ISSN: 2414-1895 DOI: 10.6919/ICJE.202111_7(11).0065

Analysis of bus passenger congestion based on MSCNN

Jiwei Lv [a], Weizhong Zhang [b,*]

College of Computer Science and Technology, Qingdao University, Qingdao 266071, China;

[a]ljw573493024@163.com

*Corresponding author Email: bzhangwz_01@aliyun.com

Abstract

In the construction of intelligent bus dispatching system, the problem of passenger congestion in buses has become an urgent problem.With the continuous development of computer vision technology and the gradual popularity of video surveillance systems in buses, it is possible to use video images in buses to detect congestion.Since the upper limit of the number of passengers on the bus is fixed, there is no need to predict the number of passengers with excessive passenger density. This paper chooses a unified multi-scale deep convolutional neural network with a relatively simple network structure to analyze bus passenger congestion. By using vgg16 as the density classification network for feature extraction, MSCNN is used to estimate the number of bus passengers and the definition and classification of the concept of congestion. According to the data set used in the experiment, the accuracy of predicting the congestion of passengers in the bus can reach 96.2%.The experimental results show that this method is of great significance to the analysis of bus passenger congestion.

Keywords

Unified neural network, multi-scale, density map, vgg16, congestion.

1. Introduction

The bus is one of the main means of transportation in today's society, and it is a fair and open convenient means of transportation that most people in the society give priority to in daily life and activities. It is also the most important component of the urban public transportation system[1]. However, the current operation and service levels of buses still need to be improved. The problem of passenger congestion in buses occurs from time to time. The traditional dispatch method based on historical passenger flow information does not consider real-time passenger flow and congestion information, which often leads to capacity. The mismatch with passenger flow makes buses often overcrowded or vacant, which greatly affects the level of bus operation and the public ride experience, and also affects the attractiveness of buses to citizens to travel.

This article focuses on the problem of detecting the degree of congestion in the bus, relying on the current intelligent vehicle-mounted surveillance camera[2] equipped with each bus, using the combination of video image processing technology and deep learning technology to complete the detection of passenger congestion[3] in the bus. Provide real-time congestion information in buses for bus companies and the public. Provide data support for the public's travel choices, the safety of corporate bus transportation capacity, route planning, dynamic control, etc., and improve the satisfaction and riding experience of bus passengers.

At present, there are still few researches on the detection of crowdedness in buses at home and abroad, and the related research mainly focuses on crowd density estimation. Therefore, this article mainly designs a method for detecting crowdedness in buses by drawing on the crowd density estimation

ISSN: 2414-1895 DOI: 10.6919/ICJE.202111_7(11).0065

method based on video images. Traditional crowd density estimation methods are mainly divided into the following categories: video-based methods, detection-based methods, regression-based methods, density map-based methods, and convolutional neural network-based methods[4].Video-based crowd density estimation is an earlier method[5]. Through a series of continuous frames in the video, the number of pedestrians is estimated based on the movement of the crowd and the characteristics of the human body. The continuous frame sequence is compared, the background and the foreground are separated, and then the comparing the characteristics of people with people[6], the limitation of this method is that it cannot estimate the crowd density in static pictures. Based on the estimation of the population density of the detection, earlier research can be traced back to 2004 after the HOG operator-based pedestrian detection method was proposed, and some scholars built a pedestrian technology algorithm based on this algorithm. These methods usually detect people or heads through a sliding window on the image. They can achieve excellent detection accuracy in sparse scenes. However, when encountering occlusion and background chaos in extremely dense crowds, their results will be unsatisfactory[7],Between 2008 and 2014, the method based on the number of people regression solved the shortcomings of the pedestrian detection method and can be effective for scenes with a large number of people. The main idea is to use various feature descriptors to extract features from the input image or video, and then use various machine learning models to regress the number of people. The previous method successfully solved the problems of occlusion and background confusion, but always ignored the spatial information. Therefore, Lemptisky et al. first adopted a method based on density estimation by learning the linear mapping between local features and the corresponding density map. These methods consider spatial information, but they only use traditional manual features to extract low-level information, which cannot guide high-quality density maps to estimate more accurate counts. Compared with traditional feature extraction, the crowd density estimation method based on convolutional neural network has made great progress[8].Benefiting from the powerful feature representation of CNN, more and more researchers use it to improve density estimation. Earlier heuristic models usually use basic CNN to predict the density of the crowd, and they will be significantly improved compared with the traditional methods of people flow statistics. For different levels of supervision and learning paradigms of different models, there are also models designed across scenarios and fields.

With the rapid development of deep learning, some classic algorithms have been applied to various scenarios in life. This paper proposes to apply the deep learning algorithm and theory to the intelligent bus dispatching system, according to the pictures collected by the intelligent camera of the bus, using the deep convolutional neural network training to generate the density map to detect and classify the degree of congestion in the bus.

2. Fundamental

2.1 MSCNN

A unified deep neural network, denoted the multi-scale CNN(MS-CNN), is proposed for fast multi-scale object detection[9]. MSCNN is a two-stage object detector, composed of an object proposal network and an accurate detection network. In the proposal subnet, detection is performed at multiple scale output layers, so that the receptive field matches objects of different scales. Combining the results of these different scales can produce a powerful multi-scale detector. By optimizing the multi-task loss, the end-to-end learning of the unified network is realized. In addition, the deconvolution feature upsampling is also discussed as an alternative method of input upsampling to reduce memory and computational cost.

MSCNN multi-scaled Faster RCNN to improve the ability to distinguish small targets. This is because RPN generates proposals of multiple scales by sliding a set of fixed filters on a set of fixed convolution feature maps. This creates an inconsistency between the object size (variable) and the filter acceptance domain (fixed).

ISSN: 2414-1895 DOI: 10.6919/ICJE.202111_7(11).0065

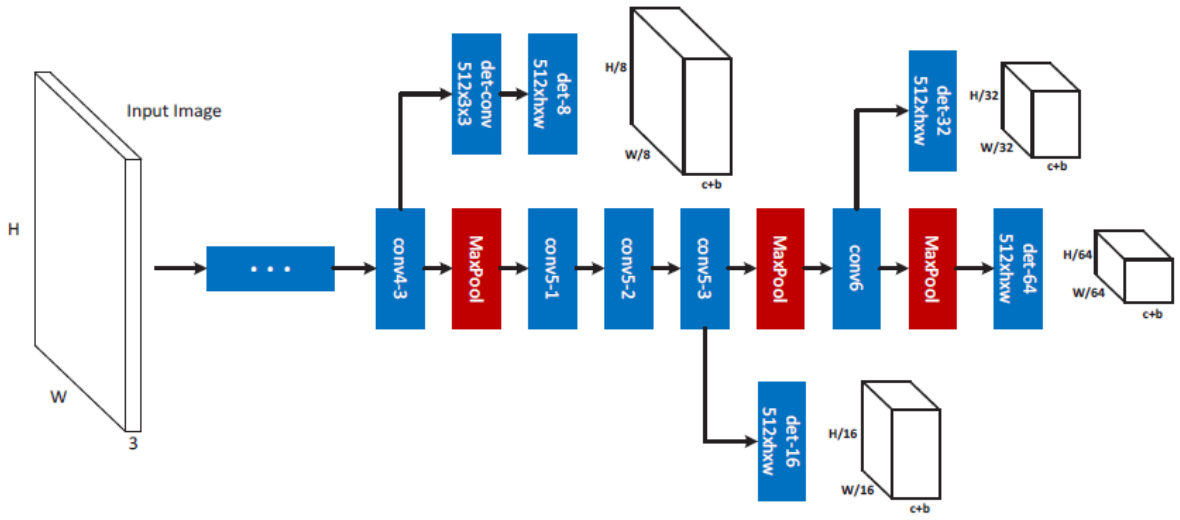

Fig 1 Proposal sub-network of the MS-CNN. The bold cubes are the output tensor of the network. h×w is the filter size, c the number of classes, and b the number of bounding box coordinates.

2.2 Loss function of MSCNN.

For a detection network with multiple branches, the total loss value is calculated by weighting:

$$\mathcal{L}(\mathbf{W}) = \sum_{m=1}^{M} \sum_{i \in S^m} \alpha_m l^m(X_i, Y_i \mid \mathbf{W})$$

The loss function in each branch is defined as follows:

$$l(X, Y \mid \mathbf{W}) = L_{cls}(p(X), y) + \lambda[y \geq 1] L_{loc}(b, \hat{b})$$

Among them, the loss of positioning is defined as:

$$L_{loc}(b, \hat{b}) = \frac{1}{4} \sum_{j \in \{x,y,w,h\}} \text{smooth}_{L_1}(b_j, \hat{b}_j)$$

In order to solve the problem caused by the unbalanced sampling of positive and negative samples in the detection process, the loss function of the classification is modified to a weighted form:

$$L_{cls} = \frac{1}{1+\gamma} \frac{1}{|S_+|} \sum_{i \in S_+} -\log p_{y_i}(X_i) + \frac{\gamma}{1+\gamma} \frac{1}{|S_-|} \sum_{i \in S_-} -\log p_0(X_i)$$

2.3 Classification of crowdedness in buses

In 2004, the National Standard of the People's Republic of China "Technical Conditions for the Operation of Motor Vehicles" stipulated that urban buses and trolleybuses are 0.125 square meters for one person, and a maximum of 8 people per square meter. In fact, no matter how crowded the bus is, it is difficult to squeeze 8 people in 1 square meter. Crowding is a relatively subjective concept. It is affected by personal factors such as body shape, personality characteristics and behavior habits, as well as external factors such as space size, temperature, and so on. There is currently no clear and unified definition. Congestion is a qualitative problem, but it is necessary to formulate its quantitative standards in the establishment of a sample library of images of congestion in buses and subsequent research on congestion detection. The definition of congestion in the "Manual on Public Transportation Capacity and Service Quality" (TCQSM2013) by the American Transportation Research Council is shown in Table 2-1:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202111_7(11).0065

Tab. 2-1 Level of service criteria for standing passenger space in TCQSM2013

Service level	Standing passenger area (㎡/person)	Passenger experience
A	≥1.00	Passengers can move freely, most or all passengers have seats
B	[0.50,1.00)	There is a distance between passengers, standing comfortably
C	[0.40.0.50)	Standing passengers have no physical contact, standing passengers and sitting passengers have basically the same personal space
D	[0.30,0.40)	Standing passengers have occasional physical contact, and the personal space of standing passengers is smaller than that of sitting passengers
E	[0.20.0.30)	Approaching uncomfortable conditions, frequent physical contact with standing passengers
F	<0.20	Extreme congestion

Since the above table divides the standing passenger area service level into six levels, there are too many levels for crowding analysis. In order to provide passengers with an intuitive description of the degree of congestion, this article divides the degree of congestion into four levels based on the above table, as shown in Table 2-2: comfortable, normal, crowded, and very crowded.

Tab. 2-2 Rough classification standard for crowding in bus

Crowdedness	Standing passenger area (㎡/person)	Description of passenger congestion
Comfortable	≥1.00	Passengers can move freely in the carriage, most or all passengers have seats
normal	[0.40.1.00)	There is a lot of space left in the compartment, and there is little physical contact with standing passengers
Crowded	[0.20.0.40)	Standing passengers have more physical contact, and passengers feel slightly crowded
Extremely crowded	<0.20	Standing passengers have frequent physical contact and passengers feel very crowded, which affects getting on and off the bus

2.4 Establishment and processing of bus passenger training set.

The source of the data in this article is from the newly launched new energy buses in Qingdao in 2020. At present, the number of cameras in most buses is 6-8. Among them, there are 6 cameras in Qingdao city transport buses as shown in Figure 2, and the data samples are collected from 6 cameras in various buses. The video images taken by the front cameras inside the buses are selected for congestion detection research. At the same time, considering the high similarity between adjacent video frames of the video in the bus, the video frames are extracted from the video at 30-frame intervals as sample images. Some sample images are shown in Figure 3. In order to protect the privacy of passengers,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202111_7(11).0065

and at the request of the data provider, the images in the bus shown in this article have already occluded the faces of passengers and vehicle information.

Fig 2 Bus and in-car camera

Comfortable normal

Crowded Extremely crowded

Fig 3 Part of the sample data set

3. Training and prediction of mscnn

3.1 The generation of density map

If the position of a marked point is x_i,the changed point can be expressed as $\delta(x - x_i)$,where $\delta(x - x_i)$ is the impact function. Therefore, a label with N heads can be expressed as:

$$H(x) = \sum_{i=0}^{N} \delta(x - x_i)$$

The density map generated in this way has a serious problem, which will cause the generated density map to be very sparse, resulting in the overall output approaching 0 when the network calculates the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202111_7(11).0065

loss, and it is not conducive to counting the number of people when the population density is high. Therefore, the Gaussian function can be used to convolve the above formula, and the position marked as the human head becomes the density function of the area. This not only solves the sparseness of the picture to a certain extent, but does not change the way the number of people in the picture is counted. The choice of the Gaussian kernel convolution kernel is very important. Due to the perspective distortion of the picture, the scale of the pixels and the surrounding samples are inconsistent in different scene areas. Therefore, in order to accurately estimate the population density function, perspective transformation needs to be considered. Assuming that the population around each head is uniformly distributed, the average distance in the image between the head and the nearest k neighbors can give a reasonable estimate of geometric distortion.

3.2 Density grade classification network

The purpose of the density classification network is to roughly classify images according to the situation of the data set, filter out the density of people of different levels, reduce the difficulty of training the number of people estimation network and improve the accuracy. The density level is divided into 3 categories, namely none (no people in the picture), low density (1 to 99 people in the picture), and high density (the number of people is greater than 100). Use 0,1,2 as the category label, adopt one-hot coding, and use cross entropy as the loss function of the network. Because the classification goal is simple, the vgg16 network with the fully connected layer removed is used as the feature extraction network, followed by 3 custom fully connected layers to finally classify the output. The vgg16 network does not participate in training, and only trains a custom three-layer fully connected network.

3.3 Prediction and output

Since this project will output all pictures with more than 100 people to 100 people, which is very crowded, the first step is to use the density level network to divide the pictures into 3 categories when outputting. Then send the output picture with label 1 to mscnn to predict the number of people, and finally give the number of people contained in all the pictures and the degree of congestion. Here, by setting a threshold for the number of people, 1~15 people are considered comfortable, 15~35 people are normal input, 35~55 people are crowded, and 55~100 people are very crowded. Under my 1080ti graphics card, it outputs 20 to 30 frames per second.

4. Summary

This article mainly studies the detection and classification of passenger congestion in buses in the field of urban transportation, and has done a lot of attempts and work through field investigation of this research point and how to integrate deep learning algorithms with the practical application. Based on the basic model framework of the convolutional neural network MSCNN, this paper uses density map classification to apply the crowd density estimation originally used in high-density to the relatively small number of bus passengers, thereby solving the problem of the large scale of the original model and the large amount of calculation. Finally, this paper applies the predicted number of people to the problem of bus congestion detection, and then optimizes the overall accuracy of the congestion detection.

References

[1] Wang Wei, Yang Xinmiao, Chen Xuewu. Urban Public Transportation System[M]. (Science Press, China, 2002).

[2] Huang Kaiqi, Chen Xiaotang, Kang Yunfeng. Overview of Intelligent Video Surveillance Technology[J]. 2015.

[3] Yang Zeng. Research and Realization of Congestion Classification Algorithm in Bus Based on Deep Learning[D] (Master, Chongqing University, China, 2019). p.2-3.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202111_7(11).0065

[4] Chen Lei, Wang Guodong.Multi-level fusion convolutional neural network for crowd density estimation[J].Journal of Qingdao University (Natural Science Edition),2020,33(04):31-36.

[5] Paul Viola and Michael J. Jones and Daniel Snow. Detecting Pedestrians Using Patterns of Motion and Appearance[J]. International Journal of Computer Vision, 2005, 63(2) : 153-161.

[6] Sheng-Fuu Lin, Jaw-Yeh Chen, Hung-Xin Chao. Estimation of number of people in crowded scenes using perspective transformation.[J]. IEEE Transactions on Systems, Man & Cybernetics: Part A, 2001, 31(6):645-654.

[7] Congdon Peter. Quantile Regression for Area Disease Counts: Bayesian Estimation using Generalized Poisson Regression[J]. International Journal of Statistics in Medical Research, 2017, 6(3):92-103.

[8] Shelhamer Evan and Long Jonathan and Darrell Trevor. Fully Convolutional Networks for Semantic Segmentation.[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(4) : 640-651.

[9] Zhaowei Cai, Quanfu Fan, Rogério Schmidt Feris, et al. A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection.[J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2016, abs/1607.07155