Real-Time Panoramic Multi-Target Detection Based on Mobile Machine Vision and Deep Learning

Anzhou Wen
Changchun University of Science and Technology China
w782579376@gmail.com

Abstract. Research on computer vision technology based on deep learning has become an important research direction in the field of artificial intelligence. Among them, the target detection task of images and videos has occupied a pivotal position in many intelligent vision research and applications. The purpose of this article is to learn from the existing object recognition and target detection technology foundation, to conduct in-depth research on the current traffic patrol deficiencies, to achieve real-time patrol and monitoring of key roads. This article chooses to use the target detection algorithm of Faster R-CNN in deep learning, applies drones and mobile machine vision to traffic patrol inspection, and detects various targets in the PASCAL VOC data set. The AP of the bus is 68.20%; The AP of the car is 75.50%; The moving object recognition and tracking method based on Faster R-CNN and Hungarian matching method is verified, and traffic flow data can be calculated to realize real-time monitoring of road traffic. The research in this paper helps to overcome the defects of human eye recognition in the traditional monitoring system and improve the efficiency of traffic inspection.

Key words: Deep Learning, Panoramic Multi-Target, Real-Time Detection, Mobile Machine Vision

1. Introduction
In computer vision, multi-target detection in video sequences is the purpose of detecting area of interest, it is a video target tracking, video segmentation, and other follow-up study foundation work [1-2]. With the development of panoramic vision sensor technology, the panoramic vision system is more and more applied in the field of intelligent video surveillance. Based on static image target detection task currently has made great progress, its detection precision and speed were obvious ascending [3-4]. But in video surveillance, vehicle assisted driving, robot research and other fields, video target detection has more extensive and practical application needs. The recognition of moving objects in video plays a key role in the follow-up research of target classification, target tracking and behavior understanding in intelligent video surveillance technology [5-6].

In the world, with the widespread deployment of video surveillance system, the related fields of video surveillance have been widely studied in recent years. It is becoming increasingly important to obtain information from video surveillance systems, especially the acquisition of content information by video surveillance systems in some specific environments [7-8]. The rapid development of economy, the urbanization process of everywhere to also more and more fast, followed by various problem
increasingly prominent, the traffic problems become the focus of the cities have a headache problem. Increase in vehicles, narrow streets, traffic congestion problems to be solved. Traffic inspection, planning, management and command guidance method also needs to optimize, to adapt to the new era. Mobile machine vision is a kind of portable visual image processing method, on the road traffic, the traffic in the test the traffic state identification. Mobile machine vision can make up for the defect of traditional inspection methods, does not depend on capital equipment, can be flexible to regional testing and monitoring, reduce the amount of monitoring equipment, greatly reduces the cost of equipment, is a supplement to perfect the existing traffic system inspection method of [9-10].

In order to obtain real-time road traffic status information, this paper has carried out research on road inspection methods with mobile machine vision. Take advantage of the flexibility and maneuverability of drones, combined with the characteristics of dynamic perception of mobile machine vision. A moving target recognition and tracking method based on Faster R-CNN and Hungarian matching method is proposed, which can be used to realize real-time monitoring of road traffic after verification.

2. Urban Road Traffic Detection Method Based on Deep Learning

2.1. Target Detection Algorithm Based on Faster R-CNN
Aiming at the airborne blurred image, a secondary fuzzy image evaluation algorithm based on improved grayscale variance is proposed, which can better evaluate the quality of the clear image. Assuming that the resolution of the initial image is an image of size m*n, \( f(x_i, y_j) \) represents the pixel at position \( (i, j) \), and the grayscale variance of the initial image is:

\[
SMD_0 = \sum_{i=1}^{n} \sum_{j=1}^{m} \left( f(x_i, y_j) - f(x_i + 1, y_j) \right) + \left| f(x_i, y_j) - f(x_i, y_j + 1) \right| \tag{1}
\]

In the counterclockwise direction by 90 degrees object motion, motion 30 pixels. Motion blur uses a blur operator. After obtaining a blurred image, its gray scale score \( SMD_t \) is calculated to calculate the difference \( TSMD \) between the two images.

\[
SMD = \sum_{i=1}^{n} \sum_{j=1}^{m} \left| g(x_i, y_j) - g(x_i + 1, y_j) \right| + \left| g(x_i, y_j) - g(x_i, y_j + 1) \right| \]

\[
TSMD = SMD_0 - SMD_t \tag{2}
\]

Faster R-CNN is a target detection algorithm developed based on CNN. For each image to zoom in, scale for long side is not more than 1000, short edge is not more than 600; Image pixels between 0-255, the mean of image pixel is 0. In this paper, the model pre-trained by VGG16 on ImageNet is used. The learning rate of the first four layers is 0. After feature extraction, the image is sampled 16 times. In RPN layer, Anchor mechanism is proposed, including three proportions and three sizes, namely 128, 256, 512, a total of nine Anchor types. Rois is also provided to RoIHead as a training sample. RoIHead in RPN given above 2000 candidate box are classified and the return of the location parameters, selected 128 sample_rois, USES the RoIPooling pooling all these different size of the area on the same scale (7 * 7).

2.2. Moving Target Recognition and Tracking Method Based on Faster R-CNN and Hungary Matching Method
The core of the Hungarian algorithm is to find augmented paths, it is a kind of with augmented path to calculate the maximum matching algorithm of binary chart. Recognition algorithm is divided into three steps: The first step, image preprocessing. To improve the efficiency of operation, the inspection before every image is compressed, the aspect ratio is 640 x 480, the transformation of color images to grayscale. The second step, the image input trained Faster - R - CNN network model, detect the target location. The third step, tracking moving targets. On the premise of known motion target location, the use of the
Hungarian matching algorithm of moving target with labels, localization of the moving target tracking can be achieved.

3. Experimental Environment and Experimental Data
Experimental data is an important reason for affecting network performance. In this paper, the commonly used PASCAL VOC data set of Faster R-CNN is selected, which includes 20 categories such as person, dog, bus, and car. On the basis of the original more than 9,000 training pictures, continue to add road traffic pictures captured by drones. The pictures are framed from the shooting video, about 1000, and are marked in the format of VOC data set, label The name of the picture is indicated. The resolution range is between 320*180 and 1442*1080. The shooting angles are different from top to bottom. The occlusion pictures between the targets are also included, which fully simulates various situations that may occur in actual shooting. The four-rotor UAV used in this article is the simplest mechanism in the multi-rotor, and it is a friendly carrying platform. It is more stable than other aircraft, simple and easy to maintain, and relatively refined in software and easier to control. As shown in Table 1, it is the relevant environmental information of the experiment in this article.

Table 1. Experimental environment

| Name       | Content                        |
|------------|--------------------------------|
| CPU        | Intel Core i5                  |
| GPU        | Nvidia GTX 960                 |
| RAM        | 64G                            |
| System     | Windows 7                      |
| Learning framework | Tensorflow 1.4          |

4. Simulation results and analysis

4.1. Analysis of Target Detection Algorithm Based on Faster R-CNN
In this paper, the model that Vgg16 trained on ImageNet is selected as the feature extraction network. VOC and the increased data set are used for transfer learning to fine-tune the training of the network. After 30,000 iterations, the training is stopped. The resulting average accuracy (AP) is shown in Table 2 and Figure 1.

Table 2. Average accuracy of various targets

| Numbering | Target class  | Accuracy | Numbering | Target class | Accuracy |
|-----------|---------------|----------|-----------|--------------|----------|
| 1         | Acroplane     | 66.32%   | 11        | Diningtable  | 62.21%   |
| 2         | Bicycle       | 72.20%   | 12        | Dog          | 68.75%   |
| 3         | Bidd          | 59.37%   | 13        | Horse        | 78.03%   |
| 4         | Boat          | 47.22%   | 14        | Motorbike    | 70.64%   |
| 5         | Botle         | 38.93%   | 15        | Person       | 70.86%   |
| 6         | Bus           | 68.20%   | 16        | Pottedplant  | 36.31%   |
| 7         | Car           | 75.50%   | 17        | Sheep        | 61.90%   |
| 8         | Cat           | 74.00%   | 18        | Sofa         | 58.33%   |
| 9         | Chair         | 40.03%   | 19        | Train        | 72.14%   |
| 10        | Cow           | 68.03%   | 20        | tv/monitor   | 65.63%   |
It can be seen that the detection accuracy of certain types of targets varies greatly, one of the reasons is that the sample sizes of various types of targets included in the training data set are inconsistent, which makes the test results have a large gap. Judging from the relative position and size of the bounding box and the detection target, the detection results are good, indicating that using Faster R-CNN to detect road traffic has certain significance.

4.2. Analysis of Moving Target Recognition Results Based on Faster R-CNN and Hungarian Matching Method

In this paper, for the recognition of airborne moving images, the scene is aerial photography of the hovering drone above the endless highway. The characteristic of this scene is that the background is static and the target is a moving vehicle. Therefore, the multi-target tracking method is used to detect the target based on Faster R-CNN, and then the Hungarian matching algorithm is used to locate and track the moving target to complete the detection of the vehicle. Identify. The sample is taken from aerial video of a certain section of the road by drone. A total of three sections of video are taken from different angles under sufficient light conditions. The first step of the experiment is to compress the video to an aspect ratio of 640×480 and convert it to a grayscale image. Input each frame of image into the trained network model, identify the moving vehicles, and then use the Hungarian matching algorithm to match the label to each target, so that the moving vehicles are identified from the image sequence. After recognizing the vehicle, after recording the trajectory of each vehicle, it can be judged that the vehicle has passed the virtual line, and the vehicle count value of the lane is incremented by one. Vehicle count value. The length of the lane in the video is selected as the vehicle speed measurement interval. After calibration, the actual physical length can be obtained from the video. Through the quantitative relationship between the appearance frame and the disappearance frame of the vehicle, the passing time of the vehicle can be obtained, and then the average speed of the vehicle can be calculated. The final calculation results are shown in Table 3.
Table 3. Traffic flow calculation results

| Name           | Content  |
|----------------|----------|
| Left lane PCU  | 9        |
| Right lane PCU | 3        |
| Speed km/h     | 36.7682  |
| Density PCU/KM | 34.0012  |

After experimental verification, the actual case is tested, and the moving target recognition and tracking method based on the Faster R-CNN and Hungarian matching method is verified, which can check and identify moving vehicles in road traffic, and then calculate traffic flow data, realize real-time monitoring of road traffic.

5. Conclusion

Target detection is an important research and application hotspot in the field of computer vision. It comprehensively uses knowledge of many disciplines such as image processing, pattern recognition, artificial intelligence and machine learning. It promotes intelligent traffic management, automatic video surveillance, unmanned driving technology, and lay the foundation for advanced human-computer interaction and other aspects. Based on existing open source data sets and existing Internet resources, this paper proposes a mobile machine vision deep learning target detection network model based on quadrotor drones. Innovative method of moving target recognition based on Faster R-CNN and Hungarian matching method, and proved its effectiveness through experiments, applied to real-time traffic monitoring.

References

[1] He Y, Mendis G J, Wei J. Real-Time Detection of False Data Injection Attacks in Smart Grid: A Deep Learning-Based Intelligent Mechanism[J]. IEEE Transactions on Smart Grid, 2017, 8(5):2505-2516.

[2] Wang R, Liang Y, Xu J W, et al. Cascading classifier with discriminative multi-features for a specific 3D object real-time detection[J]. The Visual Computer, 2019, 35(3):399-414.

[3] Grimm D A P, Gorman J C, Stevens R H, et al. Demonstration of a Method for Real-time Detection of Anomalies in Team Communication[J]. Proceedings of the Human Factors & Ergonomics Society Annual Meeting, 2017, 61(1):282-286.

[4] Siyu S, Yuchao L, Xueming L, et al. Optimization of 3D Surfaces of Dextran with Different Molecule Weights for Real-Time Detection of Biomolecular Interactions by a QCM Biosensor[J]. Polymers, 2017, 9(12):409.

[5] Stergiopoulos G, Chronopoulou G, Bitsikas E, et al. Using side channel TCP features for real-time detection of malware connections[J]. Journal of Computer Security, 2019, 27(2):1-14.

[6] Zhongyuan Chen, Yueting Peng, Xiaoxue Xie, 等. Dendrimer-Functionalized Superparamagnetic Nanobeacons for Real-Time Detection and Depletion of HSP90α mRNA and MR Imaging[J]. Theranostics, 2019, 9(20):5784-5796.

[7] Ji X, Zhang G, Chen X, et al. Multi-Perspective Tracking for Intelligent Vehicle[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(2):518-529.

[8] Meng W, Huang Y. Panoramic multiple-moving-target recognition and tracking for autonomous-navigation agricultural vehicles[J]. Agro Food Industry Hi Tech, 2017, 28(1):2225-2230.

[9] Tomislav Medič, Holst C, Kuhlmann H. Towards System Calibration of Panoramic Laser Scanners from a Single Station[J]. Sensors, 2017, 17(5):1145.

[10] Nguyen H T, Langstrand J P, Hildebrandt M. ReClass: An Application of Deep Learning for Real-time Reading Detection Using an Eye Tracking Device[J]. Proceedings of the Human Factors & Ergonomics Society Annual Meeting, 2018, 62(1):307-310.