July-August 2020
ISSN: 0193-4120 Page No. 5238 - 5246

Analysis on SSD Real Time Highway Congestion Based on Convolutional Neural Network Target Detection Algorithm

DongmeiLiu[1], MingDeng[1], XuejunLi[1], MingjingLi[1] and ChuanliYin[2,*]
[1]Electronic Information EngineeringCollege, Changchun University, Changchun, Jilin, China, 130022
[2]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin, China, 130000

Article Info
Volume 83
Page Number: 5238 - 5246
Publication Issue:
July - August 2020

Article History
Article Received: *25 April 2020*
Revised: *29 May 2020*
Accepted: *20 June 2020*
Publication: *28 August 2020*

Abstract

Target detection is to identify and locate the object of interest from static images or video sequences, which is one of the key tasks in the field of computer vision. Target detection uses image processing, machine learning, deep learning and artificial intelligence technology, which has been widely used in many fields, such as intelligent transportation, military exercises, medical image analysis, industrial detection and so on. However, in real-time highway congestion analysis, there are still many environmental interference factors, such as brightness, shape, color, occlusion and so on, which makes the research opportunities and challenges of target detection algorithm coexist. Firstly, this paper analyzes the basic concepts of convolution neural network target detection algorithm. Then, this paper analyzes the real-time road congestion model of SSD. Finally, an empirical model is proposed.

Keywords: *Convolution Neural Network, Target Detection Algorithm, Ssd, Real-Time Road Congestion;*

1. Introduction

Object detection is the core of computer vision. With the robustness of detection algorithm becoming stronger and stronger, we have made a great breakthrough in deep learning technology, which has been applied to the field of computer vision. At present, target detection framework has become the mainstream. However, the detection accuracy and detection speed always restrict the practicability of the target detection algorithm, which may lead to problems such as missing detection, especially small target detection.

For the detection target, the early industrial focus is mainly on the face, people, cars and so on, which are very important for video monitoring, road traffic and other fields. So far, computers need to understand scenes more comprehensively, and the categories of detection have been extended to all aspects of life[1]. For different application scenarios, target detection technologies include single target detection and multiple target detection. In the field of transportation, target detection technology has a wide range of applications, including automatic vehicle detection and recognition, road vehicle and pedestrian positioning, bayonet and road video monitoring, etc.). Through the image processing technology, we can identify the road situation, which not only needs to detect the car, pedestrian signal, but also to judge the traffic signal. Finally, based on these information objectives, we can guide the operation of the car[2].

Visual information is the most important and efficient form of human receiving external information. Images and videos are the most important sources of human information, which can describe the objective things vividly[3]. People try to simulate the human brain through computer to recognize images, which will lead to a discipline, such as computer vision[4]. Through computer simulation of human brain, computer vision can automatically identify the effective information carried by image and video data, which can identify and locate the target of interest in static image or video sequence[5]. Through the comprehensive use of image processing, machine learning, deep learning and artificial intelligence technology, target detection has been widely used in military exercises, medical image analysis, intelligent transportation[6]. Through target detection, we can achieve auto focus, which will automatically adjust the contrast and saturation of the region and get better picture quality.

2. Related research status

2.1. Research status of vehicle detection

At present, the commonly used vehicle detection methods based on video can be divided into two categories: Vehicle Detection Based on motion information and vehicle detection based on feature information[7]. Vehicle detection methods based on motion information mainly include optical flow method, frame difference method and background difference method. Optical flow method has a large amount of calculation, poor anti-interference and real-time performance[8]. The frame difference method is easy to form a large hole in the moving target, and the moving target speed is too fast or too slow, which may lead to unable to segment the moving target[9]. The background subtraction method uses the subtraction of the current image and the background modeling image to complete the detection of moving objects[10]. This method can not only detect the short-term stationary vehicles (long-time stationary vehicles are classified as the background), but also not limited by the speed of the vehicle, and meet the requirements of real-time traffic video detection. However, with the change of weather, light and other factors, the background will change accordingly, so the background modeling should be adaptive.

Common background modeling methods: Gaussian mixture model, codebook, self-organizing background detection, sample consistency background modeling vibe model, pbas detection algorithm, etc. Gaussian mixture model is better than other background modeling algorithms. Many new or improved algorithms are based on its different variants. However, the disadvantage of Gaussian mixture algorithm is that the calculation is relatively large, resulting in slow detection speed and sensitivity to light. Codebook is sensitive to illumination, and sobs is robust to illumination, but map model makes the calculation much more. Sacon adopts pixel based and block based update strategies, but the sample set replacement strategy is FIFO, which has great limitations. The speed of vibe algorithm is very fast, the amount of calculation is relatively small, and it has certain robustness to noise, and the detection effect is also good. Pbas algorithm combines the advantages of Sacon and vibe algorithm, and introduces the idea of cybernetics and the measurement method of background complexity. However, the computational complexity of pbas algorithm is also increased, which is time-consuming.

2.2. Research status of vehicle behavior recognition

Vehicle behavior recognition and analysis is an active and challenging research field. At present, there are two kinds of methods: vehicle behavior learning based on topic model and vehicle behavior learning based on trajectory.

Vehicle behavior recognition based on topic model does not need very accurate trajectory information. In recent years, there are quite a lot of important researches using topic model to do behavior analysis. The general methods include latent Dirichlet allocation (LDA) and probabilistic latent semantic analysis (PLSA). In addition, tdpmm includes the time change information in the trajectory modeling. Dynamic HDP is also applied to track learning behavior distribution, which can update shared semantic regions. However, topic model-based behavior recognition is mainly used to describe panoramic behavior, which is not suitable for real-time judgment of vehicle behavior.

Aiming at the problem of real-time vehicle behavior detection, researchers generally adopt the vehicle behavior recognition method based on trajectory. First of all, according to the trajectory of a series of vehicles, the similar motion behaviors are divided into a group. Then, we establish a unified behavior model for the motion behavior in the same category. Finally, we detect the abnormal trajectories of some real-time trajectories. If part of the trajectory passes the detection, we can predict the trajectory along the motion mode. Morris and Trivedi proposed a three-level learning process model, which uses different resolutions to specify specific behaviors at different levels. The first layer only considers the node information of the start and end points; the second layer provides the spatial distribution information between nodes; the third layer uses the spatiotemporal dynamic coding hidden Markov model (activity HMM) probability model to describe the behavior.

The quality of trajectory based vehicle behavior learning largely depends on the robustness of vehicle tracking influencing factors, such as noise, illumination change, shadow and occlusion. Trajectory clustering requires pairwise computation of the similarity between trajectories, which consumes time and memory. Although the vehicle behavior modeling method based on HMM model has a high recognition rate, it only considers the positive samples of this type of vehicle trajectory, which ignores the influence of negative samples of other types of vehicle behavior. By using the maximum likelihood value for classification and recognition, we limit the classification ability of HMM vehicle behavior modeling method, which leads to great limitations in multi category vehicle trajectory recognition.

3. The development of convolution neural network

3.1. Theory stage

In 1962, Hubel and Wiese found that there were very complex hierarchies in the lower layers of the visual system when they were experimenting with the visual cortex of animals. Most neurons selectively respond to local stimuli, which indicates that visual signals are transmitted from the retina to the brain through multiple layers of local receptive fields. In 1980, Fukushima proposed a self-organizing model based on the transmission of visual signals, which is called neurocognitive machine. The model is a multilayer neural network. The response of each neuron is excited by the neurons in the local receptive field of the previous layer, which simulates the cell function and hierarchical structure of the visual cortex. In this stage, the discovery and application of local receptive field lay a theoretical foundation for the development of convolutional neural network.

3.2. Model implementation phase

In 1988, based on the neurocognitive machine, Yann Lecun proposed the convolutional neural network LeNet-5 model. In this model, local receptive field, shared weight and pooling structure are added to the traditional neural network model, which makes the convolution neural network have translation invariance, rotation invariance, scale invariance and anti deformation ability to a certain extent. In convolution neural network LeNet-5 model, convolution layer and pooling layer are set alternately, which can abstract the input image into a set of feature maps through multiple nonlinear transformations. Through the fully connected neural network, we can classify the features, which will complete the image recognition. The training algorithm of convolution neural network uses BP algorithm to train the convolution neural network supervised. Convolution neural network LeNet-5 model has achieved good recognition effect in handwritten numeral data set, which is widely used in handwritten character recognition in commercial

TEST Engineering & Management
July-August 2020
ISSN: 0193-4120 Page No. 5238 - 5246

field. At this stage, convolutional neural network is gradually applied to face detection and face recognition.

3.3. Extensive research phase

Since Geoffrey Hinton emphasized the strong feature learning ability of multilayer neural networks in Science in 2006, many researchers have devoted themselves to the research of multilayer neural networks including convolutional neural networks. In 2012, krizhevsky proposed the convolutional neural network AlexNet model, which won the championship in the image recognition task of Imagenet large-scale visual recognition competition with great advantages, making convolutional neural network become a hot topic in academic and industrial circles. At present, convolutional neural network has been applied to many fields, such as target detection, video classification, face recognition, behavior recognition, crowd density estimation, image quality evaluation and so on.

4. Related concepts

4.1. Mainstream convolution neural network model

Among many convolutional neural network models, the most classic is the AlexNet model proposed by Hinton team in 2012. The AlexNet model has attracted extensive attention of researchers, which became the champion of Imagenet large-scale image recognition competition in 2012. After that, many researchers have developed more complex convolutional neural network models based on AlexNet model, such as Vggnet, Google net, RESNET and so on. The AlexNet model structure diagram is shown in Figure 1.

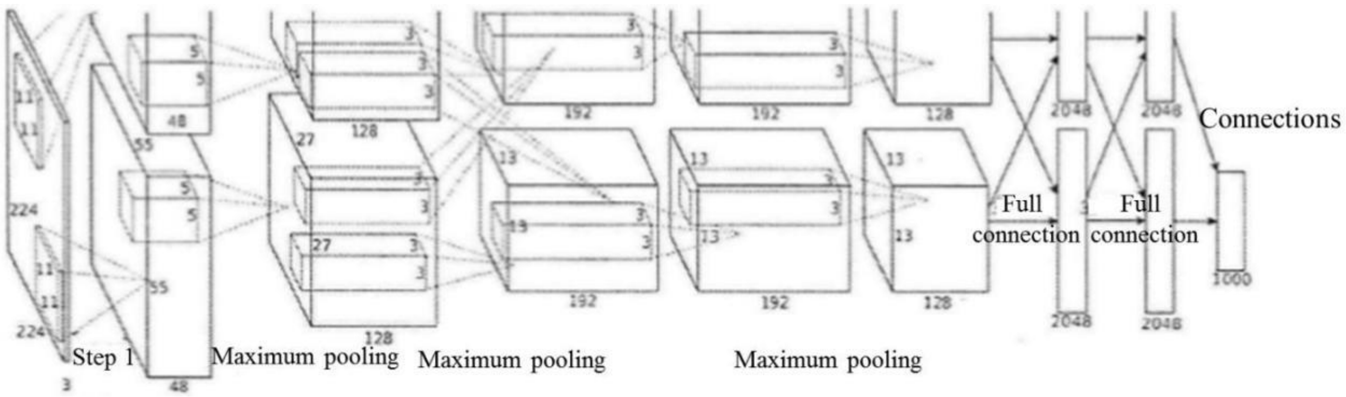


Figure 1. The AlexNet model structure diagram.

TheAlexNet model is a deep convolution network model designed by Alex krizhevsky, which is far ahead of the traditional one. There are eight learnable layers, including five convolution layers and three full connection layers. In training, the activation function of AlexNet model is relu, which greatly shortens the learning cycle of the model. The speed and efficiency of AlexNet model are improved, which is about 6 times faster than that of tank function.

4.2. Pool operation

Real time highway congestion analysis has a huge amount of data information and calculation, which requires us to add pooling layer between convolution layers of network structure. Through the pooling layer, we can improve the overall detection accuracy, including reducing parameters, preventing over fitting, and improving model performance. We can adjust the network structure by maximizing the original image features. By changing the location of the pooling layer, we can improve the network structure of squeezenet, which will achieve higher accuracy. Through the pooling operation of Max pooling, we can sample the data. If the output size of the convolution layer is 4*4, the filter size of the pooled layer filter is 2*2, and the step size is 2. After being processed by pooling layer, the output data size is 2*2, which is reduced to 1/4 of that before

Published by: The Mattingley Publishing Co., Inc.

pooling operation. The size calculation formula of the output feature map after pooling operation is shown in Formula 1.

$$out_{size} = \frac{input_{size}\text{-}filter_{size} + 2\,padding}{stride} + 1 \quad (1)$$

4.3. SSD network model

SSD is a single detector, which can identify multiple objects. By predicting the class fraction and box offset of the fixed set, SSD can be applied to the default bounding box of the small convolution filter for feature mapping. During the training period, SSD only needs one input image and ground block diagram, which can evaluate a small set. In multiple feature maps, the default box with different aspect ratio at each location. For each default box, we can predict shape offsets and confidence levels for all object categories. When training, we can first match the default box with the actual block diagram. By identifying regions of interest, we can discard regions of interest. In addition, the model loss is the weighted sum between local loss and confidence loss.

By using vgg-16-atrous as the basic network, Conv8_2, Conv9_ 2, Conv10_2, Conv11_2 can be used as feature extraction layer. Through target detection, we can select five feature maps for prediction, which is not only carried out on the added feature map, such as conv8_2, Conv9_ 2, Conv10_2, pool _11. The SSD network structure is shown in Figure 2.

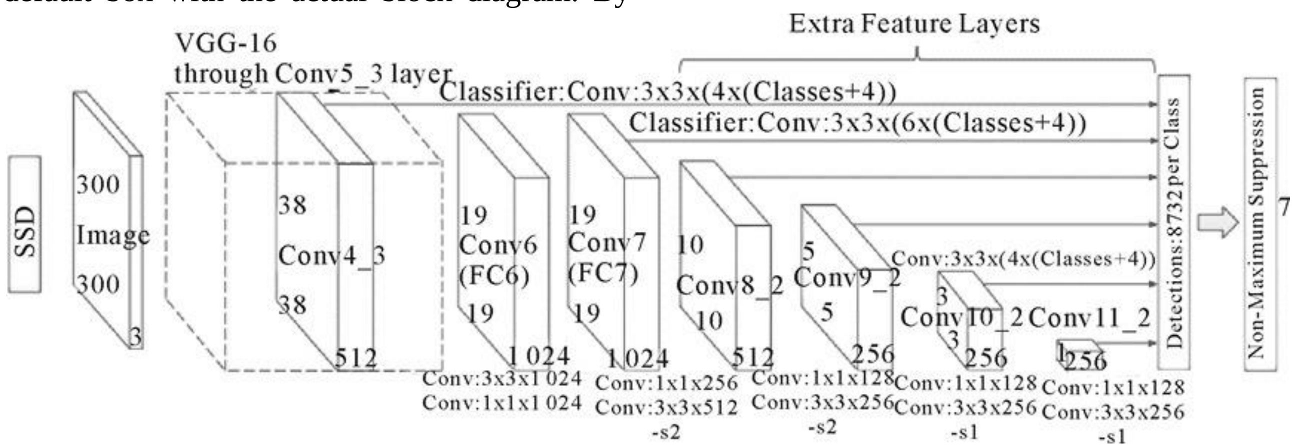


Figure 2. SSD network model.

5. Real time highway congestion analysis process

5.1. Deep learning target detection algorithm based on regression

Through SSD model, we can make convolution neural network to convolute the image, which will generate a series of frames with different sizes and aspect ratios on different levels of feature maps. By predicting the type of objects in the border, we can adjust the shape of the border to fit the size of the object. Finally, it is tested on Pascal VOC, mscoco and ilsvrc data sets, and the experimental results show that the average accuracy is about 75%. Through pre training, we can improve the SSD network model. Through the dataset. Training, we can generate a new network model, which can accurately extract the characteristics of road vehicles and detect road congestion. The basic flow of target detection is shown in Figure 3.

July-August 2020
ISSN: 0193-4120 Page No. 5238 - 5246

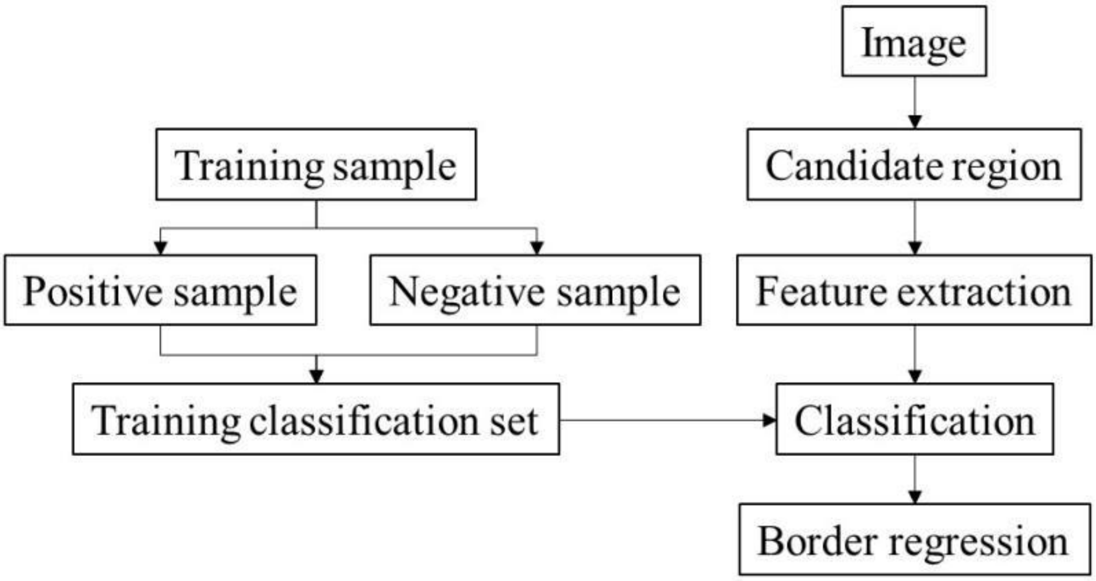


Figure 3. The basic flow of target detection.

5.2. Framework of detection method

Through the lightweight squeezenet network, we can train the image classification model based on vehicle density, which will detect the video. We use interval sampling method to extract video frames for detection, which will record the vehicle density classification state sequence of each lane at each time. By classifying the state time series, we can analyze and judge whether each lane is congested at the current time. The overall detection method flow chart is shown in Figure 4.

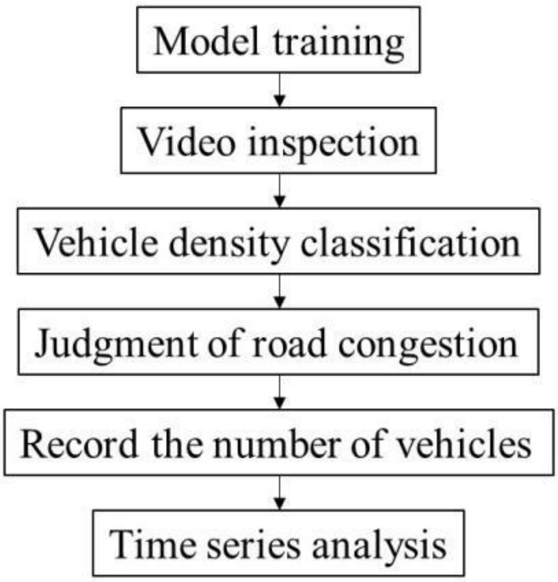


Figure 4. The overall detection method flow chart.

5.3. Vehicle target task construction

By installing cameras at road intersections, we can capture video in real time. Through video streaming, we can detect vehicles at intersections in real time, which is our vehicle target task construction mode, as shown in Figure 5.

Published by: The Mattingley Publishing Co., Inc.

July-August 2020
ISSN: 0193-4120 Page No. 5238 - 5246

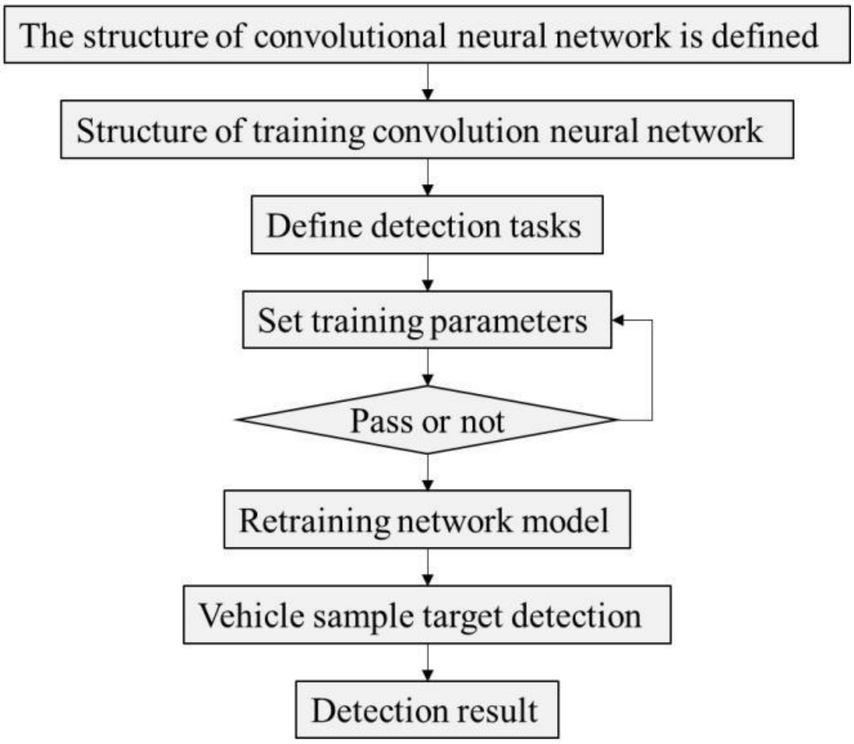


Figure 5. Vehicle target task construction.

5.4. Vehicle density classification

Based on the information of vehicle density data, we can realize image classification by convolution neural network. The main task of vehicle density classification is to classify and detect the vehicle density of each lane at each time. According to the density of vehicles in each lane, eight kinds of labels are made. If the density of the current lane is large and the vehicle covers the whole lane, it is marked as many. If the density of the current lane is low, the vehicle does not cover the whole lane, and there are more uncovered areas behind the road, the mark is less.

5.5. Design of vehicle detection process based on SSD

In this paper, the vehicle detection problem can be regarded as a regression problem, which can design a vehicle detector. Based on the multi-scale feature map, we can predict the SSD target detection flow chart, as shown in Figure 6.

The steps of target detection based on SSD model are as follows.

First, input an RGB image. we can adjust the original image to 300x300x3, where 3 represents its R, G and B channels.

Second, we input the adjusted image into the detection network. Through the convolution layer and pooling layer, we can combine in a certain order to form the main branch of the network. By combining some feature layers in the main branch together, we can form an additional feature layer, which will be used as input and use convolution kernel to do convolution operation in the way of sliding window.

Thirdly, in convolution processing, we can predict the set of boundary boxes with different aspect ratio at each position of several characteristic graphs on different convolution layers. For each default box, we can predict shape offset and confidence for all target categories.

Fourth, the threshold is set according to the predicted bounding box. The detection window whose score is lower than the threshold is discarded, and the redundant window is eliminated by non

Published by: The Mattingley Publishing Co., Inc.

July-August 2020
ISSN: 0193-4120 Page No. 5238 - 5246

maximum suppression algorithm. Finally, we can get the test results.

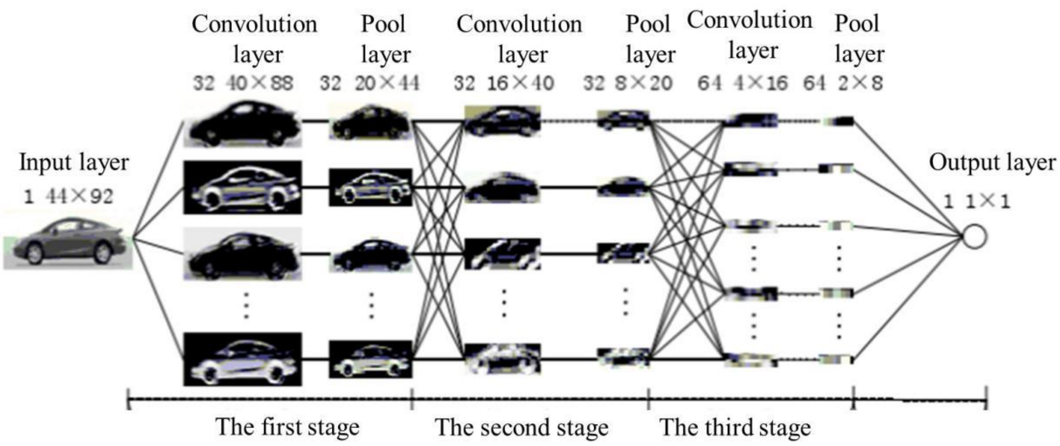


Figure 6. Flow chart of target detection based on SSD.

6. Experimental results and analysis

6.1. Video data model

SSD is a target detection network model, which can detect multiple targets. This paper analyzes the results of a video, and the vehicle detection situation is shown in Figure 7.

Figure 7. The vehicle detection situation.

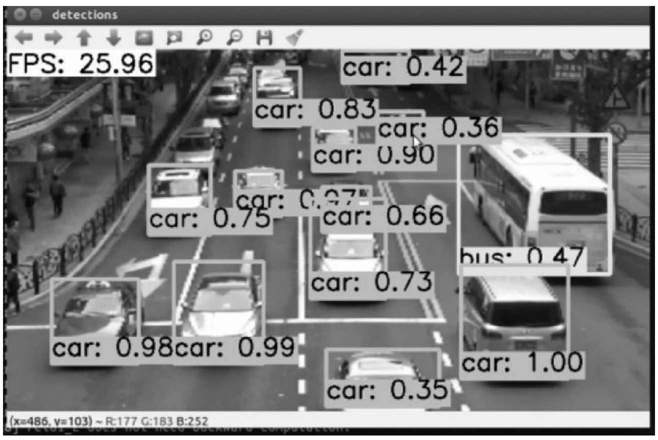
6.2. Classification of results

Through the model, we can classify the detected information, which will mark the specific category and confidence level. By setting two types of detection vehicles and pedestrians in the test, we show the basic results, as shown in Figure 8.

Figure 8. Classification of results.

6.3. Real time speed display

Based on the experimental information, we indicate the category and confidence level, which further shows the speed of operation. The speed is about 26 frames per second, which can meet the requirements of real-time processing. The real-time speed display is shown in Figure 9.



Published by: The Mattingley Publishing Co., Inc.

Figure 9. Real time speed display.

7. Conclusion

The traditional vehicle detection method is difficult to meet the needs of practical application. Based on multi-scale feature map prediction SSD vehicle detection algorithm detection, we can have a good accuracy and speed detection results for large targets. However, SSD model is not ideal for small targets. Through the analysis of SSD target detection algorithm, we design a vehicle detector based on SSD. By training the SSD network model, we can transform the problem of target detection into the problem of classification. Through the analysis of experimental results, we can improve the accuracy and detection efficiency.

Acknowledgments

Jilin Province science and technology department project: research on multi-objective coordinated optimization control of idle start-stop system for intelligent connected passenger cars (Project No. 20200201285JC).

References

[1] Ojala T, Pietikainen M, Harwood D.A comparative study of texture measures with classcification basedon feature distribution [J].Pattern Recognition, 1996, 29:51-59.

[2] Wang, R. C, Lien, et al. Automatic vehicle detection using local features—a statistical approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2008, (1):83-96.

[3] Sivaraman S, Trivedi M M. A General active-learning framework for on-road vehicle recognition and tracking [J]. Nllgn Ranoraon Ym Ranaon on, 2010, (2):267-276.

[4] Wang Xiangxin, Shi Yang, Wen Mei. Acceleration of CNN convolution on mobile GPU [J]. Computer engineering and science, 2018, 40 (1): 34-39.

[5] Zhuang bin, Yang Xiaoguang, Li Keping. Criterion and detection algorithm of road traffic congestion events [J]. Chinese Journal of highway, 2006, 19 (3): 82-86.

[6] Yan Yinhui, Zhou Zhengjin, Xiang Tao. An improved SSD model based on squeezenet and its application [J]. China transportation informatization, 2018 (S1): 189-191.

[7] Gan Wei, Zhang Suwen, Lei Zhen, et al. SIFT feature detection parallel algorithm for mobile intelligent terminal [J]. Computer science, 2016, 43 (S1): 165-167.

[8] Tang Yiping, Huang Leilei, Yan hangchen, et al. Research on lightweight holographic road traffic state visual detection [J]. Computer science, 2014, 41 (5): 308-314.

[9] Yuan Bin, Zhang Yong. Traffic congestion detection algorithm based on image texture analysis [J]. Journal of Shanghai Institute of shipping science, 2015, 38 (4): 77-81.

[10] Zhang Hui, Wang Kunfeng, Wang Feiyue. Application progress and Prospect of deep learning in target visual detection [J]. Acta automatica Sinica, 2017,43 (8): 1290-1304.