A New Method for Pedestrian Detection with Lightweight Backbone based on Yolov3 Network

Qirui Dong*

KangChiao International School East Campus

Abstract: The main purpose of YOLOv3, aiming to improve the detection speed and accuracy from current detection models, is to predict the center coordinates of (x, y) from the Bounding Box and its length, width through multiple layers of VGG Convolutional Neural Network (VGG-CNN) and uses the Darknet lightweight framework to process images at a faster speed. More specifically, our model has been reduced part of YOLOv3's complex and computationally intensive procedures and improved its algorithms to maintain the efficiency and accuracy of object detection. By this method, it performs a higher quality on mass object detection tasks with fewer detection errors.

1 Introduction

In the development of the modern automobile industry, automotive automation or Advanced Driver-Assistance Systems (ADAS) has been in the popular field. In recent years, various manufacturers have presented their research results. By supporting with sensor equipment, many novel features have appeared in cars. From the earliest Adaptive Cruise Control (ACC) features to the Front Collision Warning (FCW) feature and now to Tesla's autopilot series. Everyone who is involved in this field to devote their fantastic ideas to the world.

However, current autopilot functions are very limited. Based on the Automation levels which are defined by SAE J3016, most of these "smart" vehicles are still in the L1 or L2 stage. Although these vehicles are able to support drivers in some fields, most of them work within little success under current traffic circumstances. In other words, current automation technologies in vehicles are mainly based on human drivers, the driver still needs to monitor the surrounding environment at any time. They provide partially automatic assisted driving, which can realize automatic steering, automatic acceleration, and deceleration, etc.

Our research is based on the third version of You Only Look Once (YOLO) series, a detection framework by Redmon J. et. al. research theory focused on pedestrian detection in the traffic road. Compared with other object detection framework, YOLO series reach a faster processing speed to detect objects because it has one input stage as their detection strategy. On the basis of the classification network, they removed the full connection layer but added the convolution layer and the upsampling layer.

2 Method

We establish an object detection model modified by YOLO-based Convolutional Neural Network (CNN) family of models for object detection. In this model, it improves from the most recent variation version YOLOv3, by replacing CNN structure to the lightweight backbone of VGG11. As the figure 2 shown, our invention could be specific described as following 5 steps:

Step 1. This model uses INRIA and Penn-Fudan datasets to specifically simulate pedestrian objects on the road. Before training, the model initials pedestrian detection weights by pre-training configurations from ImageNet dataset.

Step 2. Randomly divided each dataset into training sets and testing sets. Setting training super parameters and net's parameters, inputs training set to the net.

And put uploaded weights configuration model to calculate and store the training weights iteratively.

Step 3. While testing, the object detection model resolves the input images by VGG-11 convolution neural network and obtains multiple scales of feature by passing through convolutional layers and upsampling layers.

Put images which needed to be identified into CNN, obtain multiple scales of feature by passing through several convolutional layers and upsampling layers.

Step 4. After model calculated weights parameters, the logical function shall activate these values by IoU threshold judgement to new parameters of the Bounding Box.

Step 5. By the final calculation results, the program will produce the target detection box and recognition results through non-maximum value suppression processing.

This method is a subset feature of Advanced Driver-Assistance Systems (ADAS), it modified from YOLOv3 object detection model. This feature could improve the detection speed and accuracy of the images in the vehicles to identify pedestrians on the road traffic circumstance, especially in coping with unexpected situations.

3 Experimental verification

Our pedestrian detection model is based on lightweight backbone with VGG11 convolutional neural network, which variance from YOLOv3 real-time object detection system. To improve the processing speed, we replace the Darknet-53 CNN to VGG-11 CNN in our model. Doing by this, it significantly cut down the processing speed than the original model. Now in this lightweight model, it reaches 50fps processing speed under 608 image size of object detection in our testing experiment environment.

To reach this experimental result, we first use transfer learning skills to initial pre-trained pedestrian detection weights by ImageNet dataset. Then, we divide INRIA and Penn-Fudan datasets by a certain percentage of train set and test set to learn the pedestrian detection weights iteratively.

In our experiment, we use NVIDIA GeForce GTX 1080 as our testing environment. Under this circumstance, our training weights could afford around 50 fps processing speed within a moderate detection accurate for 608 image size. Also, we test original YOLOv3 model in our testing environment, it only reaches around 20fps processing speed within a moderate detection accurate. Thus, we assume our model achieve a great success specific in processing speed field.

4 Conclusion

Compared with processing speed, VGG 11 CNN model processes images in a faster speed than YOLOv3 under the 608-pixel image size. And, Titan X is an advanced GPU platform than NVIDIA GeForce GTX 1080. Thus, we conducted that our model reaches a faster processing speed than original YOLOv3 CNN model for resolving 608-pixel image size figure.