Pedestrian Detection in Crowded Crowd Scene Based on Deep Learning

Zhenbin Le\textsuperscript{1,}\textsuperscript{*} and Zhaohui Meng\textsuperscript{1}

\textsuperscript{1}Department of Computer and Information Hohai University, Nanjing, China, 211100

*Corresponding author e-mail: wangyu_@hhu.edu.cn

Abstract. Detection-based methods typically detect and locate each person on a crowd image by using a designed pedestrian target detector and obtain counting results by accumulating each detected person. However, these methods require a large amount of computational resources and are often limited by human occlusion and complex background in real scenes, resulting in inaccurate detection. Based on the characteristics of computer depth learning and population density distribution map, a more optimized pedestrian detection approach is proposed.

Keywords: Deep Learning, Pedestrian Detection, Monitoring

1. Introduction
Economic development and social prosperity have brought about rapid population mobility, not only the economic density of society is increasing, but also the density of people in many public places is also very large for a long time. In daily life, we see a lot of crowded places, shopping malls, supermarkets, railway stations, stadiums, life squares, fitness and entertainment places are crowded places, long filled with crowded people\cite{1}. Not only that, some places where people gather on weekdays, in festivals or activities is even more crowded, the size of the crowd and crowd dynamics affect people's safety. After entering the new century, there have been many crowded stampede in crowded places, and even terrorist violence in some crowded places. The development of modern science and technology also brings solutions to this situation. Video surveillance is one of the technologies that give the supervisor a certain degree of control over his place, but in crowded places, Ordinary video surveillance is also often dizzying. Video number statistics involves a variety of algorithms such as image processing, computer vision and data mining\cite{2}. The purpose is to effectively monitor the crowd in the monitoring area and estimate the density information of the crowd in the scene in real time or to count the specific number of the crowd in the scene. Some places use traditional light, electricity, pressure and other sensors to count the number of people in crowded places, but this kind of method is seriously missed and false detection, and the application of the scene is limited, after the emergence of video statistics has been gradually eliminated. The existing video number statistics method can achieve good results in a specific scene, but it is difficult to achieve strong robustness and high accuracy due to the problems of light change, occlusion and pedestrian scale change in video images\cite{3}. The emergence of deep learning provides a new possible method for pedestrian detection in places where people gather.

2. Overview of deep learning
2.1. Concept
Deep learning is a new concept in recent years. Tracing back to its source has been in application for a long time. It is a branch of machine learning. It means that the machine has multiple levels of circular
logic resolution through the training of data. The machine mechanism for deep learning is deep learning neural network. The construction of deep learning network involves many fields, needs to have special advanced professional knowledge, is even more complicated in the original design, and it is difficult to divide each other. A wide range of vector transformations and features can only be realized on the basis of long-term analysis and research by multi-domain experts.

### 2.2. Features

The process of deep learning is carried out within the processor of the computer. Only those who are proficient in this profession can know the context of it. For users of deep learning machines, nothing can be seen except the intelligent superiority of deep learning machines. For non-professional people, deep learning has a strong concealment, and this concealment comes from its natural abstract invisible characteristics. Through multi-level learning and mapping, we can gradually get high-level abstract features from the bottom features such as edge, color and so on. These abstract features have high dimension and strong discrimination. Using simple classifiers, we can achieve high accuracy classification, regression and other tasks. Most of the time, at the beginning of the design of deep learning, the designers are result-oriented, reverse reasoning step by step to the complex learning network, from simple to complex, using the way to break down stems, leaves, words from the root, from the purpose of the problem, and then solve it\[^4\].

### 2.3. Steps

In the text introduction, the steps of deep learning are very concise, that is, network design, goal setting and cumulative training. Deep learning can be used for different fields, different industries, different professional service, and each different design needs a database and a set of functions to design different tree neural networks according to the actual needs. This function is understood in the same way as the function in mathematics, and different variables are mapped to different solution sets. The problem of mapping results at each step is the source of the goal setting. Then the correspondence of the mapping is verified in continuous training to ensure that the network structure is consistent with the target.

### 3. Neural structure of deep learning network

The development of computer hardware is the material basis of deep learning neural network. Post-painting, high-speed CPU and performance-enhanced graphics cards make it possible to process a large amount of data at high speed. Deep learning has opened a new AI era in many application fields, especially in image processing, speech recognition, and natural language processing. Figure 1 shows the deep learning neural network structure.

![Figure 1](image_url)

**Figure 1.** Schematic diagram of deep learning neural network.
3.1. Convolutional layers
It is also a filter of network structure, which can extract target features well and analyze each small area of neural network more deeply to obtain more abstract features. According to the step size set in advance, for each step size, the convolution kernel makes a convolution calculation with the feature map of its corresponding position, and the output feature map is used as the input of the next layer.

3.2. Pool layer
The dimension of the feature graph obtained by convolution layer is too large, so adding pool layer to reduce the matrix, the compressed feature graph size can become smaller, which makes the network calculation more simplified and speeds up the calculation speed. Pooling operation also ensures the robustness of network learning to affine transformation such as scale change or translation. The common pooling methods include maximum pooling and average pooling. The maximum pool operation diagram is shown in figure 2.

![Figure 2. Schematic diagram of maximum pool operation.](image)

3.3. Full connection layer
The full connection layer plays a simpler role. Each layer before it works independently, and only the parts related to the latter layer will produce a connection, otherwise the convolution layer and the pool layer will stay at their respective levels. With the full connection layer, each neuron in the first layer feature map is connected with the neuron of the input layer, which can be connected with the convolution layer or the full connection layer, which needs to be determined according to the need.

4. Population density map
4.1. Fixed Gaussian distribution of population density
Using the property that the sum of all values in the Gaussian distribution is 1, the position of each head in the population density map is represented by a two-dimensional Gaussian distribution. The number of people in the image can be obtained by integrating the population density map.

4.2. Population density perspective
In practical applications, the Gaussian distribution cannot accurately correspond to the size of all human heads in the scene. Since the population density map is the training label for convolutional neural networks, it may lead to problems such as wrong optimization direction, difficult convergence and poor generalization ability of convolutional neural network model, which ultimately affects the performance of convolutional neural network\textsuperscript{[5]}. Therefore, in the data set with this perspective, the size of Gaussian kernel can be designed according to the information given by the perspective, so that the size of Gaussian distribution accords with the size of the head in the real scene.
4.3. Making of population density map of label

The population density map of the input image is estimated by convolution neural network, and the number of people is calculated by the population density map. Among them, the training set generation scheme plays an important role in the accuracy of the number statistics. To some extent, the learning density map can save more information of the image and help to improve the counting accuracy. This requires the production of training data through the data set label population density map.

5. Pedestrian detection in crowded crowd scene based on deep learning

Pedestrian detection in crowded crowd scenes based on deep learning is rarely studied. The development of traditional methods in the progress of science and technology has improved the detection accuracy to a certain extent, but it still has not got rid of the problems existing in the traditional pedestrian detection method: it is easy to be affected by occlusion. The performance is very poor when the crowd is crowded. Some researchers have tried to solve these problems, focusing on some of the characteristics of the human body, and tend to solve the population count from the overlooking perspective[6]. On the basis of deep learning, a global convolution neural network model is proposed to detect the position of human head, and a pair of object models is used to train together. It performs well in a small number of life scenes, but the model is complex and the processing speed is slow. Recently, the method of using deep learning to return to number has proved to be effective. Convolutional neural networks are used to alternately learn population density maps and numbers. These algorithms based on single-column convolutional neural networks are limited by extracting scale-related features and are difficult to solve the scale change problem on crowd images. The structure of a multi-column convolution neural network using deep net and shallow net is proposed, which combines the characteristics of different dimensions and improves the resolution of the output population density map[7]. Classifying the number of people to each level is equivalent to roughly estimating the total number of people in the image, thus merging the advanced features into the density estimation network, that is, the layers in the network can learn the global features.

6. Conclusion

At present, deep learning is a hot research direction in the field of machine learning. Using its application in pedestrian detection equipment, it can successfully overcome the problems of poor pertinence of sliding window strategy based on image and a wide variety of windows. Improve the accuracy of detection, in crowded crowd scene pedestrian detection application prospects.

References

[1] Lu Kun. A Study on Pedestrian Detection in Monitoring [D].Scene 2019.
[2] Jiao Huiying. A Study on the Statistical Method of Video Number Based on Machine Learning Algorithm [D].2019.
[3] Huang Jingwen. Video Surveillance Pedestrian Detection [D].Based on Depth Regression 2017.
[4] Wang Jing Jing. A Study on Pedestrian Detection Method Based on Multi-feature Learning [D].
[5] Ding Wenxiang. Pedestrian detection in real-time video based on deep neural network [D].2016.
[6] Chang Lingling. A Study on Pedestrian Detection Based on Depth Learning [D]. 2017.
[7] Zhang Yang, Liu Weiming, Wu Yihu. pedestrian detection method based on deep belief network classification algorithm Computer Applied Research [J].2016,33(2):594-597.