Research on Multi-feature and Machine Learning Hierarchical Pedestrian Detection Method Based on Deep Learning

Yitong Hao
Harbin Engineering University, Harbin, China
657820855@qq.com

Abstract. With the rapid development of science and technology, computers have become an important way for people to obtain information. As an important means of transmitting visual information, image plays a very important role. Pedestrian detection is a key research field of machine vision and a key technology in intelligent video surveillance and intelligent transportation system. Pedestrian detection is the basis and premise of pedestrian tracking, behavior analysis, gait analysis and pedestrian identification, and a good pedestrian detection algorithm can provide strong support and guarantee for the latter. Pedestrian detection is not an independent problem, it is closely related to video pedestrian tracking, human behavior recognition and pedestrian identification. The method of machine learning is to let the program have the ability of learning. Through continuous training and iteration, the parameters of the model are optimized. Deep learning can give full play to the speed and rationality of the machine and improve the accuracy of pedestrian detection.

1. Introduction
In recent years, intelligent transportation, intelligent video surveillance and other fields have developed rapidly, which has made people's lives more convenient. As a key technology in these fields, pedestrian detection has attracted more and more attention and research from relevant professionals in the industry [1]. Pedestrian detection is to segment pedestrians appearing in video sequences or images from the background and accurately locate them. It is currently one of the most active research topics in the field of computer vision. With the rapid development of science and technology, computers have become an important way for people to obtain information. As an important means of transmitting visual information, images have a very important position. Every day people can obtain a large amount of image information through computers, and this information affects our lives all the time. As a key science and technology in smart devices, computer vision applications can allow machine devices to obtain surrounding video or image information, and automatically analyze and process these perception information, so as to have the ability of intelligent visual perception and thinking [2]. The core idea of machine learning is to train the machine to simulate the human brain, learn and judge [3]. If there is a pedestrian on the road ahead, the vehicle assisted driving system can detect and judge in advance, prompt the driver of the road ahead, and automatically take relevant emergency braking measures to avoid car accidents as much as possible.

Pedestrian detection is a key intelligent technology in computer vision applications, which refers to the precise calibration of the position coordinates of pedestrians from the collected images [4].
Pedestrian detection is not an independent problem, it is closely related to video pedestrian tracking, human behavior recognition, and pedestrian identification [5]. Pedestrian detection is to identify pedestrians in the collected video in general traffic scenes and give the results of the frame level. In further applications, such as tracking and re-recognition, each pedestrian needs to be distinguished [6].

Pedestrian detection is the basis and prerequisite of pedestrian tracking, behavior analysis, gait analysis, and pedestrian identification. A good pedestrian detection algorithm can provide strong support and protection for the latter [7]. The purpose of monitoring is to accurately judge travelers and their behaviors, and predict pedestrians with abnormal behaviors and possible unsafe accidents, so as to ensure the safety of people's lives and property [8]. Although people can easily obtain various image information through various shooting devices, these images can also be conveniently stored in various storage devices. The method of machine learning is to make the program have the ability of learning, and through continuous training and iteration, the parameters of the model are optimized. Deep learning can give full play to the speed and rational advantages of machines and improve the accuracy of pedestrian detection.

2. Basis of visual target detection

Visual object detection is a basic topic of computer vision, and its purpose is to study the image recognition and positioning functions in machine intelligence. Pedestrian detection is a typical target detection problem in the field of computer vision. In fact, it is a binary classification problem. It is mainly used to distinguish pedestrian targets from background through statistical learning methods. The idea of the candidate area is to extract a small rectangular area from the image to be detected for target positioning. This small rectangular window is the research area of interest. In this rectangular area, some feature description operators are used to calculate the correlation of the area. Features, and finally use the trained classifier to classify and recognize the test examples. In any area of the image, the most common way is to use windows of multiple scales and aspect ratios to slide across the image [9].

Pedestrian detection means that the computer judges whether there is a pedestrian in a given image and video, and if there is a pedestrian, the specific location of the pedestrian needs to be given. Pedestrian detection is the basis and premise of research on pedestrian tracking, behavior analysis, gait analysis, and pedestrian identification. A good pedestrian detection algorithm can provide strong support and protection for the latter.

In the training phase, a training sample set is first given, where the positive sample set is composed of pedestrian targets, and the negative sample set is obtained from background image collection, and then a specific feature extraction algorithm is used on the sample set to convert from image space to feature space. The feature dimension conversion is then trained by a machine learning algorithm to obtain a classification model to determine the sample category. When testing the test pictures, integrate their classification results for pedestrian detection. The integration method uses a weighted sum method, as shown in the formula:

$$C(x) = \sum_{m=1}^{M} a_m C_m(x)$$

Among them, $C_m(x)$ is the m-th global classifier, $a_m$ is its weight, and M is the number of global classifiers.

Traditional manually set features are difficult to express the target accurately. Secondly, traditional features can only extract the underlying image features. Because the deep network can learn features from data unsupervised, and this learning method also conforms to the mechanism of human perception of the world, the features learned through the deep network often have certain semantic features when there are enough training samples, and are more suitable for the recognition of targets and behaviors. In the detection process, the pre-calibrated detection window is mainly obtained by sliding window strategy, and the feature extraction algorithm is also applied to transform into feature space, which is screened by the trained classification detection model. In terms of efficiency, the traditional detection method uses sliding windows to traverse to achieve positioning, and the windows traverse the original
image in turn according to a certain sequence and step size, and each step is judged. The biggest problem brought by this method is excessive computational redundancy, and the positioning deviation is large, and the performance will be greatly reduced.

3. Pedestrian detection based on machine learning

3.1 Feature extraction method

The input of the lowest layer of the network is the local receptive field of the image. The convolution kernel operation is used between the layers to obtain the output feature image of the latter layer. Each convolution layer will be followed by an activation function layer, by introducing a nonlinear transformation, To make the expression and fitting ability of the model stronger, and then introduce the pooling layer to reduce the dimensionality of the features of the convolutional layer. At present, the pedestrian detection methods based on machine learning mostly learn from fixed-size training samples. The detectors obtained by this type of method can only detect fixed-size images, and cannot directly determine the location of pedestrians in the image [10]. Data expansion is an effective method to solve the problem of small samples, and the classification ability of the model can be further enhanced by expanding the sample size. In the convolutional layer, in order to make full use of the more potential information of the image, the sparsity of the feature extraction of the convolutional layer is strengthened. After the convolution operation of the last layer, a part detection map is generated, and the structure of the sub-sampling layer that should be Change to the deformation processing layer, deal with the possible deformation of the part, and finally get the corresponding part detection score.

According to different actual needs, some follow-up processing of the detection results may be required, which may include analysis and processing of pedestrian movement paths, judgment and prediction of human action behavior, and system alarms. The specific process is shown in Figure 1.

![Figure 1 Pedestrian detection process](image)

Based on the limited information sample that can be applied, the expected risk cannot be reasonably calculated and analyzed. Therefore, the traditional method is mainly to use the empirical risk minimization standard, and the risk is defined by the sample:

$$R_{emp} = \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i, w))$$  \hspace{1cm} (2)

It can be obtained that the probability \( l \) between actual risk and empirical risk meets the following conditions:

$$R(\omega) \leq R_{emp}(\omega) + \sqrt{\frac{h(\ln(2l/h+1) - \ln(\eta/4))}{l}}$$  \hspace{1cm} (3)
Among them, \( l \) is the number of samples in the training set, and \( h \) is the VC dimension of the function set, reflecting the high and low complexity.

Because the local neighborhood pixels of the image are closely related, and the distant pixels have poor correlation, after the multi-layer mapping, the high-level neurons merge the information of the local units to obtain the global characterization information of the image. The effect of pedestrian detection directly depends on the selection of features and the selection of classifiers. Generally, if you have a good feature, you can achieve good detection results even if you choose a simple classifier. In traditional image recognition and classification tasks, complex artificial feature extraction processes are usually required to reduce dimensionality, while CNN can directly use images as the input of the network, thereby avoiding artificial subjectivity, and the weights on the mapping surface are shared, thereby greatly reducing the training parameters and further reducing the complexity of the network.

3.2 Pedestrian detection
In the training stage, random deactivation is equivalent to sampling the complete neural network, randomly extracting some neurons every time, and then updating the parameters of this sub-network according to the input data. Different from traditional image processing methods, training set also becomes a factor that affects the performance of detection methods based on statistical learning. When the size of pedestrian is different from the size of sliding window, it is necessary to zoom the picture in order to detect pedestrian accurately. However, the size of pedestrians cannot be predicted in advance for an image to be detected, so the method based on multi-scale sliding window needs to detect every scaled image to ensure that pedestrians in the image will not be missed. In the test of target detection methods, because different applications pay attention to different aspects of detection results, it is necessary to use multiple indicators to characterize the performance of different aspects, and use complex curves to characterize the overall level of the detector [11]. There is a defect in the sliding window method. If the size of pedestrians in the image to be detected is the same as the size of the image that the detector can detect, the detector can detect correctly. If they are different, they cannot be detected correctly. After data preprocessing, the data has appeared in the form of an analyzable two-dimensional table. When the developed system structure system is put into market, professionals should conduct system test to prevent system installation defects. The research on the reliability of network topology needs a comprehensive, complete and reasonable analysis from two aspects of network devices and traffic load. The algorithm flow is shown in Figure 2.
After the vision-based detection method enters the era of statistical learning, the demand for samples in different scenes is extremely urgent, because the classification part of its detection is essentially transformed into the classification problem of high-dimensional data space. With the development of statistical learning, the combination of feature extraction and classifier has become the core of detection framework, and motion extraction has become the preprocessing part.

4. Conclusions
Vision is an important sense of human beings, and human beings perceive changes in the outside world mostly through vision. It is of great practical significance to use computers to simulate human visual cognitive ability. After using two kinds of features to build detectors respectively, the fusion method of these two kinds of detectors is better than the fusion method of two kinds of features directly connected. When detecting a new window, these global classifiers are integrated, and the final detection result is determined by weighted voting. Different algorithms have their own unique advantages for the specific recognition object of pedestrians. Although the application of deep learning algorithm is simpler, it doesn't need to determine a fixed mathematical model, but it needs to be trained on the basis of massive data to reduce the error rate and get a better recognition effect. The training speed of deep network needs to be further optimized. By studying the relationship between deep learning model, traditional computer vision system and image processing technology, new model training methods may be inspired. In the process of pedestrian detection, there may be a stronger tendency to some features. Therefore, in the next step, we can add the corresponding weight feedback mechanism, and judge the robustness and stability of features by the weight, so as to obtain more accurate detection results.

References
[1] Zhang Yang. Pedestrian detection algorithm combining texture features and deep learning[J]. Journal of Liaoning Technical University: Natural Science Edition, 2016, 210(02):206-210.
[2] Deng Wei, Liu Binghan. Pedestrian detection based on deep learning with semantics[J]. Computer System Applications, 2018, 27(006):165-170.
[3] Tian Qing, Yuan Tongyang, Yang Dan, et al. Pedestrian detection method based on dark channel dehazing and deep learning[J]. Progress in Laser and Optoelectronics, 2018, 55(11):181-189.
[4] Taiana M, Nascimento J, Bernardino A. On the purity of training and testing data for learning: The case of pedestrian detection[J]. Neurocomputing, 2015, 150(20):214-226.

[5] Alom M Z, Sidike P, Taha T M, et al. State Preserving Extreme Learning Machine: A Monotonically Increasing Learning Approach[J]. Neural Processing Letters, 2017, 45(2):703-725.

[6] Brunetti A, Buongiorno D, Trotta G F, et al. Computer vision and deep learning techniques for pedestrian detection and tracking: A survey[J]. NEUROCOMPUTING, 2018, 300(26):17-33.

[7] Qiu Dong, Liu Deyu. Pedestrian detection method based on fuzzy deep learning network[J]. Computer Technology and Development, 2018, 028(010):22-26.

[8] Zhong Chong, Xu Guangzhu. Moving pedestrian detection method combining foreground detection and deep learning[J]. Computer and Digital Engineering, 2016(12): 2396-2399.

[9] Hoang V D, Jo K H. Joint components based pedestrian detection in crowded scenes using extended feature descriptors[J]. Neurocomputing, 2016, 188(5):139-150.

[10] Goldhammer M, Khler S, Zernetsch S, et al. Intentions of Vulnerable Road Users—Detection and Forecasting by Means of Machine Learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(7):3035-3045.

[11] Liu Li, He Wubin, Yang Sai, et al. Traffic safety intelligent warning system based on pedestrian detection[J]. Electronic Design Engineering, 2020, 28(04):96-99.