Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 5, No. 2, 2023

Based on YOLO v3 Target Recognition Algorithm as Vehicle Tracking Algorithm Analysis

Zili Zhang *

Sydney Smart Technology College, Northeastern University, Qinhuangdao, 066004, China
*** Corresponding author Email:** zzl17731725321@hotmail.com

Abstract: Traditional traffic information acquisition and acquisition are mainly implemented by sensors, and these traditional acquisition and acquisition systems have some great drawbacks. However, with the popularity of traffic monitoring, computer vision technology gradually has a platform foundation that can be applied to identify and track traffic conditions. In this paper, through the research of traditional and Deep learning-based multi-target recognition algorithms and two common multi-target tracking algorithms, a solution of YOLO v3 network combined with deep-sort algorithm is proposed. In this paper, a video of traffic information of urban roads is directly collected for areas with relatively large traffic flow. Interval frames are extracted from the video data set to make relevant data sets for training and verification of YOLO v3 neural networks. Combined with the test results, an open source vehicle depth model dataset is used to train the vehicle depth feature weight file, and Deep-SORT algorithm is used to achieve the target tracking, which can realize the real-time and more accurate multi-target recognition and tracking of moving vehicles.

Keywords: Vehicle Identification; Vehicle Flow Monitoring; Multi-target Tracking; Yolov3; Deep-Sort.

1. Introduction

In the study of intelligent transportation system, it can be found that due to the demand for road target recognition, the collection, acquisition and recognition processing of main subject information such as vehicles and pedestrians on the road have become an extremely important subject. The process of collecting and acquiring traffic information in the past is mainly based on sensor systems, including wave frequency acquisition system and magnetic frequency acquisition system. These traditional data acquisition and acquisition systems have some major shortcomings in design, and both methods are easily interfered with by other factors, resulting in less accurate results, and their maintenance costs are also high. With the emergence of advanced image acquisition equipment, computer vision technology has obtained a large number of data sets and has been rapidly developed. At the same time, it has become more and more widely used in the field of intelligent transportation system. In addition, in the case of actual traffic video, because most of the vehicles themselves are in a moving state, it is very necessary to study the vehicles in a moving state. At the same time, the vehicles in most traffic videos appear in parallel and exist at the same time. The requirements for real-time robustness and accuracy are high. However, a large number of current algorithms cannot meet this requirement. Therefore, the research of real-time recognition and tracking algorithm based on the actual situation of multi-target vehicles is of great significance for the automatic extraction of road traffic information and operation status management.

The purpose of this paper is to study how to use the appropriate target recognition algorithm to identify the moving vehicle in the actual traffic video in the city, so as to achieve the requirement of obtaining the vehicle target information with high accuracy and fast. Based on the current research on major traffic problems and intelligent transportation systems, the acquisition of such information will effectively determine whether a vehicle is speeding or violating regulations in specific traffic conditions, and can determine and track the specific location of illegal vehicles [1]. To a large extent, it can realize the intelligence of the traffic field.

2. Research Status at Home and Abroad

With the development of intelligent optimization algorithm, deep learning and other technologies, relevant technicians began to focus on the application of this technology in the field of computer vision. With the efforts of many scholars, object recognition technology in traffic has also been vigorously developed and widely applied to all aspects of people's lives [2].

Based on the analysis of modern traditional moving target recognition algorithms in China and abroad, it is found that the main research suggestions of researchers are basically to carry out practical application operations on how to introduce some new operators on the basis of fully considering the relevant theories of some motion recognition algorithms introduced above. Therefore, the traditional algorithms are improved to solve the difficulties in practical application research. For example, Kim et al. They put forward a research proposal that uses Harris operator to replace the concept of sports field in traditional optical flow field algorithm [3]. Using this calculation method can effectively reduce the calculation amount of traditional optical flow method and greatly improve the efficiency of target recognition. German scholars Bernd Kitt and Benjiamin Ranft [4] put forward a research proposal that combines extended Kalman filtering with optical flow field calculation. This method can be used in the recognition algorithm of automatic monitoring system to identify multiple targets at the same time. Minoh M put forward the three-frame difference method, which is improved based on the frame difference method. Gan Minggang combined edge recognition and three-frame difference for target recognition, and the recognition results

were more accurate. Lin Mingxiu et al. proposed a research method that first fused the fused inter-frame difference and background difference, and then added gray correlation analysis to determine the target state.

However, for this topic, when the target to be identified is in a moving state, there may be some cases such as occlusion, so there are some problems and shortcomings in the above methods, resulting in the above algorithm is not particularly suitable for target recognition in this case. In order to solve the above problems, this paper introduces the target recognition algorithm based on neural network, which does not need to consider the relative motion state of the target and the camera under the driving environment and the occlusion, etc., and directly recognizes the vehicle target.

The 1980s, with the development of computer hardware technology, video tracking technology entered a relatively rapid stage of development. Wang Qing, et al, they put forward a research proposal of integrating multidimensional hybrid Gaussian model and interframe difference algorithm to realize an adaptive background modeling method [5]. In the early 21st century, search algorithms began to become the main research objects of target recognition and tracking algorithms, such as mean filter, particle filter and Kalman filter [6,7], and researchers began to consider the use of relevant filtering methods to predict the motion state of the target. By making corresponding prediction of target position, the prediction result was generated, and then some optimization operations were performed on the prediction result, to be used for target tracking. Then the error rate of the detector is given for the tracking results between different frames and the tracking is carried out continuously.

When the target to be tracked is in a moving state, the methods such as interframe difference and background difference can not accurately judge the coherence of the target between the front and back frames, so this topic needs to use a different method to achieve. In this paper, on the basis of target recognition, I will introduce the concept of Deep learning and use Deep-SORT algorithm, so as to achieve the purpose of real-time target tracking.

3. Main Research Contents and Methods

3.1. Main Research Contents

(1) Completed the overall construction of the moving vehicle recognition and tracking system based on video. Demand analysis and overall architecture design of the video-based moving vehicle recognition and tracking system are carried out. In the process of vehicle target recognition, a suitable target recognition algorithm for the vehicle in the traffic video of the research object of this paper is given by comparing various target recognition algorithms, and related analysis is carried out. In the process of vehicle target tracking, by comparing various target tracking algorithms, the appropriate target tracking algorithm for the vehicle in the traffic video of the research object of this paper is given, and the relevant analysis is carried out.

(2) Aiming at the demand of accuracy and real-time in traffic video, a solution of YOLO v3 network combined with Deep-SORT algorithm is proposed. The YOLO network is used as the baseline network, the network is trained with self-collected data set, and the vehicle tracking algorithm is combined to obtain the vehicle detection results that can be used to initialize the tracking algorithm. After that, the vehicle detection results obtained by the above algorithm and Deep-Sort are combined to realize vehicle tracking.

(3) Using self-collected data set, the video duration of the test set is 32s, and the training set includes 1166 pictures for training and verifying the target recognizer based on YOLO network; The open-source vehicle depth feature dataset is used to train the model, and the depth feature weight file of Deep-SORT tracking algorithm is obtained. Then on the basis of the weight file, the final measurement is raised

The scheme has the advantage of being more accurate.

3.2. Main Research Methods

First of all, by consulting relevant books and literature, learn the relevant knowledge of deep learning, install programming software and related function libraries, and complete the preliminary preparation for design.

Then, the data set is made, the input video image is scaled normalized, the neural network is trained and tested, and the vehicle recognition function is realized. The vehicle target recognition frame identified by YOLO v3 is compared with the Kalman filter prediction frame from the previous frame, and the Markov distance and minimum cosine distance of the feature vector between them are calculated respectively, and fused into a correlation matrix. The tracking prediction frame and the current frame identification frame are matched using the Hungarian algorithm. If the matching fails, it is considered that the time T from the last matching to the target exceeds the set critical value, then the track is deleted, otherwise the track is maintained. If the match is successful, it is considered that the target is in the recognized state and has not been lost, which can be used as the final result

4. System Implementation and Result Analysis

4.1. Data Set Collection and Processing

The data set used in this project mainly includes two parts. The first part is the traffic data set collected from the middle part of the South Second Ring Road of Xi 'an City, which is mainly used for the training and testing of YOLO v3 neural network. The collected vehicle data set is processed using labeling, and the vehicles are divided into five categories: car, taxi, minivan, truck and motorcycle, so as to facilitate and better identify the vehicle target. After the annotation is completed, the corresponding XML document is obtained. That is, Pascal VOC standard format annotations. In this format, the first column is the label name, the second column is the relative coordinate of the center point x, the third column is the relative coordinate of the center point y, the fourth column is the relative coordinate of the width, and the fifth column is the relative coordinate

The high relative coordinate, in this case, corresponds to the reading of *xyxy in the YOLO v3 object identifier in Chapter 3.

For this part of data set, it is divided into training set, verification set and test set according to training and identification requirements, in which the training verification set accounts for 80% of the total data set, the test set accounts for 20% of the total data set, 80% of the training verification set is the training set, and 20% is the verification set. After a trainable data set is obtained, the YOLO v3 target recognizer is trained using this data set.

Then Deep-SORT can be trained based on this part of the data set, so as to achieve the functional purpose of tracking

vehicles.

4.2. Implementation and Effect Analysis based on YOLO v3 Identifier

4.2.1. Implementation Process

Based on the design of YOLO v3 identifier in 3.2 and the data set for YOLO v3 target identifier in 4.1, the yaml file of data needs to be configured and copied from yolov3/data/ to the data set cars file. After opening, modify the corresponding image path, as well as nc (number of labels) and name (label name). At the same time, based on the universality and vastness of the coco dataset originally used by YOLO, and containing a certain number of vehicles

According to this paper, transfer learning method can be used to train the vehicle identification system. Transfer learning methods can migrate depth and features, and can improve and enhance the broad performance of the model well, even for large data sets, and grow with a change in the number of layers n where a parameter is fixed, for two tasks with little similarity. The transfer distance between them increases faster than the transfer distance between two tasks with high similarity. The more similar the two sets of data, the worse the effect of deep feature transfer training.

In the process of implementing transfer learning method, it is necessary to modify the nc number of the model file that you want to transfer training. Here, based on the performance comparison diagram of each YOLO v3 pre-training model in Figure 2-6, the accuracy and speed are not bad, and the deployment of relatively lightweight yolov3m.pt file is actually the preferred weight file for this project.

In the train.py source file given in the yolov3 source code, the main augment types are as follows, and their meanings are given here one by one:

--img unified input image size

--batch Number of input images for each network training. The minimum performance is 1 # --epochs training times

--data the relative location of the data yaml file

-- Location of the cfg model yaml file

--weights Position of the pre-trained model

--device Training device (CPU or GPU)

Under the objective condition that the GPU of this project is NVIDIA GeForce GTX 1050 Ti 4G, certain compression processing is carried out on the original data set, and -IMG 640 and --batch 4 are selected as the input image scale. After 50 training sessions, it took 7.25

The training will be completed in an hour, and its effect analysis will be given in 4.2.2.

After completing the training of YOLO v3 based target recognizer, using the single frame image of the test data set as the original image, using PowerPoint for video synthesis, thus synthesizing video for target recognition test. Specific effects will be given in 4.2.2.

4.2.2. Effect Analysis

In the process of training, the whole process can be visualized through the wandb library. The GPU is used for training. Figure 4-4 and Figure 4-5 show the CPU and GPU utilization during training. As can be seen from Figure 4-4 CPU status during YOLO training and Figure 4-5 GPU status during YOLO training, the CPU utilization reaches 44%, the overall status is stable, the GPU utilization reaches 16%, the memory usage is 2.9/4.0GB, and the overall status is stable.

In the field of multi-object recognition, mAP is used as the evaluation index. After completing the training, we can get the figure of PRC (Precision-Recall Curve) as shown in FIG. 1 below.

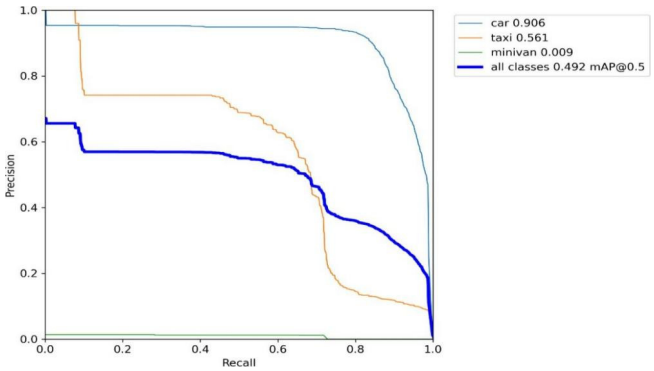

Figure 1. Precision-Recall curve

At the same time, the mAP value of the training process can also be obtained, and the change trend of the value can be seen from the training results figure 2 below.

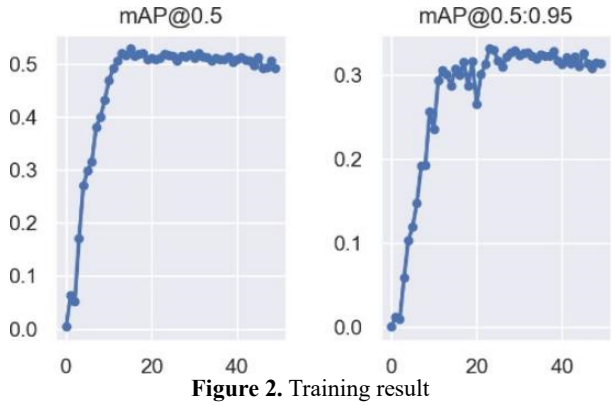

Figure 2. Training result

It can be concluded that the final training result is mAP@0.5 and 0.5567, mAP@0.5: 0.95 is 0.4919. It can be seen that mAP@0.5 is still relatively higher than mAP@0.5 of the official YOLO v3 project, which shows that the training effect is considerable. In view of the simplicity of the data set and the condition of the experimental development environment, the training effect can be achieved with 50 training times. This weight file can be used as the weight file of the object recognition function of this subject, and meet the requirements of traffic video.

4.2.3. Effect Analysis

Firstly, the effect of extracted vehicle depth feature was analyzed. During the completion of training, the function values of weight_loss, triplet_loss and total_loss of this depth feature remained stable at about 1.8, 0.58 and 4 respectively, indicating good training effect. Figure 3, 4, and 5 shows the changes of the weight_loss, triplet_loss, and total_loss loss functions.

At the same time, the image of the classification accuracy change process can also be obtained, as shown in Figure 6. We can clearly see that the final accuracy is infinitely close to 1, which shows that the accuracy is also a high degree.

Finally, the video-based vehicle identification and tracking system designed and implemented in this subject is tested. This test is based on the pre-trained YOLO v3 vehicle identification weight file Best-pt and Deep-SORT vehicle tracking weight file mars.pt to identify and track video vehicles in the test set. Figure 7 shows the effect.

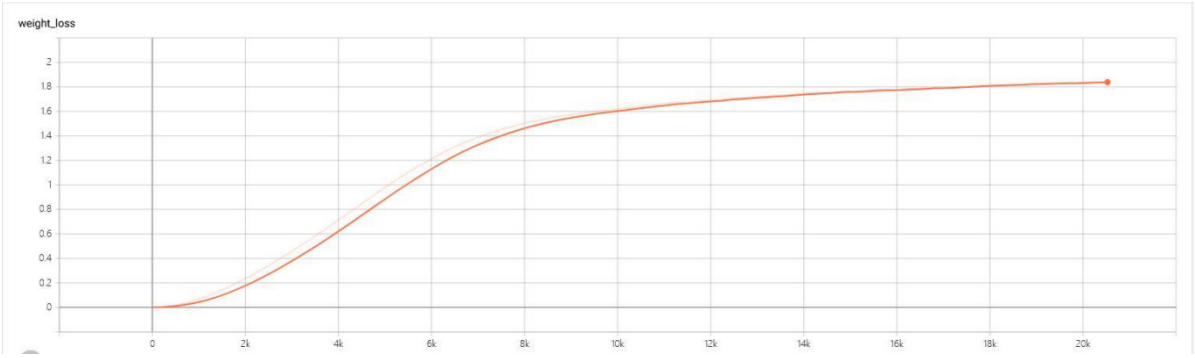

Figure 3. Changes in the weight_loss function for tracker depth feature training

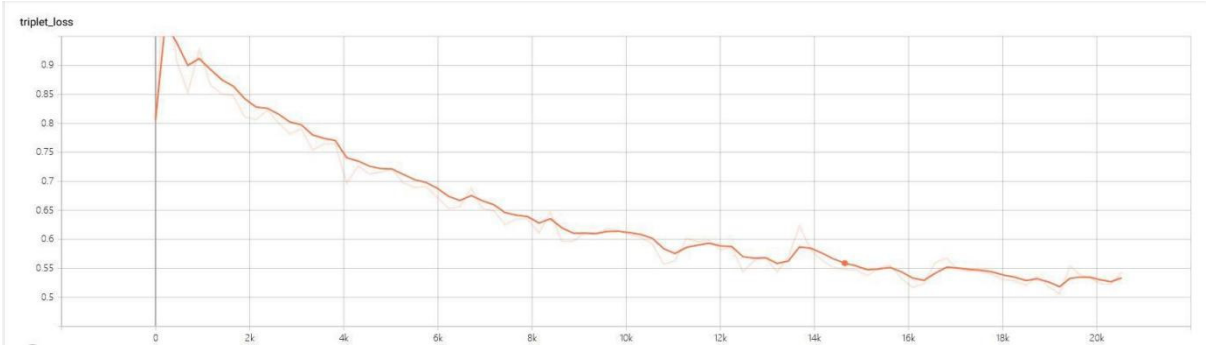

Figure 4. Changes of the triplet_loss function during tracker depth feature training

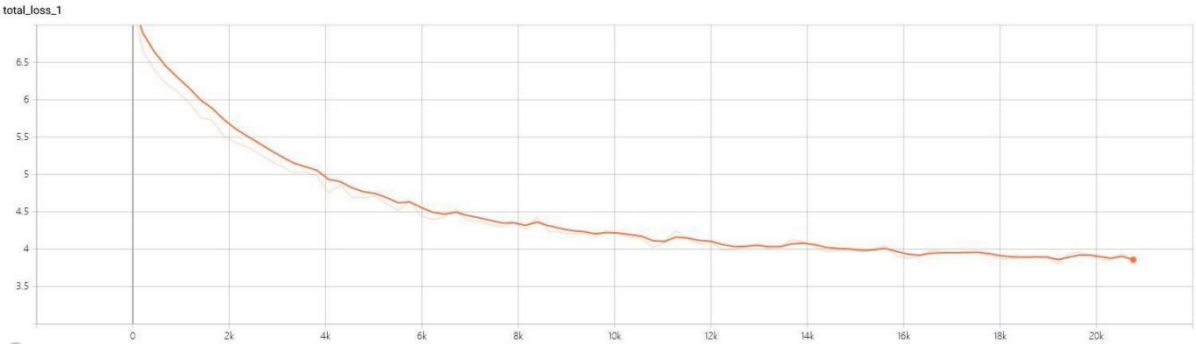

Figure 5. Change of total_loss function in depth feature training of the tracker

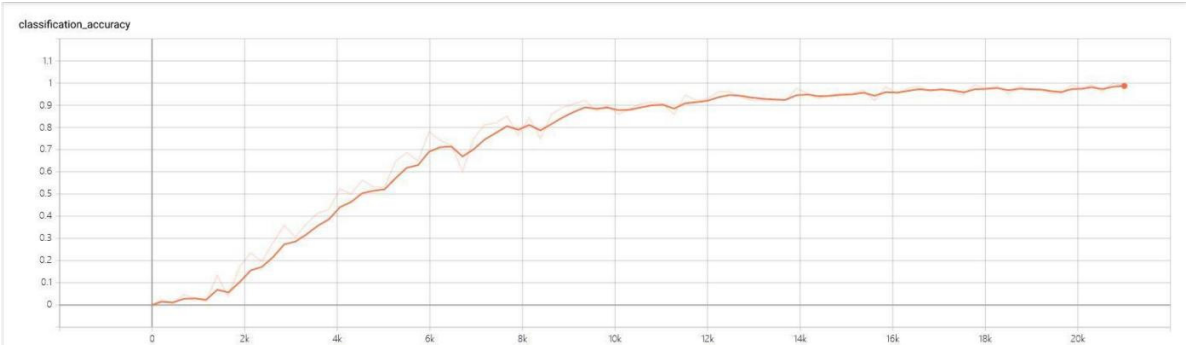

Figure 6. Changes in the classification_accuracy function of tracker depth feature training

Figure 7. Effect of video-based vehicle identification and tracking system

The box in the figure represents the target recognition box generated based on the target recognition, and the Arabic digit label represents the ID in the Deep-SORT algorithm. It can be seen that all vehicles with target identification boxes can be assigned ids to complete the tracking effect.

5. Conclusion

The urban traffic data set is used to train the target recognition neural network, and the vehicle appearance data set is used to train the depth feature neural network required for target tracking. Based on the trained weights, the target recognition system and the overall tracking system are respectively realized, and the effect analysis is given respectively. The verification achieves the purpose that the system designed in this subject can effectively identify and track vehicles in traffic scenes.

This paper has conducted in-depth research on vehicle and tracking based on traffic scenarios and achieved certain results. However, for complex and flexible traffic scenarios, the technical route and design system proposed in this paper still need further research and improvement, mainly in the following aspects. The data set used in this project is the traffic data set actually collected, with a very limited number, with a total of more than 1700 photos. The data set is limited to traffic video streams in daytime and cloudy days, and does not take into account other weather conditions such as sunny, rainy and snowy days and nighttime conditions. The data set used in this project still has certain limitations in real-time performance, which has certain defects for the whole system. The actual data set only reaches 6-8 FPS, which will lead to certain fluency in the playback of the recognition video, so it can be seen that there is a certain phenomenon of frame drop. At the same time, mAP values are heavily influenced by this aspect of the data set. It is necessary to further consider the use of better equipment to collect more real-time image data sets. The experimental equipment used in this subject is relatively backward, and there are certain calculation errors and time efficiency limitations in the training process. It can be considered to further use professional desktop equipment as training equipment to solve this problem.

References

[1] Wu, X. (2018) Research on video-based vehicle recognition and tracking. Thesis of Chang 'an University.

[2] Gong, J., Ji, S. (2017) From photogrammetry to computer vision. Journal of Wuhan University (Information Science). 42(11):21-5+118.

[3] Kim, J., Kim, D. (2010) Moving object detection under free-moving camera. 2010 IEEE International Conference on Image Processing. 4669-72.

[4] Kitt, B. (2010) Detection and tracking of independently moving objects in urban environments. 2010 International IEEE Conference on Intelligent Transportation Systems. 1396-401.

[5] Wang, Q., Chen, F., Xu, W., et al. (2012) Object Tracking via Partial Least Squares Analysis. IEEE Transactions on Image Processing, 21(10):4454-65.

[6] D R. (2013) An algorithm for tracking multiple targets. IEEE Transactions on Auto-matic Control, 24(6):843-54.

[7] Kalal, Z., Mikolajczyk, Mata, J.K. (2012) Tracking-learning-detection. IEEE Transaction on Pattern Analysis and Machine Intelligence, 34(7):1409-22.