Design and Implementation of Pedestrian Flow Monitoring and Warning Platform in Public Area based on Pedestrian Detection

Aijun Zhen¹, Qingli Wang¹, Ancheng Wang¹, Kai Wu¹
¹School of Computer and Information Engineering, Tianjin Agricultural University, Tianjin, China

zhenaj@tjau.edu.cn

Abstract—In the traditional public area pedestrian flow monitoring, a lot of manpower and material resources are needed. Only relying on people's eyes to identify and count the number of people in the video under monitoring is not only inefficient, but also has errors in accuracy and wastes a certain amount of energy and financial resources. In the rapid development trend of artificial intelligence, this system starts from the perspective of real-time monitoring, with the maturity and development of machine vision technology, and creates a cloud monitoring platform based on JavaWeb technology. The pedestrian detection technology is applied to the pedestrian flow monitoring, which provides a feasible scheme for accurate and fast pedestrian flow monitoring and warning, and develops a system to help public area managers and government departments to deal with the crowd gathering situation, analyze and monitor the crowd density, and give a quick warning to the crowd outbreak and gathering place.

1. INTRODUCTION
With the development of artificial intelligence, pedestrian detection technology based on machine vision technology[1] has been widely used in life. Whether it is a tourist attraction or a traffic hub and other public places, the monitoring of pedestrian flow has always been a hot issue. The current situation and trend of pedestrian flow density should be analyzed. In the traditional crowd flow monitoring, managers need to enter the public places to observe and monitor on the spot, or use the camera for monitoring, but it can only provide the managers with the original image, unable to analyze the image accurately and effectively, And can not avoid and stop the occurrence of accidents in time. In order to achieve early warning, managers must invest a lot of manpower and material resources. The development of monitoring video population statistics technology from manual counting to automatic counting further determines the importance of the number of people counting technology based on machine vision in public places[2]. Use the supervision platform to supervise the crowd, the number of people can be identified by the program. If the number of people exceeds the maximum capacity of the scenic spot, it will be reported to the cloud monitoring platform for warning and decision-making for government departments, leaders and manager of public area. The dispatching staff can conduct guidance and management according to the data, which is convenient and fast.

2. RESEARCH ON KEY TECHNOLOGY OF SYSTEM IMPLEMENTATION
Since the publication of RCNN algorithm in 2014, target detection has bid farewell to the traditional method of feature extraction and classifier combination, and has entered a new era of classification and
regression using convolutional neural network\cite{3}. Its task is to segment the background of "we don't care" and obtain the "we care about" foreground target\cite{4}, and the current target detection algorithm is divided into two categories; One is two stage detection, which mainly includes r-cnn\cite{5}, fast RCNN\cite{6} and fast RCNN\cite{7}. It needs to go through the region proposal stage, that is to generate 1000-2000 candidate regions, and then classify each candidate region. Therefore, the accuracy of this type of detection algorithm is higher, but its speed will be slower; The other is one stage detection, It mainly includes SSD\cite{9}, retinanet and Yolo\cite{8} series. It does not need to go through the candidate region stage, and can directly extract features in the network to predict object classification and location. Therefore, its advantages and disadvantages are opposite to the two stage detection algorithm, which is fast, but its accuracy will decline.

Yolo series of target detection is a typical representative of one stage detection algorithm. Its biggest feature is that it can directly use CNN convolution network neural network to detect the target end-to-end, and then the coordinate position, confidence and category probability of the target can be obtained, so the final detection result can be obtained after a single detection. As the latest algorithm of Yolo series, V3 is better than Yolo_V2\cite{10} is a little larger, but it is still fast and more accurate\cite{11}. It has both reservation and improvement on the previous algorithm, mainly manifested in its selection of better backbone network darknet-53, multi-scale prediction (similar to FPN algorithm) and the use of better classifiers. It has won praise for its near extreme speed and excellent accuracy, and has gradually reached the balance of speed and precision.

Yolo network is mainly divided into three parts: CNN convolution layer, regression network layer and NMS screening layer. When we input a picture of any size and keep the aspect ratio unchanged, the new graph zoomed to 416×416 and rgb3 channel is used as network input Yolo_V3 uses multi-scale prediction\cite{12}, and outputs three layers after CNN convolution neural network. Each layer has S×S grids with the size of 13×13,26×26 and 52×52, and then each grid is responsible for detecting the targets whose center points fall in the grid, and predicts the center coordinates (x, y), width and height(W,H) and confidence of B bounding boxes, The confidence level reflects whether the object is included or not and the accuracy of the position in the case of containing the object, as in (1).

\[
\text{confidence} = p(\text{Object}) \times \text{IoU}_{\text{pred}}^{\text{truth}}
\]  

Finally, through the NMS screening layer, the bounding boxes with low credibility are removed according to the score threshold, and the remaining bounding boxes are subjected to NMS non maximum suppression, and the most suitable rectangular boxes are screened out by removing the high overlap.

As a multi-target tracking algorithm, deep sort can be divided into prediction part and update part. It takes the intersection and union ratio IoU between each detection and the predicted target frame of the existing target as the measurement of the target relationship between the front and rear frames. It uses the Kalman filter\cite{13} to predict the current position, extracts the apparent features of the target for nearest neighbor matching, and associates the target frame with the target through Hungary algorithm\cite{14}, and uses Yolo_V3 is the network of target detection. For each frame of the detected video, the prediction part will read the position of the target detection frame in the current frame and the depth characteristics of each target frame image block, delete the target frame and features whose confidence level does not meet the threshold, and suppress the target frame with non maximum value to eliminate the situation of multiple frames on a target, Then the Kalman filter is used to predict the position of the target in the current frame. The update part is to update the parameters and feature set of the Kalman tracker. The Hungarian matching algorithm with weight is used to match the previous motion track with the current detection object to form the motion track of the object, The weight is obtained by the weighted sum of Mahalanobis distance of point and trajectory and the similarity of image (cosine distance of vector). When calculating Mahalanobis distance, Kalman filter is used to predict the covariance matrix of motion distribution to judge whether the target disappears or new target appears.
3. DESIGN AND IMPLEMENTATION OF PEDESTRIAN FLOW MONITORING AND EARLY WARNING SYSTEM

The system is mainly composed of edge and cloud side. The system edge is mainly responsible for image acquisition, pedestrian detection and data upload, and the system cloud side is mainly responsible for image and data display, and voice warning in case of crowd outbreak. The two systems interact with each other through HTTP protocol get request and post response, All functions of the pedestrian flow monitoring and warning platform based on pedestrian detection are realized.

3.1 Design of system function module

3.1.1 Design of system structure

The ultimate goal of the system is to realize the intelligent monitoring of pedestrian flow, improve the accuracy of monitoring, liberate a certain amount of labor, and promote the development of pedestrian detection technology application. The system uses Python language and Java language to achieve different functions of the system, Python language is used to implement the image processing of OpenCV computer database and tensorflow deep learning framework to realize Yolo_V3 algorithm is used to detect pedestrians and identify the number of people. The cloud monitoring platform uses Java language and uses springboot framework to display data and video. By building a high-performance nginx reverse proxy server in aliyun server, the processed video image is played in real-time on the cloud monitoring platform by using multimedia video processing tool FFmpeg and RTMP protocol streaming. In the database part, mybatis framework is used to realize the operation and processing of data information, and HighCharts chart library is used to draw broken line statistical chart and real-time dynamic broken line area map, which makes the display of human flow more intuitive and provides theoretical support for management personnel to rationalize resource allocation.

3.1.2 Design of system edge module

The edge of the system is composed of three functional modules: pedestrian detection, live streaming media transmission and pedestrian flow data upload, which enables managers to realize pedestrian detection at the system edge, live streaming media transmission, and upload real-time data of pedestrian flow.

3.1.3 Design of system cloud module

The cloud of the system is composed of management personnel setting, pedestrian flow threshold setting, live video monitoring, real-time dynamic display of pedestrian flow data, voice warning of pedestrian flow, statistical analysis and prediction of recent pedestrian flow. Managers can set the threshold value of pedestrian flow in the cloud, monitor the real-time video of public places, and view the real-time traffic data. When the flow of people is large, voice warning can be provided. The system can count the daily total flow of the past month and predict the trend of the flow of people in the next month.

3.1.4 Analysis of system business

After logging in to the cloud system, the scenic spot managers set the threshold of pedestrian flow according to the scene, and then transfer the video image source to the edge end, and start the program to process the image and identify the number of people. The processed real-time data of video and human flow are transmitted to the cloud. Then, the video image and the real-time dynamic broken line area map of the flow of people can be viewed after returning to the cloud. When the pedestrian flow is greater than the set maximum flow threshold, the voice alarm is immediately given to remind the management personnel to take appropriate measures. In addition, the management personnel can also view the daily total flow line chart of nearly one month. And according to the current location of all traffic data forecast line chart, access to the trend of the flow information. Finally, administrator can modify the account information. The system activity diagram is shown in Figure 1.
3.2 Realization of system edge function

3.2.1 Pedestrian detection
Anaconda, as an open source Python distribution, contains a large number of software packages and their dependencies. It can establish different running environments for different projects without interfering with each other. It can be said that anaconda is a powerful tool for Python development. The system edge is to install OpenCV, Tensorflow, Keras, Numpy and other toolkits under the created Anaconda environment, and use OpenCV to obtain video source. Read the picture of each frame, change the image into a new image filled with absolute gray "R128-G128-B128", normalize and adjust the size of 416×416, so as to meet the input format of Yolo network. Then, the neural network enters the CNN convolution network and goes through 53 layers of convolution\(^{[18]}\) to remove the last layer of full connection layer, the remaining 52 convolution layers are used as the main network. These 52 convolution layers are composed of a 32 size convolution kernel convolution and five groups of repeated residual units\(^{[19]}\). Each residual unit includes a separate convolution layer and a group of repeated convolution layers. In each repeated convolution layer, the convolution operation of 1×1 is performed first, and then the convolution operation of 3×3 is performed. And the number of repetitions is 1, 2, 8, 8 and 4. Because the step size of the first independent convolution layer operation of each group of residual units is 2, the dimension is reduced for 5 times. Finally, three feature maps with sizes of 13×13, 26×26 and 52×52 can be obtained. By K-means clustering, Yolo_V3 obtains three groups of nine prior boxes, which are assigned to the three feature maps respectively. Because the smaller the feature map, the larger the receptive field is, the more sensitive it is to the large target, so the larger the prior frame is, and the smaller the prior frame is used for the larger feature map.

Then, the width, height, central coordinates, confidence level and category of the target frame are obtained. The final width and height are obtained by calculating the width and height with the width and height of the prior box, and the center coordinates are calculated by the offset between the sigmoid function and the grid cell to obtain the final center coordinate. The final score of each bounding box is the product of the confidence degree calculated by the sigmoid function and the category probability, and then the target box whose score is less than the threshold is deleted. Call the pillow library to draw the target box, including line width, font, different targets and different colors. Finally, only the target box with the highest score and the minimum overlap is retained by NMS non maximum suppression algorithm.

3.2.2 Live streaming media transmission
Considering the high connection concurrency, the system first builds the Nginx Server\(^{[20]}\). Nginx is easy to install, has simple configuration files, has few bugs, and is easy to start, and can almost run continuously for 7×24. Even if it runs for several months, it doesn't need to be restarted. Then, use the multimedia video processing tool FFmpeg to upload the processed video image to the server with RTMP (real time messaging protocol) real-time message transfer protocol.
3.2.3 Upload real-time data of pedestrian flow
The system edge uses post request to submit the real-time data of human flow to the cloud monitoring platform to upload the data of human flow.

The recognition effect of the system edge is shown in Figure 2.

![Figure 2: Recognition effect of the system edge](image)

3.3 Realization of system cloud function

3.3.1 Live streaming
Live streaming by Video.js 5.4.1, play the RTMP protocol video stream.

3.3.2 Real time dynamic display of pedestrian flow data
The `<a>` tag in HTML5 is used for dynamic display of real-time data. The ID and initial value of 0 are set for this tag. Then, the Ajax request method is defined by JS to obtain the current passenger flow, and the value is written through the ID of the `<a>` tag. The page setting is refreshed every 2 seconds, and the Ajax request can continuously obtain the changed pedestrian flow value and realize real-time dynamic display.

3.3.3 Dynamic real time display of broken line area map
Use the HighCharts chart library to create a broken line area chart. The data also comes from the dynamic change of people flow value obtained by Ajax request, and the flow value is taken as the y-axis, and the time is set as the x-axis, which is refreshed every 2 seconds to realize the dynamic real-time broken line area map\(^{[21]}\). By obtaining the threshold set by the system, the color of different areas of the y-axis is set according to the size of the threshold, Orange means high flow, blue means normal flow, purple means low flow.

3.3.4 Voice warning of pedestrian flow
In the Ajax request of real-time dynamic display of traffic data, the returned data is converted from String to Int type, and then compared with the threshold set by the system. If the maximum threshold is exceeded, the specified alarm content is read out by using HTML5 Web Speech API to realize voice alarm of pedestrian flow.

3.3.5 Statistical chart of pedestrian flow in last month
The daily flow of people in the last month can be found out through the database, and the line chart can be drawn by using the highcharts chart library.

3.3.6 Forecast chart of pedestrian flow in the next month
The date of all dates and the daily pedestrian flow value are found through the database. A linear model of \(y = ax + B\) is simulated with the date as the independent variable and the pedestrian flow as the dependent variable. The next month's date is used as the independent variable to predict the next month's pedestrian flow value according to the model. Finally, the line chart is drawn by using the HighCharts chart library.
3.3.7 Pedestrian flow threshold setting
The user sets the threshold according to the scene, including the maximum adult flow, the median flow and the minimum flow. After the user setting, the database is updated and the value is transferred to the Input tag through the session for voice warning of pedestrian flow.

3.3.8 Administrator information settings
Administrator can change the user name, password, mobile phone number and other information of the account.

The effect of the system cloud is shown in Figure 3.

![Figure 3: effect of the system cloud function](image)

4. CONCLUSION
This paper implements a pedestrian flow monitoring and early warning platform based on pedestrian detection. With the support of machine vision and pedestrian detection theory, it solves the problem of pedestrian flow supervision in public places, greatly improves the efficiency of supervision, reduces the artificial burden, and has important guiding significance for practice. However, affected by the development environment, the system still has some deficiencies. The system function needs to be further refined, and the system will be further improved from this point of view in the future.

ACKNOWLEDGEMENT
This research was financially supported by the Key Teaching Reform Projects of Tianjin Agricultural University in 2018 (Project Number:2018-A-18). The Key Teaching Reform Project Name is “Discussion on the teaching reform of managing the software process course under the guidance of new engineering concept”.

REFERENCES
[1] Zhang Y, Zhou D, Chen S, et al. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network[C]. IEEE Conference on Computer Vision and Pattern Recognition. 2016:589-597
[2] Rengang Li. Design and implementation of Intelligent City pedestrian flow statistics system based on machine vision technology[J]. University of Electronic Science and technology.
[3] Qunli Yao, Xian Hu,Hong Lei. Research progress of deep convolution neural network in target detection[J]. Computer engineering and application, 2018, 54 (17):1-9.
[4] Hongpeng Yin,Bo Chen, Yi Chai, Zhaodong Liu. Overview of vision based target detection and tracking[J]. Acta automatica Sinica, June 21, 2016
[5] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//CVPR, 2014: 580-587
[6] Girshick R.Fast R-CNN[C]//ICCV,2015:1440-1448.
[7] Ren S Q,He K M,Girshick R,et al.Faster R-CNN:towards real-time object detection with region proposal networks[J].IEEE Transactions on Pattern Analysis and Machine Intelligence.2017,39(6):1137-1149.
[8] Liu W,Anguelov D,Erhan D.SSD: single shot multibox detector[C] //ECCV,2016:21-37.
[9] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C] // CVPR, 2016.
[10] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [J]. arXiv: 1612.08242v1, 2016.
[11] Redmon J, Farhadi A. YOLO, V3: an incremental improvement [J]. arXiv: 1804.02767v1, 2018.
[12] Fukai Zhang, Feng Yang, CE Li. Fast vehicle detection method based on improved yolov3 [J]. Acta automatica Sinica, 2016-06-21
[13] Dan Yang. Design and application of Kalman filter [D]. Xiangtan University, 2014
[14] Jun Bao. Overview of target tracking algorithm based on deep sort [a]. Network application branch of China Computer Users Association. Proceedings of the 23rd annual meeting of network new technology and application of China Computer Users Association in 2019 [C]. Network application branch of China Computer Users Association, 2019: 5
[15] Jia Huang. Research on computer vision technology based on OpenCV [D]. East China University of science and technology, 2013
[16] Minmin Zhang, Heping Xu, Xiaojie Wang, Mengyun Zhou, Shuyue Hong. Google tensorflow machine learning framework and application [J]. Microcomputer and applications, 2017, 36 (10): 58-60
[17] Yin Zhang. Design and implementation of streaming media system based on RTMP protocol [D]. University of Electronic Science and technology, 2015
[18] Mingyu Guo, Shi Liu. Optimization and implementation of yolov3 image recognition tracking algorithm [J]. Electronic test, 2019 (15): 65-66+86
[19] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // CVPR, 2016.
[20] Xing Du. Research on the theory and technology of lightweight web server nginx [D]. Nanjing University of Posts and telecommunications, 2016
[21] Zhong Xie, Hongxia Yan. JQuery + highcharts realize dynamic statistical chart [J]. Computer programming skills and maintenance, 2014 (13): 64-65+86