Japan Society of Civil Engineers 2020 Annual Meeting

VEHICLE DETECTION AND COUNTING FOR VEHICLE TYPES OF MOTORCYCLE AND AUTO RICKSHAW USING DEEP LEARNING

Nihon University Student Member ○Daichi Tashiro
Nihon University Regular Member Tetsuhiro Ishizaka
Nihon University Regular Member Atsushi Fukuda

1. INTRODUCTION

The accuracy rate of the vehicle counting by the conventional traffic sensor in the Paldi intersection of Ahmedabad city, India, is around 60%. This low accuracy is caused that most vehicle of the auto rickshaw and motorcycle is not detected with high accuracy. This insufficient rate should be improved in aspect to examine an appropriate traffic demand control in Ahmedabad. Recently, vehicle detection on image processing has been developing and achieved good performance. But it has not been examined and demonstrated for auto rickshaw and motorcycle in Ahmedabad city.

Therefore, the object of this study is to verify whether the vehicle detection and counting of motorcycle and auto rickshaw is performed using deep learning for the picture of the CCTV camera in Ahmedabad.

2. LITERATURE REVIEW

Sekiya[1)] measured the traffic measurement of a motorcycle and the car using image processing for the purpose of the road maintenance corresponding to the rapid increase of the motorcycle in Indonesia. Konnno[2)] applied the recognition technique of the general object by the deep learning, and detected the transportation modes such as a Southeast Asian tuk-tuk or the auto rickshaw and clarified the distribution in the city of these vehicles.

But, both of them did not investigate the complex traffic volume such as a car, a motorcycle, and auto rickshaw.

Therefore, in this study, inspect the complex traffic volume such as car, motorcycle, auto rickshaw, bus, and truck that were not worked on in the literature review by using deep learning

3. METHOD OF STUDY

3.1 Measurement procedure of the traffic volume

This study employs the technique called Convolutional Neural Network or CNN which has used in a lot of field on the image recognition. CNN is broadly classified into two categories "one-stage detector" such as YOLO,SSD and "two-stage detector" such as R-CNN. Generally speaking, "one-stage detector" is not good at accuracy of the detection and "two-stage detector" is not good at speed of the detection.

As a solution of these weakness, this study use epoch-making model called "Retina Net". This model's feature is unique loss function called "Focal Loss". This function divides into background or target as an object. The classified target is focused on learning to identify object. Therefore, the Retina Net is expected to obtain higher accuracy and speed.

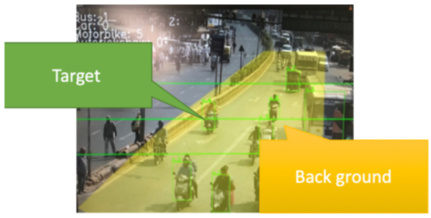

Fig. 1 "Focal Loss" role

The detection model with the Retina Net detects a vehicle and classifies vehicle type into five types (car, bus, motorbike, auto rickshaw and truck) from an image of CCTV. To classify these vehicle types, the model need feature of each vehicle type. So, we taught the model feature each vehicle type (generally called "annotation") like Fig.2.

Fig.2. Annotation

The vehicle counting was improved from conventional method to consider unique traffic condition at the Paldi intersection and low angle of CCTV camera.

The Paldi intersection has the vehicle queue form stop line generated by high traffic demand at peak time. So, accuracy of detection is probably very low because occlusion over vehicles making it difficult to classify into background or object. To solve problem in this study, the detected vehicle which run the part surrounded with yellow area at figure 3 is tracked and counted on one of three green lines. Even if the first line which drawn in nearest side of stop line at intersection is occupied by stopping vehicles, the second or third line could help to identify object for vehicle counting. Finally, the detected vehicles were counted as traffic volume by each vehicle type.

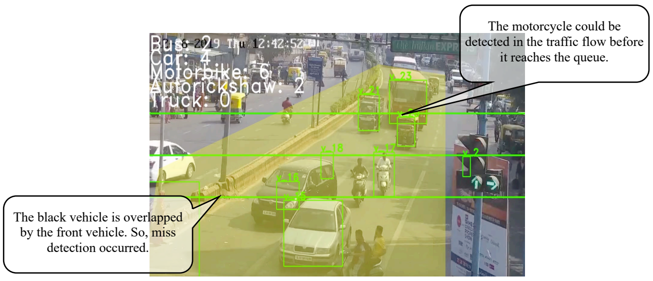

Fig. 3 Vehicle measurement model

3.2 Video outline

In this study, we use CCTV camera video on the Paldi

Keywords: Image Processing, Vehicle Detection, CNN, Traffic Volume
Contact address: Narashinodai 7-24-1, Funabashi city, Chiba, 274-0063, Japan, Tel: +81-47-469-5335

intersection which was recoded at 12:00 pm to 1:00 pm on December 26. The counted traffic volume manually is limited vehicle run the part surrounded with yellow area at fig. 3. The same condition was applied on the model counting. And this video was applied on not only the training data but also verification data due to limited of recorded video data.
Finally, the count traffic volume by the developed model is compared with manually counted it.

4. RESULT

The traffic volume counted by model and manual are shown in the table 1 and 2. And the accuracy is shown in the table 3. The traffic volume of auto rickshaw and motorcycle is 81.3% in traffic volume of all type. That is why motorcycle and auto-rickshaw is an important factor to effectively control the intersection.

Table 1 Manual Counted Volume

type / time	car	motorbike	auto rickshaw	bus	truck	total
12:00~12:20	91	483	181	33	8	796
12:20~12:40	112	421	191	34	20	778
12:40~13:00	99	404	160	22	3	688

Table 2 Model Counted Volume

type / time	car	motorbike	auto rickshaw	bus	truck	total
12:00~12:20	45	204	85	25	3	362
12:20~12:40	53	156	102	17	14	342
12:40~13:00	58	218	92	9	4	381

Table 3 Accuracy of Traffic Volume by Vehicle Type

type / method	car	motorbike	auto rickshaw	bus	truck	total
model count	176	702	312	56	24	1270
manually count	302	1308	532	89	31	2262
accuracy	58.3%	53.7%	58.6%	62.9%	77.4%	56.1%

The accuracy of our model is around 60%. As the percentage shown in this table, motorbike and auto rickshaw are recognized as well as other vehicle type and it means the model has the ability to count the volume of motorcycle and auto-rickshaw. When desired accuracy of vehicle detector used in Japan however considers that it is approximately 98% or more, the result of our model was not achieved well.

It is difficult for a model to detect a vehicle under the influence of the large vehicles such as buses definitely like Fig. 4. a big vehicle such as a bus overlapped behind vehicle and is a cause of low accuracy. In many times, the vehicle queue length used to reach the third line. Therefore, all of three lines was occupied by the stopping vehicles and the detection is not gained well because of vehicle occlusion. In the Paldi intersection, CCTV was installed at the corner of intersection because it should cover all direction. Of course, CCTV can shoot video just under the CCTV pole. But the angle condition of CCTV cannot cover direction traffic volume.
In the motorbike, the auto rickshaw, problems include the point where a lot of vehicles slipping to a stop line through a left turn traffic lane were founded.

Fig. 4 The situation of detection looks difficult

5. CONCLUSION

This study developed and improved the vehicle detection and counting model using Retina Net as a deep learning technologies. And it was also configured for localizing unique traffic condition and CCTV camera setting. The detection and counting of motorcycle and auto-rickshaw was achieved as the same performance of normal cars. But the total accuracy over vehicle type is still short of the desired value. Especially, although this study improved the counting lines, it doesn't contribute the accuracy well to meet the desired value in the case of long queue and blocked bus.
In further study, the training data should be increased and the verification data should be used as not same with the training data.

REFERENCES

1)関谷浩孝，上坂克巳，橋本浩良，高宮進：「画像処理による交通量計測手法」についてのインドネシア公共事業省道路研究所との共同研究，土木技術資料，Vol, 55，No.7，pp.24-27，2013
2) 金野貴紘，荒井勇輝，屋井鉄雄：R－CNN による途上国を対象とした車両検出方法に関する研究，土木学会論文集（土木計画学），Vol，74，No.3，pp.193-202，2018
3) Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar : Focal Loss for Dense Object Detection, ICCV 2017, 2017