Vehicles and Pedestrians Detection by Using Deep Transfer Learning

Ei Phyu Myint
University of Computer Studies,
Yangon, Myanmar
eiphyumyint@ucsy.edu.mm

Myint Myint Sein
University of Computer Studies,
Yangon, Myanmar
myint@ucsy.edu.mm

***Abstract* — This paper presents a comparison of two customized models for object detection which are generalized by fine tuning the pre-trained models. The already trained models used are Faster RCNN with Inception V2 and Single Shot Detector with MobileNet V2. The images are obtained by using the PoE-supported IP camera and raspberry pi. As a result, the mean average precision (mAP) values of Faster RCNN with Inception V2 and Single Shot Detector with MobileNet V2 based on transfer learning technique are 0.556 and 0.267. So, Faster RCNN with Inception V2 customized model gives better detection result.**

Keywords — object detection, transfer learning, fine tuning

I. Introduction

Object detection is an important and challenging task in computer vision. It includes the tasks such as recognizing and classifying every object in an image and localizing each one by drawing the appropriate bounding box around it. Along with the rise of facial detection, autonomous vehicles, smart video surveillance and various people counting applications, fast and accurate object detection systems are rising in demand.

Recently, deep learning has substantially advanced the object detection field. By using deep learning, features and contents of various objects from digital images can be automatically determined. But deep learning requires a very large amount of dataset. By using transfer learning, the problem of insufficient training data can be solved. The training data and test data are not required to be identically distributed in transfer learning. In addition, the model in target domain does not need to train from scratch. In this way, the demand of training data and training time in the target domain can be significantly reduced. Figure 1 shows the transfer learning process.

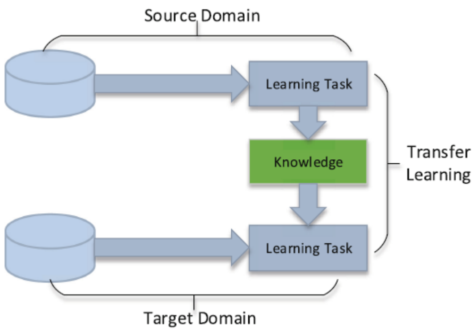

Fig. 1. Transfer Learning based learning process [1]

Chuanqi Tan et al. [1] presented a survey on Network-based, Instances-based, Mapping-based and Adversarial-based deep transfer learning categories. Nobuyoshi Yabuki et al. [2] utilized the transfer learning technique based on Single Shot Detector with VGG-16 model to detect the objects in construction sites and disaster areas. Zhen Zeng et al. [3] presented transfer learning based license plate recognition system by using pre-trained Xception model. Ajeet Ram Pathak et al. [4] presented about the deep learning applications for object detection including with frameworks and services, benchmarked datasets, state-of-the-art approaches, etc. Xinrui Zou [5] presented the comparison of traditional machine learning techniques and deep learning techniques.

The research question for this paper is: Which of the existing models is the best for this application in the field of object detection? Two models for object detection based on fine-tuning the pre-trained models are compared to answer this question.

II. Objectives

The main objective of this research is how to apply the existing deep learning models for an application rather than creating a new one. It gives us an effective result without requiring high computational resource.

III. Methodology

The object detection models are created by using transfer learning technique based on pre-trained models. Tensorflow object detection API is used to identify and detect objects in images. As shown in figure 2, steps are as follows:

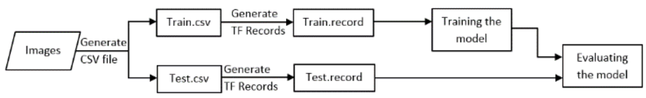

Fig. 2. Basic Flow Diagram

A. Gathering and Labeling of Data

A simple and good tool called labelImg is used to label these images. As shown in the figure 3, creating a rectangle box on the target objects is the process of labeling the images. Six labels are created. These labels consist of five vehicles and pedestrian.

Fig. 3. Labeling images by using LabelImg tool

B. Creating TFRecords files

Tfrecord files for both train and test images are needed to generate after labeling all the images and saving their corresponding XML files. Firstly, the .csv files from the XML files are created. And then, tfrecord files are created from these .csv files according to the labels such as bicycle, bus, car, motorcycle, person and truck.

C. Training and Evaluating the model

At first, a label map file which maps an id to class name is created. The training process uses transfer learning technique which uses an already trained model to train on custom data. Faster RCNN with Inception V2 and Single Shot Detector with MobileNet V2 models are fine tuned based on the latest checkpoint that has been pre-trained on the COCO dataset. Finally, the inference graphs are generated based on the saved summary files with the highest step number after the models are successfully trained. In this way, these inference graphs can be used by applications that want to run these customized models.

IV. RESULTS AND DISCUSSION

Training dataset contains 881 images and testing dataset contains 378 images. The training processes are done over five thousands iterations. To evaluate the accuracy of these models, COCO evaluation metric is used. A mean Average Precision (mAP) evaluation metric is used when evaluating object detector performance. This value shows that how accurate the detection is over the images. Table 1 presents the mAP for each of the customized models.

Table 1. Mean Average Precision of Different Customized Models

Customized Model	**mAP**
Faster RCNN with Inception V2	0.556
Single Shot Detector MobileNet V2	0.267

Table shows that Faster RCNN with Inception V2 customized model is more precise than Single Shot Detector with MobileNet V2 customized model in detecting objects for this application. A few of the results produced by these models are shown in figure 4 and figure 5.

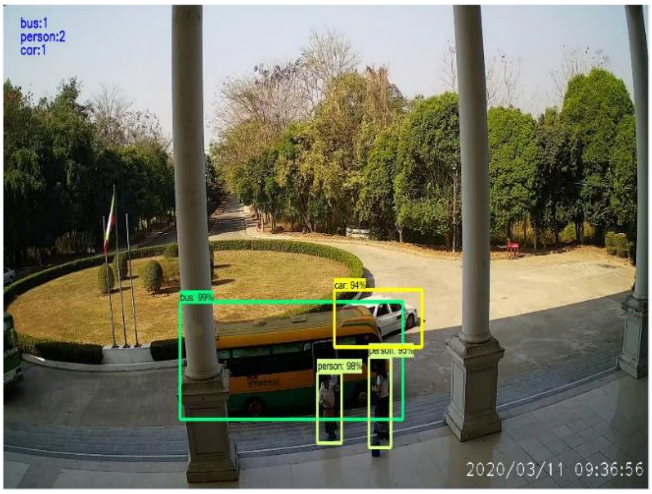

Fig. 4. Detection Result of Faster RCNN with Inception V2 customized model

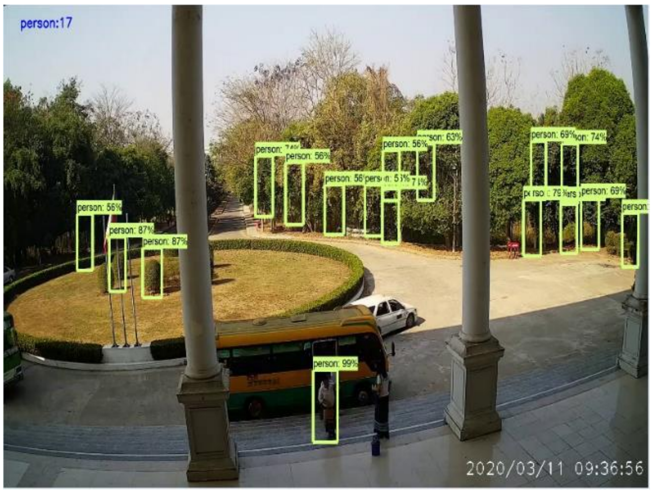

Fig. 5. Detection Result of Single Shot Detector with MobileNet V2 customized model

V. CONCLUSION

The deep transfer learning technique is implemented by using already trained models such as Faster RCNN with Inception V2 and Single Shot Detector with MobileNet V2 models. The application is implemented by using PC with Intel i5 processor. The results suggest that fine tuning the Faster RCNN with Inception V2 model gives better results than fine tuning the Single Shot Detector with MobileNet V2 model according to the mAP.

REFERENCES

[1] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang and C. Liu, “A Survey on Deep Transfer Learning”, Artificial Neural Networks and Machine Learning – ICANN 2018.

[2] N. Yabuki, N. Nishimura and T. Fukuda, “Automatic Object Detection from Digital Images by Deep Learning with Transfer Learning”, Advanced Computing Strategies for Engineering, 2018.

[3] Z. Zeng, P. Gao and S. Sun, “License Plate Recognition System Based on Transfer Learning”, International Conference on Signal and Information Processing, Networking and Computers, ICSINC 2018, pp. 42-49.

[4] A. Ram Pathak, M. Pandey and S. Rautaray, “Application of Deep Learning for Object Detection”, International Conference on Computational Intelligence and Data Science ICCIDS, 2018.

[5] X. Zou, “A Review of Object Detection Techniques”, International Conference on Smart Grid and Electrical Automation, ICSGEA, 2019.