Research on Object Detection Based on Road Safety Considerations by Deep Learning Techniques

Cheng-Fa Tsai [*] Shu-Huei Huang Hsiu-Ting Kung

National Pingtung University of Science and Technology, Department of Management Information Systems

Abstract

This paper proposes a convolutional neural network called MYOLO, which is based on the convolutional neural network YOLO and fine-tuned to correctly distinguish speed bumps and rumble strips. Develop an object detection model that is most suitable for detecting speed bumps and rumble strips for self-driving cars, and establish a complete vehicle dynamic analysis system. It is hoped that the object detection model of the speed bumps and rumble strips of the self-driving car can be constructed by employing deep learning techniques, so as to establish a complete vehicle dynamic analysis system to avoid damage to the vehicle chassis and shock absorbers, resulting in high vehicle maintenance costs for car owners. It is expected that by increasing the functionality and economic value of self-driving cars, the safety of self-driving cars may be improved, and the road safety of pedestrians can be enhanced. The main purpose of road obstacles is to remind car drivers to reduce their driving speed. Road obstacles such as speed bumps and rumble strips can provide safety for passers-by. However, their disadvantages are that they may cause damage to the vehicle chassis and shock absorbers. Advanced driving assistance systems have gradually become popular with the development of self-driving cars. However, nowadays advanced driving assistance systems have the functions of detecting speed bumps and rumble strips, but they are extremely rare. This work first collects images of speed bumps and rumble strips at two different resolutions, and then utilizes various well-known deep learning algorithms, such as SSD, CenterNet, Yolov3, and so forth, to detect speed bumps and rumble strips. A series of experimental results demonstrate that after fine-tuning the YoloV3 model, the accuracy rate of

[*] Corresponding author: Cheng-Fa Tsai e-mail: cftsai@mail.npust.edu.tw

identifying the speed bumps and rumble strips is the highest, which can be as high as 95%.

Keywords: Deep Learning, Object Detection, Speed Bump, Rumble Strips

1. INTRODUCTION

In the development of information technology, artificial intelligence, machine learning and deep learning may be the most deeply felt and most frequently heard terms for the progress of technology in recent years. Deep learning technology is evolved from neural networks (Bochkovskiy et. al, 2020; He et al, 2015). Nowadays, deep learning has become a very popular and popular application and research topic in many applications (Shah et al., 2019). It has been proved to be extremely prominent and successful in various application fields, such as image analysis, speech recognition and text recognition (Ren et al., 2016). In general, deep learning uses supervised and unsupervised strategies to learn multi-level representations and features in hierarchical structures and is used for informative work in classification, speech recognition, and image recognition. In recent years, many deep learning technologies have been widely used in industrial-related fields (Tan et al., 2019; Yan et al., 2020), and researchers have also actively developed several new deep learning models (Tang et al., 2019; Wang et. al, 2020; Yu et. al., 2018), which can be applied to manufacturing intelligence, component failure prediction, factory monitoring and management, and so forth (Yun et. al, 2019; Zhou et al., 2019).

Artificial intelligence and even the recent deep learning technology have achieved many practical results in vehicle manufacturing and intelligent vehicle driving (Tang et. al., 2019). In the driving of intelligent vehicles, the setting of road obstacles is to hope that the driver will reduce the driving speed. The advantage of setting road obstacles is that it can protect the safety of passers-by, such as setting up speed bumps and rumble strips, but the disadvantage is that it may cause damage to the vehicle chassis and shock absorbers. Moreover, the maintenance cost of the vehicle must be spent. Today's advanced assisted driving systems are becoming more and more common with the development of autonomous vehicles. This system usually combines a variety of sensors such as cameras, radars, signal transmitters and receivers to assist the driver of the car to drive more safely, for the sake of future safety. The arrival of the era of autonomous and driverless cars is pre-paving the way (Tang et. al, 2019).

2. MATERIALS AND METHODOLOGY

The purpose of this study can be illustrated as follows. Firstly, assist in the study of self-driving car selection based on road safety considerations and propose the most suitable object detection model for detecting speed bumps and rumble strips to establish a complete vehicle dynamic analysis system. Secondly, avoid damage to the chassis or suspension system caused by the inability to accurately identify the speed bumps and rumble strips, resulting in excessive maintenance costs. Finally, protect the safety of pedestrians. In this study, the authors construct a convolutional neural network called MYOLO, which is based on the convolutional neural network YOLOV3 (Redmon et al., 2018), and makes some adjustments and improvements. The convolutional neural network mainly processes and learns the features by the convolutional layer, the linear rectified layer, and the pooling layer, and correctly distinguishes the speed bumps and rumble strips to develop self- driving is the most suitable object detection model for detecting speed bumps and and rumble strips, and establishes a complete vehicle dynamic analysis system.

Fig. 1. The speed bumps.

Fig. 2. The rumble strips.

This paper collected images of speed bumps and rumble strips and marks of speed bumps at two different resolutions, and used deep learning object detection algorithm combined with convolutional neural network to detect them. The algorithms utilized are SSD (Liu et al., 2016), CenterNet (Zhou et. al., 2019), Yolov3 (Redmon et al., 2018). Each method trained two image sets of speed bumps and rumble strips at different resolutions, and the object detection model was evaluated when all algorithms were trained.

This paper was divided into two parts for the detection of the target object in the image set. The first part used the deep learning object detection algorithm to train the target object in the image set, and the second part was the target object detection test in the image set of this work.

The operation steps of the deep learning techniques proposed in this research in the object detection model training are as follows:

Step 1: Using the concept of transfer learning, the deep learning object detection model that has been trained on the voc2007 data set is used as the initial weight for each model training in this work. Train the initial weight model of the speed bumps and rumble strips images.

Step 2: The adjustment categories of training model parameters can be divided into adjustment of loss function, adjustment of training times, adjustment of training learning rate, training image and xml file path to record the coordinate position of objects in the image, and model storage after training. The file path is adjusted to set the value of the anchor box.

Step 3: The model reduces the image size according to the set conditions, which is beneficial to the reduction of training time and hardware resources.

Step 4: After the reduced image is binarized, it is sent to the feature extraction network of the model for feature extraction to generate multiple feature layers.

Step 5: The anchor box will appear at the grid points of each feature layer according to the set width and height to classify and calculate the category and position of the target object.

Step 6: The anchor box will update the weight of the model after the calculation of the classification and position of the grid point of each feature layer, such as the loss rate and the learning rate reduction.

Step 7: Check whether the termination condition is met. If the termination condition is not met, go back to step 3.

The operation steps of the deep learning techniques proposed in this research in the object detection model testing are as follows:

Step 1: Load the trained model.

Step 2: Read the test data set of the image set and the coordinate file of the corresponding target object in the image.

Step 3: Extract the relevant feature layers from the images of the test set through the feature extraction network.

Step 4: Use the anchor box to determine the object category for the feature layer and calculate the position of the target object

Step 5: Use NMS (Non-maximum suppression) to select the prediction box with the largest IoU (Intersection over Union) value of the target box from multiple prediction boxes generated from the same object category in the image.

Step 6: Save the recognition result, and determine whether all the images have completed the test. If so, end the test process; otherwise, continue the test.

3. IMPLEMENTATION DETAILS

The experimental environment of this study is illustrated as below. This investigation employed a computer with Intel® Core™ i9-9900K 3.60GHz to 5GHz CPU, 64 GB RAM, and the model is trained on NVIDIA TITAN RTX 24G GPU. The implementation of this work utilized Python 3.7.10, while the part of the training employed the Pytorch 1.7. Furthermore, the computer operating system environment was Ubuntu 18.04.

The relevant collection methods of the image collection of this paper have been mentioned in the previous section. The field environment of the image collection in this paper was divided into two parts: the speed bumps and the speed bumps marking on the road section of the National Pingtung University of Science and Technology in Taiwan, and the plastic material speed bumps and the speed bumps marking on the road section in San Antonio, Texas, USA. Most object detection models do not support target objects with Chinese label names, so this study assigned four target objects to English label names.

The number of original images in the image set of this paper was 3366, and there were

6949 target objects in total. Corresponding to the research purpose of this paper, it is necessary to adjust the image resolution from the original (1920×1080) to (1080×720), and store it as another set of image sets. Due to the insufficient number of image sets, the experimental results may be affected. Therefore, the image enhancement and horizontal flip method of this paper was used as the image enhancement method in the experiment. Finally, the high-resolution and low-resolution image sets were combined with the corresponding horizontally flipped image sets to form an image set with the number of target objects and the number of images doubled.

In this paper, 70% of the high-resolution and low-resolution images were divided into training sets, 20% were divided into validation sets, and 10% were divided into test sets by random numbers. In order to ensure the fairness of model training and testing, the random number division data set was performed twice. Since there are two sets of high-resolution and low-resolution image sets in this work, after two random number data division, a total of four sets of training, validation and test sets will be formed.

To evaluate the performance and robustness of the object detection, the performance of the MYOLO was compared with those of a well-known SSD and CenterNet in terms of execution time, precision, accuracy. This paper used several well-known deep learning object detection models, and each model was trained for 100 epochs, of which the first 50 epochs freeze some architectural layers in the model for training. Three models (SSD, CenterNet, MYOLO) and two sets of test sets were used as evaluation objects for evaluating IoU, precision, recall, F1 score, Mean Average Precision (mAP), and frames per second (fps). In addition, the Adam optimizer and Relu activation function were utilized for SSD, CenterNet to perform object detection. However, the presented MYOLO adopted the Adam optimizer and Mish activation function for detection of speed bumps and rumble strips.

4. EXPERIMENTAL RESULTS

After a series of experiments, it can be observed that the MYOLO (based on YoloV3) model has the highest accuracy in judging and finding the target object as the speed bumps and rumble strips. The mAP value obtained by MYOLO after two tests of high-resolution target objects was 98% and 94%. The mAP values obtained by MYOLO after testing low-resolution target objects twice were 95% and 94%. At the same time, it can be found that the average accuracy of MYOLO in detecting high-resolution was 98% and 93%, and the average obtained when detecting two low-resolution test sets was 93% and 92%, so that the

MYOLO model may be good at object detection tasks. It has always maintained strong performance, and its detection speed achieved 77fps. In brief, The presented MYOLO outperforms numerous well-known deep learning approaches in object detection comparisons, such as SSD and CenterNet.

5. CONCLUSIONS

This paper employed two speed bumps and rumble strips image sets with different resolutions to train three deep learning object detection models, hoping to discover a cost-effective model to offer a driving assistance system that can detect speed bumps and rumble strips at low cost. After a series of experiments, it can be demonstrated that the MYOLO model has the highest accuracy, precision and execution time in judging and finding the target object as speed bumps and rumble strips comparing with SSD and CenterNet techniques.

Acknowledgement: This research was funded by the Ministry of Science and Technology, Republic of China, Taiwan, grant numbers MOST-110-2221-E-020-016.

References

Bochkovskiy A. Wang C.-Y. & Liao H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934.

He K. Zhang X. Ren S. & Sun J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9), 1904-1916.

Liu W. Anguelov D. Erhan D. Szegedy C. Reed S. Fu C. Y. & Berg A. C. (2016, October). SSD: Single shot multibox detector. The European conference on computer vision (pp. 21-37). Amsterdam, Netherlands .

Redmon J. & Farhadi A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Ren S. He K. Girshick R. & Sun J. (2016). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine* *intelligence*, 39(6), 1137-1149.

Shah S. & Deshmukh C. (2019). Pothole and bump detection using convolution neural networks. The 2019 IEEE Transportation Electrification Conference (ITEC-India)(pp1-4) Chennai, India.

Tan M. & Le Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946.

Tan M. Pang R. & Le Q. V. (2020). Efficientdet: Scalable and efficient object detection. The 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10781-10790). Seattle Washington, United States of America.

Tang Z. Naphade M. Liu M. Y. Yang X. Birchfield S. Wang S. & Hwang J. N. (2019). Cityflow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification. The IEEE Conference on Computer Vision and Pattern Recognition (pp. 8797-8806). Long Beach, California, United States of America.

Wang C.-Y. Mark Liao H.-Y. Wu Y.-H. Chen P.-Y. Hsieh J.-W. & Yeh I.-H. (2020). CSPNet: A new backbone that can enhance learning capability of cnn. The IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 309-391). Seattle, Washington, United States of America.

Yu F. Wang D. Shelhamer E. & Darrell T. (2018). Deep layer aggregation. The IEEE conference on computer vision and pattern recognition (pp. 2403-2412). Salt Lake City, Utah, United States of America.

Yun S. Han D. Oh S. J. Chun S. Choe J. & Yoo Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. The IEEE International Conference on Computer Vision (pp. 6023-6032). Seoul, South Korea.

Zhou X. Wang D. & Krähenbühl P. (2019). Objects as points. arXiv preprint arXiv:1904.07850.