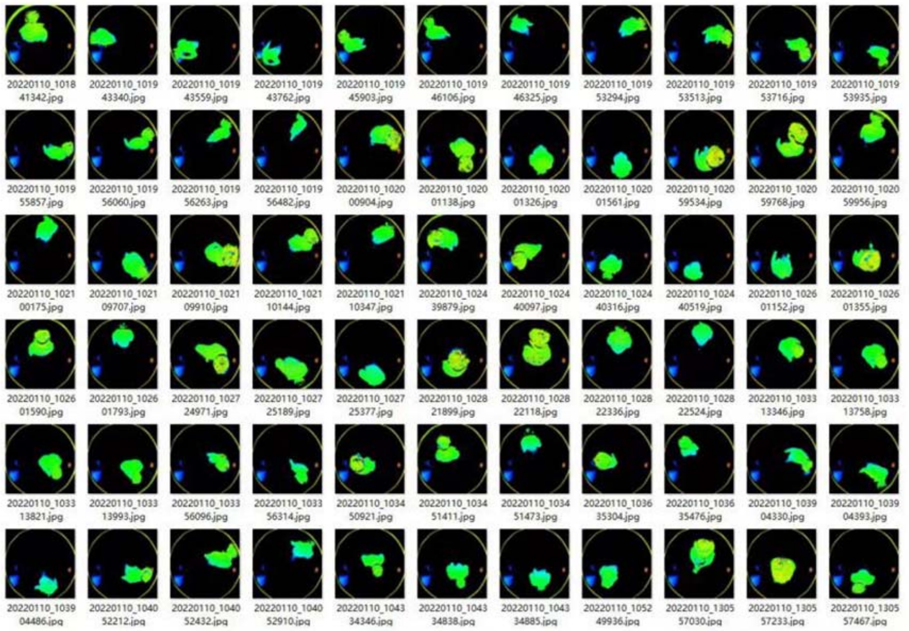

Figure 6: Image Overlays

2.5.1. Test Network Performance Evaluation Criteria

The evaluation indexes of classification network capability include global classification accuracy or average accuracy. Similarly, the commonly used evaluation standard for detecting network performance is map [12]. However, the detection network is a multi label image classification, so the calculation steps are more cumbersome than the classification network. Next, we will introduce some common concepts of accuracy, recall and average accuracy in target classification evaluation indicators. For the classification task, the relationship between the real value and the detection result will be as shown in Table 1.

Table 1: Confusion matrix of classification results

		Estimate	
		Positive	Negative
True value	Positive	TP (positive sample is judged as positive)	FN (positive samples are judged as negative)
	Negative	FP (negative samples are judged as positive)	TN (negative samples are judged as negative)

The confusion matrix is calculated for each class in the data set. Suppose that in the classification problem, there are two types of data sets: cats and dogs. For the confusion matrix of cats, TP represents correctly classified as cats, FP represents that dogs are classified as cats, FN represents that cats are classified as dogs, and TN represents that dogs are correctly classified as non cats. Accuracy is defined as the proportion of positive samples correctly predicted in all positive classes predicted. Recall rate is defined as the proportion of correctly predicted positive samples in all positive samples. The accuracy rate measures the ability of the model to judge the target category, while the recall rate measures the ability of the model to identify the target. Its calculation formulas are as follows (1) and (2):

$$precision = \frac{TP}{TP + FP} \tag{1}$$

$$recall = \frac{TP}{TP + FN} \tag{2}$$

Calculate the accuracy and recall corresponding to top-1 to top-n. Taking the recall rate as the abscissa and the accuracy rate as the ordinate, the precision recall curve can be obtained. Set a set of thresholds s = {0,0.1,..., 1} to obtain a corresponding maximum accuracy, as shown in formula (3),

$$p(s) = \max_{s'>s} p(s') \tag{3}$$

The average accuracy calculation formula of this kind is shown in formula (4):

$$AP = \frac{1}{11}\sum_{S\in\{0,0.1,\ldots,1\}} P(S) \tag{4}$$

The average accuracy measures the quality of the model in each category, so the average accuracy average (map) is needed to measure the detection performance of the model in each category. The calculation formula of map is shown in formula (5),

$$mAP = \sum_{i=1}^{n} AP_i \tag{5}$$

Where n represents the number of classes and API represents the average accuracy of class I. Similar to the classification network, the detection network also needs to calculate TP and FP during evaluation. For the case where there are multiple targets in a graph, the detection result is TP or FP, which needs to be determined according to the following rules:

(1) If there is no real value in the picture corresponding to the detection result, that is, there is no target, then the detection result is determined as FP; (2) There is no real value of this detection result category in the corresponding picture of the detection result. This detection result is determined as FP (3) the detection results are sorted according to the probability prediction value from large to small, and then match the real box of the corresponding category in turn. According to the set IOU threshold, the detection result with the real box IOU greater than the threshold is left, and the detection result lower than the threshold is determined as FP. If only one test result is left, the test result is marked as TP. If there are multiple, the one with the largest probability prediction value is marked as TP, and the others are marked as FP.

According to the above rules, after all test results are determined in turn, the precision recall curve can be obtained. Then the average accuracy (AP) of each category is obtained according to certain integration rules to measure the quality of the model in each category. Finally, the AP of each type is averaged to obtain the mean accuracy (map).

3. Development Results

3.1. Experimental Simulation and Analysis

There are many kinds of deep learning frameworks. The experiment of this paper is based on Ubuntu 16 04 operating system. Cafe is developed based on the convolutional neural network framework implemented by C + + / CUDA / python. The experiment is based on NVIDIA geforce gtx1080gpu.

The specific parameters of network training are set as follows. The random gradient descent (SGD) optimization method is used to reduce the loss function and update the network weight. The initial value of training learning rate is 0.001, and the learning rate is reduced to 1 / 10 times when the number of iterations is 40000 and 50000 respectively. The size of batch training is 16, the parameters momentum and weight day are set to 0.9 and 0.0005 respectively, and the total number of training iterations is 60000. The motorcycle training set is used to train the original SSD model and mobilenet SSD model respectively, and the detection performance of the two models is compared. The results are shown in Table 2. It can be seen from the table that after changing the distribution of regional candidate boxes, the map of the network is improved from 84.2721% to 87.088%, and the detection performance is greatly improved by about 2.8%. And the parameters of the model are reduced by 13%. The reduction of parameters can not only reduce the computational complexity, but also reduce the difficulty of solving the optimization problem, making the local optimal solution easier to approach the global optimal solution.

Table 2: Comparison of detection performance between SSD model and mobilenet SSD model

	SSD	MobileNet-SSD
mAP	84.2721%	87.088%
Parameter quantity	93.8M	82.6M

After setting the appropriate region candidate box, the matching threshold should be modified to achieve an optimal balance between accuracy and recall. Next, for the changed regional candidate box distribution, explore the optimal matching threshold. The thresholds are set to 0.5, 0.6, 0.7 and 0.8 respectively. The comparison diagram of network detection performance is shown in Figure 7.

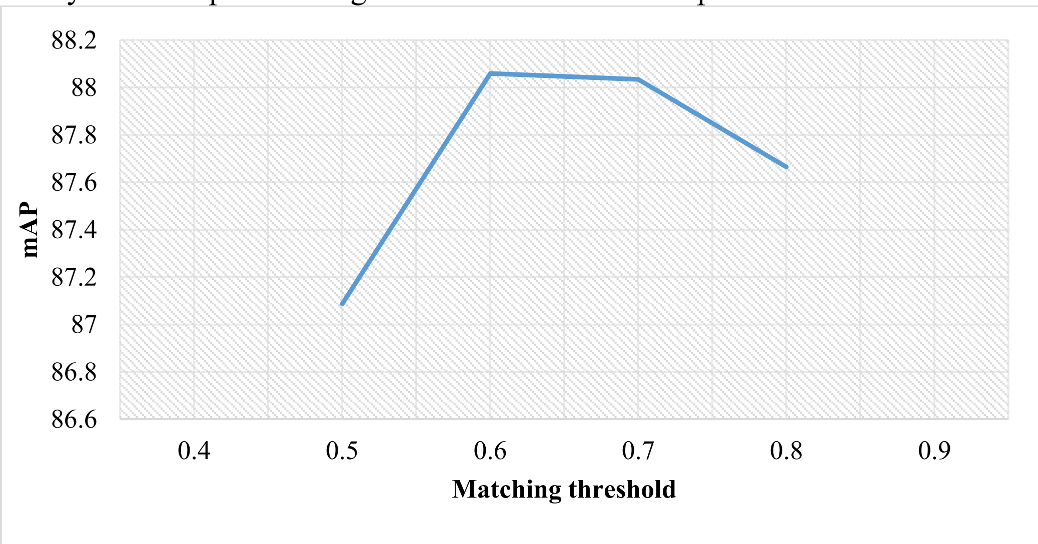

Figure 7: Relationship between matching threshold and detection performance

It can be found from Figure 7 that when the IOU threshold is 0.6, the network detection performance is the best, and the map reaches 98.0589%. When the threshold is 0.5, the restriction of the two frame relationship is too loose, which may introduce too much background noise, affect the feature learning, and lead to the decline of detection performance. Therefore, after setting the appropriate region candidate box, the matching threshold can be appropriately improved, which is beneficial to the improvement of detection accuracy and recall. In this example, the optimal threshold is 0.6 or 0.7.

Figure 8 shows the SSD original model and SSD_ Recall precision curve of our model (IOU threshold is set to 0.6), in which the red line represents the mobilenet SSD model and the blue line represents the SSD original model.

Through the above experiments, it can be seen that the regional candidate box distribution of the original SSD uniformly discretizes all the output space, but for a specific data set, the original distribution cannot effectively discretize the output space. Therefore, the distribution of specific data sets is analyzed first, so that it can effectively discretize the output space. Then, for the reset area candidate box, we explored the optimal matching threshold. We trained the mobilenet SSD target detection model with Caffe framework on the host of the laboratory, and finally screened the weight file. Finally, the detection accuracy of mobilenet SSD model increased from 84.27% to

98.06%.

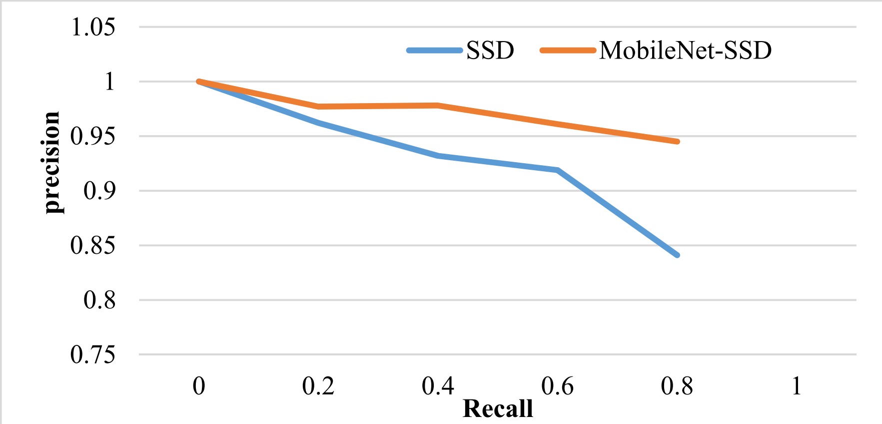

Figure 8: Recall precision curve of SSD original model and mobilenet SSD model

3.2. Accuracy

In order to verify the recognition accuracy of the developed passenger counter, we recruited a total of 10 volunteers, divided into two groups, standing on both sides of the simulated subway door. Each person passes through the subway door 20 times back and forth, that is, the total data should be 100 boarding passengers and 100 alighting passengers. During this period, volunteers are required to restore their actual state at random, that is, they can stop briefly due to congestion or pass through the subway door quickly. After the experiment, the passenger counter detects 98 times of getting on and 99 times of getting off, and the accuracy is as high as 98%.

3.3. Advantages and Practical Application Significance

Advantage 1: Based on TOF technology, it can eliminate the interference of shadow, height, luggage, multiple people, color, light, temperature, pressure and other factors. It has high accuracy and fits the actual application scene.

Advantage 2: the trained target detection model is mobilenet SSD. This lightweight network model has very low requirements for hardware performance. While ensuring the accuracy of the model, it has smaller volume and faster speed, making it possible for mobile terminals and embedded devices to run the neural network model.

Advantage 3: the project is implemented in C + + language, which is friendly to the later hardware transplantation, and uses tensorrtint8 quantization to further compress the operation power consumption and memory occupation, making it inevitable that the work can be put into practical use.

Advantage 4: it makes up for the blank of domestic rail transit passenger counter products and breaks the current situation that the product has been monopolized by foreign brands for a long time and the price of the product has been raised without an upper limit.

4. Application Prospect

4.1. Front End Transplantation

At this stage, we use a desktop computer to realize the reasoning process. In the actual application scenario, the passenger counter should be distributed above each door of the subway as an independent unit and connected with the switch and host on the train respectively. This requires us to add a chip with neural network reasoning function on the basis of TOF depth camera and make it into an integrated board in the later stage, The program directly records the data at the terminal, As shown in Figure 9.

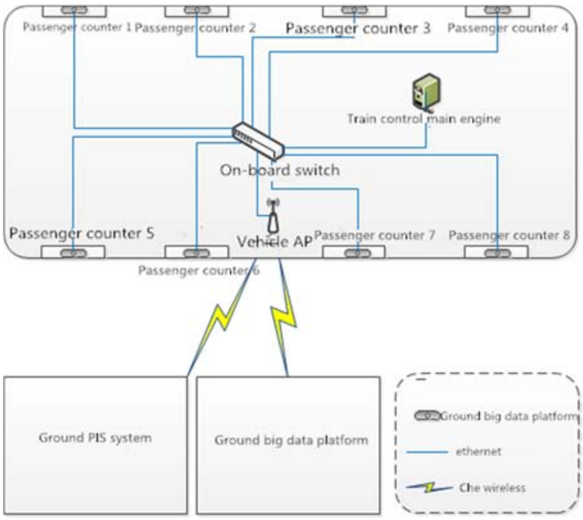

Figure 9: The desktop computer implements the reasoning process

After preliminary research, Hisilicon chip, Ruixin microchip and anba chip all have similar smart chips, but because we expect to use a chip that fully fits the needs, customization is the best choice.

4.2. Subway Passenger Count

As transportation students, we realize that all modes of transportation will evolve from non intelligent to intelligent. Just like the signal lights of road traffic, once upon a time, China was still in the situation of "fixed timing walking the world", but now vehicle detector and induction control have generally existed. The same is true for rail transit. Although most rail transit operations are still based on fixed departure intervals, one day intelligent adaptive control will also be applied to rail transit. At that time, a passenger counter with stable counting function will inevitably become the standard configuration of all trains.

4.3. Other Scene Extensions

The passenger counter based on TOF can not only be used in the field of rail transit. It can be actually applied to any location where passenger flow needs to be counted, such as shopping mall gate, airport security inspection, amusement park access control, public transport and so on. It can be said that it has a wide range of applications and strong expansibility.

5. Conclusions

In this paper, the passenger counter of urban rail transit is designed and analyzed, and the current situation of passenger count and passenger flow technology at home and abroad are investigated and analyzed; Through the investigation, it can be found that the method of binocular passenger counter or video detection is not accurate enough to determine the passenger flow, so Caffe framework is proposed to train the mobilenet SSD target detection model, which improves the detection accuracy, By analyzing the distribution of motorcycle samples in the vehicle data set, the clustering center of the data is obtained according to the kmeans method, and the distribution of regional candidate boxes is reset, so that the threshold of matching degree can be improved and useless regional candidate boxes can be deleted, which greatly reduces the amount of network parameters. The experimental results show that after selecting the appropriate region candidate box and raising the matching threshold, the accuracy of the network is improved from 84.27% map to 98.06% map; It is verified that based on TOF technology, it can eliminate the interference of shadow, height, luggage, multiple people, color, light, temperature, pressure and other factors, has high accuracy and fits the actual application scene.

References

[1] Sun Lihua. Research on Key Technologies of 3D visual simulation of urban road network. Huazhong University of science and technology, 2006.

[2] Krell H, Jaeschke H, Pfaff E. Highway Alignment Construction Comparison Using Object-Oriented 3D Visualization Modeling. Journal of Construction Engineering & Management, 2010, 140(10): 05014008.

[3] Xu Zhenhui, Qin Tao, Liu Shikuan, Fei Yun. Research on 3D road modeling and visualization method Highway, 2011.03-161

[4] Tian Bijiang, Yang Guoqing, Su Yu, Lu Yingzhi, Hu Chengyu. Discussion on road 3D visualization technology and its application. Highway Traffic Technology (Applied Technology Edition), 2016,12 (11): 49-51

[5] Yang Xiao. Research on the application of BIM visualization technology in a complex building project. North China University of water resources and hydropower, 2019

[6] Li Hao. Research on urban trunk road design optimization based on visualization technology. Xi'an University of technology, 2020

[7] Jing Y., Hu H., Guo S., et al. Short-Term Prediction of Urban Rail Transit Passenger Flow in External Passenger Transport Hub Based on LSTM-LGB-DRS. IEEE Transactions on Intelligent Transportation Systems, 2020, PP(99):1-11.

[8] Xiong Z., Zheng J., Song D., et al. Passenger Flow Prediction of Urban Rail Transit Based on Deep Learning Methods. Smart Cities, 2019, 2(3):371-387.

[9] Zhu K., Xun P., Li W., et al. Prediction of Passenger Flow in Urban Rail Transit Based on Big Data Analysis and Deep Learning. IEEE Access, 2019, PP (99):1-1.

[10] Jiang X., Feng J., Jia F. Modeling and Simulation of Passenger Distribution in Large-scale Urban Rail Transit Network. Tiedao Xuebao/Journal of the China Railway Society, 2019, 40(11):9-18.

[11] Pan Z., Wei Q., Torp O., et al. Influence of Evacuation Walkway Design Parameters on Passenger Evacuation Time along Elevated Rail Transit Lines Using a Multi-Agent Simulation. Sustainability, 2019, 11(6049):1-17.

[12] Guo Y., Wang X., Xu Q., et al. Weather Impact on Passenger Flow of Rail Transit Lines. Civil Engineering Journal, 2020, 6(2):276-284.