VEHICLES DETECTION USING DEEP LEARNING WITH IMPROVED SINGLE SHOT DETECTOR

WONG NGEI SIONG

UNIVERSITI TEKNOLOGI MALAYSIA

VEHICLES DETECTION USING DEEP LEARNING WITH IMPROVED SINGLE SHOT DETECTOR

WONG NGEI SIONG

A project report submitted in fulfilment of the
requirements for the award of the degree of
Master of Engineering (Computer and Microelectronic Systems)

School of Electrical Engineering
Faculty of Engineering
Universiti Teknologi Malaysia

JULY 2022

DEDICATION

This thesis is dedicated to my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

ACKNOWLEDGEMENT

In preparing this project report, I want to thank my family who always by my side and support me throughout the journey. I wish to express my sincere appreciation to my main thesis supervisor, Dr. Mohd Azhar Abdul Razak for encouragement and guidance. Without continuous support and guidance from my beloved family and supervisor, this thesis would not have been the same as presented here.

Next, I appreciate all the comments from my panels of examiners, Dr. Mohd Afzan Othman and Ts. Dr. Zaid Omar. They provide precious suggestion and feedback during the presentation.

Lastly, I appreciate my fellow postgraduate coursemates who provide helps and support when I face some obstacles throughout the journey. We always remind among each other about the important deadline of project report. Besides, their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. Finally, I would like to express my gratitude to all of you who help me throughout the development of this project report.

ABSTRACT

Deep learning is an important element of data science to automate predictive analysis for a computer to detect and classify the objects into different classes based on trained datasets, either through supervised learning, semi-supervised learning, or unsupervised learning. The aim of this research work is to use a deep learning algorithm, improved single shot detector (SSD) which is capable to detect vehicles, so that the proposed algorithm not only achieve fast detection speed, but also achieve high accuracy in object detection. Although there is other algorithm available in the context of deep learning such as You Only Look Once (YOLO) and Faster-Region Based Convolutional Neural Networks (Faster R-CNN), most of them have trade off between accuracy and speed in object detection. The accuracy also degrades when detecting small objects or objects that are further away. Besides, current network models have difficulties in identifying objects by relying solely on the pre-trained datasets, as the traffic participants may vary across cities, with different colours and shapes. Furthermore, training datasets manually with a variety of car models' images would be time consuming due to the huge datasets. Hence, one of the research objectives is to implement mobilenet V2 network architecture on existing SSD network to improve the detection accuracy (mAP, mean average precision), inference time (s, second) and sensitivity towards small objects in complex backgrounds without increasing the computation complexity. The second research objective of this project is to apply transfer learning mechanism for the custom dataset to increase detection accuracy against small objects and reduce training time. In this research, custom datasets are used for training and testing, where the datasets are annotated using labelImg. Google Colab and some open-source libraries, Tensorflow and Keras will be used in model training. The performance of improved-SSD in object detection is evaluated based on inference time (second) and mean average precision (mAP). All models are pretrained using Common Objects in Context dataset (COCO). On top of that, own custom dataset is used to archive better accuracy. Based on the result obtained, 1.76 seconds needed for Faster R-CNN model to perform inference per image whereas 1.24 seconds needed for proposed model to perform the same tasks. The inference time of proposed model is approximate 30% faster than the Faster R-CNN model. The mean average precision of the proposed model is 73.4% whereas the average recall rate of the proposed model is 80%. Besides, the proposed model obtains approximately 10% improvement in terms of mAP detecting small object if compared with Faster R-CNN model. The proposed model able to detect vehicles with shorter inference time and good accuracy. The model shows improvement in detecting small objects and objects that are further away with better accuracy. In short, the trained model can serve as a good starting point for the development of autonomous car.

ABSTRAK

Pembelajaran mendalam ialah elemen penting sains data untuk mengautomasikan analisis ramalan bagi komputer untuk mengesan dan mengelaskan objek ke dalam kelas yang berbeza berdasarkan set data terlatih, sama ada melalui pembelajaran diselia, pembelajaran separa penyeliaan atau pembelajaran tanpa penyeliaan. Matlamat kerja penyelidikan ini adalah untuk menggunakan algoritma pembelajaran mendalam, *single shot detector* (SSD) yang diubah suai yang mampu mengesan kenderaan, supaya algoritma yang dicadangkan bukan sahaja mencapai kelajuan pengesanan pantas, tetapi juga mencapai ketepatan tinggi dalam pengesanan objek. Walaupun terdapat algoritma lain yang tersedia dalam konteks pembelajaran mendalam seperti *You Only Look Once* (YOLO) dan *Faster-Region Based Convolutional Neural Networks* (Faster R-CNN), kebanyakannya perlu membuat pilihan antara ketepatan dan kelajuan dalam pengesanan objek. Ketepatan juga merosot apabila mengesan objek kecil atau objek yang lebih jauh. Selain itu, model pembelajaran mendalam hari ini menghadapi kesukaran dalam mengenal pasti objek dengan bergantung semata-mata pada set data yang sedia ada, kerana peserta trafik mungkin berbeza-beza di seluruh bandar, dengan warna dan bentuk yang berbeza. Tambahan pula, melatih set data secara manual dengan pelbagai imej model kereta akan mengambil banyak masa kerana set data yang besar. Oleh itu, salah satu objektif kajian adalah untuk mengunakan seni bina rangkaian mobilenet V2 pada rangkaian SSD sedia ada untuk meningkatkan ketepatan pengesanan (mAP, *mean average precision*), masa inferens (s, saat) dan kepekaan terhadap objek kecil dalam latar belakang kompleks tanpa melibatkan kompleks pengiraan. Objektif penyelidikan kedua kertas ini adalah untuk menggunakan mekanisme pembelajaran pemindahan untuk set data tersuai untuk meningkatkan ketepatan pengesanan terhadap objek kecil dan mengurangkan masa latihan. Dalam penyelidikan ini, set data tersuai digunakan untuk latihan dan ujian, di mana set data dianotasi menggunakan *labelImg*. *Google Colab* dan beberapa perpustakaan sumber terbuka, *Tensorflow* dan *Keras* akan digunakan dalam latihan model. Prestasi SSD yang diubahsuai dalam pengesanan objek dinilai berdasarkan masa inferens (saat) dan *mean average precision* (mAP). Semua model dipralatih menggunakan set data *Common Objects in Context* (COCO). Selain itu, set data tersuai sendiri digunakan untuk meningkatkan ketepatan. Berdasarkan keputusan yang diperolehi, masa diperlukan untuk model *Faster R-CNN* melakukan inferens per imej adalah 1.76 saat manakala model *mobilenet V2 SSD* hanya mermerlukan 1.24 saat untuk melaksanakan tugas yang sama. Masa inferens model yang dicadangkan adalah lebih kurang 30% lebih cepat daripada model *Faster R-CNN*. *Mean average precision* bagi model yang dicadangkan adalah 73.4% manakala *recall rate* bagi model yang dicadangkan adalah 80%. Selain itu, model yang dicadangkan memperoleh kira-kira 10% peningkatan dari segi pengesanan mAP objek kecil jika dibandingkan dengan model *Faster R-CNN*. Model yang dicadangkan dapat mengesan kenderaan dengan masa inferens yang lebih pendek dan ketepatan yang baik. Model ini menunjukkan peningkatan dalam mengesan objek kecil dan objek yang berada lebih jauh dengan ketepatan yang lebih baik. Ringkasnya, model terlatih boleh berfungsi sebagai titik permulaan yang baik untuk pembinaan kereta autonomi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SSD	-	Single Shot Detector
mAP	-	Mean Average Precision
COCO	-	Common Objects in Context
CNN	-	Convolutional Neural Network
Fast R-CNN	-	Fast Region based Convolutional Neural Network
KNN	-	K-Nearest Neighbors
RPN	-	Region Proposal Network
SVR	-	Support Vector Regression
SURF	-	Speeded Up Robust Feature
HOG	-	Histogram of Oriented Gradients
DoG	-	Difference of Gaussians
PCA	-	Principal Component Analysis
LDA	-	Latent Dirichlet Allocation
Mask R-CNN	-	Mask Region based Convolutional Neural Network
SVM	-	Support Vector Machines
FAST	-	Features from Accelerated Segment Test
RBF	-	Radius Basis Function
HSV	-	Hue Saturation Value
RoIs	-	Regions of Interests
SIFT	-	Scale-Invariant feature transform
VGG 16	-	Visual Geometry Group
IoU	-	Intersection over Union
Faster R-CNN	-	Faster Region based Convolutional Neural Network
AP	-	Average Precision
YOLO	-	You Only Look Once

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Problem Background

A report presented by Standard Law School [1] stated that the root cause of at least 90% of motor vehicle accidents is human error, and a large percentage is due to driver negligence such as speeding, aggressive and reckless driving, distracted driving, drowsy driving, and drunk driving. In the National Motor Vehicle Crash Causation Survey reported to the US Congress [2] that did not state driver negligence as the sole cause of the crash, it was mentioned that human error is still a significant facet in the sequence of events that lead to a crash. Vehicle's detection using deep learning have the potential to reduce fatalities on the road by eliminating human errors to improve the safety and efficiency of the transportation systems [3].

For a vehicle to operate in a complex dynamic environment, it needs to be able to receive the information and react in a timely manner to reach human-level reliability. With the current upturn in object detection in traffic scenes for vehicles, deep learning has seen many breakthroughs in the past years. While the task of object detection may be perceived as basic as to classify and locate objects in the image, even the most powerful approach will face a great deal of challenges, such as complex weather conditions, changeable lighting, complex environment backgrounds and real-time detection.

In this project, object detection with the application of deep learning will be discussed, particularly on the aspect of vehicles. Various object detection algorithm will be studied and reviewed, which involved traditional computer vision (feature-based detection), two-stage detector and one-stage detector. The review will see the application that related works have implemented with these methods in traffic scenarios.

1.2 Problem Statement

Vehicle's detection often refers to detection based on picture or video frames which contains vehicles [4]. The existing methods become less adaptive in specific road scenarios such as bad weather, traffic condition, occlusion occurs on objects, varying road scenes and diverse vehicle characteristics [5].

In ensuring that the object detection model has a high accuracy, the choice of a suitable dataset is critical. The state-of-the-art model has difficulty in identifying objects by relying solely on the pre-trained datasets, as the traffic participants may vary across cities in the forms of shapes, colors, sizes, lighting and more. Furthermore, training datasets manually with a variety of car models would be time consuming due to the huge datasets.

Therefore, there is a need for an improved method to address the issue of vehicle detection accuracy as the current on-road object detection result of the state-of-the-art methods is not accurate in some scenarios. The accuracy also degrades when detecting small objects or objects that are further away.

1.3 Research Aims and Objectives

1.3.1 Research Aim

The aim of this research work is to improve single shot detector (SSD) by fine-tuning hyperparameter and apply image augmentation technique, which is capable to detect vehicles, so that the model not only achieve fast detection speed, but also achieve higher accuracy towards small object in object detection.

1.3.2 Research Objectives

The objectives of the research are :

(a) To improve the existing SSD network model by adjusting certain training parameter to improve the detection accuracy (mAP, mean average precision), inference time (s, second) and sensitivity towards small objects.

(b) To apply transfer learning mechanism for the custom dataset to increase detection accuracy and reduce training time.

1.4 Project Scopes

The scopes of this project are :

(a) Implementation of Improved-SSD for object detection in vehicles with the performance measured by mAP and inference time.

(b) Custom dataset is built and trained with transfer learning mechanism.

1.5 Project Contribution

The sensitivity of the original SSD network toward smaller object can be improved by fine tuning certain hyperparameter [6] and building own custom dataset.

1.6 Report Organization

This report consists of five chapters. Chapter 1 discusses about the research's background, describes the problem statement, aims, objectives, project scopes, and expected project contribution. Chapter 2 elaborates the literature review, covering the background studies on related topics and related works conducted with similar methods or objectives. Chapter 3 presents the research methodology, explaining the project workflow and the algorithms and methods to be implemented. Chapter 4 presents the result and discussion. Lastly, Chapter 5 concludes the report with the summary of findings and ideas on future works.

REFERENCES

1 B. W. Smith, “Human error as a cause of vehicle crashes,” Center for Internet and Society, 18-Dec-2013. [Online]. Available: http://cyberlaw.stanford.edu/blog/2013/12/human-error-cause-vehicle-crashes. [Accessed: 18-Aug-2021].

2 Washington, D.C.: U.S. Dept. of Transportation, National Highway TrafficSafety Administration, 2008.

3 W. Schwarting, A. Javier and R. Daniela, "Planning and decision-making for autonomous vehicles," in Annual Review of Control, Robotics, and Autonomous Systems 1, 2018.

4 J. Cao, C. Song, S. Song, S. Peng, D. Wang, Y. Shao, and F. Xiao, “Front vehicle detection algorithm for smart car based on improved ssd model,” Sensors, vol. 20, no. 16, p. 4646, 2020.

5 W. Zhang, L. Tian, C. Li, and H. Li, “A ssd-based crowded pedestrian detection method,” 2018 International Conference on Control, Automation and Information Sciences (ICCAIS), 2018.

6 Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9. 2015.

7 Deepanshu Tyagi, "Introduction to SIFT (Scale Invariant Feature Transform)," 16-Mar-2019. [Online]. Available: https://medium.com/data-breach/introduction-to-sift-scale-invariant-feature-transform-65d7f3a72d40. [Accessed: 12-Feb-2022].

8 Lowe, David G. "Object recognition from local scale-invariant features." In Proceedings of the seventh IEEE international conference on computer vision, vol. 2, pp. 1150-1157. Ieee, 1999.

9 Truong, Mai Thanh Nhat, and Sanghoon Kim, "A Review on Image Feature Detection and Description," In Proceedings of the Korea Information Processing Society Conference, pp. 677-680. Korea Information Processing Society, 2016.

10 Cheung, Warren, and Ghassan Hamarneh, "N-sift: N-dimensional scale invariant feature transform for matching medical images," In 2007 4th IEEE international symposium on biomedical imaging: from nano to macro, pp. 720-723. IEEE, 2007.

11 Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," In European conference on computer vision, pp. 404-417. Springer, Berlin, Heidelberg, 2006.

12 Deepanshu Tyagi, "Introduction to SURF (Speeded-Up Robust Features)," 20-Mar-2019. [Online]. Available: https://medium.com/data-breach/introduction-to-surf-speeded-up-robust-features-c7396d6e7c4e. [Accessed: 12-Feb-2022].

13 Lowe, David G, "Object recognition from local scale-invariant features," In Proceedings of the seventh IEEE international conference on computer vision, vol. 2, pp. 1150-1157. Ieee, 1999.

14 Deepanshu Tyagi, "Introduction to BRIEF (Binary Robust Independent Elementary Features)," 19-Mar-2019. [Online]. Available: https://medium.com/data-breach/introduction-to-brief-binary-robust-independent-elementary-features-436f4a31a0e6. [Accessed: 12-Feb-2022].

15 Mikolajczyk, Krystian, and Cordelia Schmid, "A performance evaluation of local descriptors," IEEE transactions on pattern analysis and machine intelligence 27, no. 10 (2005): 1615-1630.

16 Hua, Gang, Matthew Brown, and Simon Winder, "Discriminant embedding for local image descriptors," In 2007 IEEE 11th International Conference on Computer Vision, pp. 1-8. IEEE, 2007.

17 Ozuysal, Mustafa, Michael Calonder, Vincent Lepetit, and Pascal Fua, "Fast keypoint recognition using random ferns," IEEE transactions on pattern analysis and machine intelligence 32, no. 3 (2009): 448-461.

18 Onel Harrison, "Machine Learning Basics with the K-Nearest Neighbors Algorithm," 11-Sep-2018. [Online]. Available: https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761. [Accessed: 11-Feb-2022].

19 Sunil Ray, "Understanding Support Vector Machine (SVM) algorithm from examples," 13-Sep-2017. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code. [Accessed: 11-Feb-2022].

20 Amari, Shun'ichi. The handbook of brain theory and neural networks. MIT press, 2003.

21 JavaTpoint, "Random Forest Algorithm," 5-Sep-2019. [Online]. Available: https://www.javatpoint.com/machine-learning-random-forest-algorithm. [Accessed: 11-Feb-2022].

22 Rebecca Bevans, "Simple Linear Regression | An Easy Introduction & Examples," 19-Feb-2020. [Online]. Available: https://www.scribbr.com/statistics/simple-linear-regression. [Accessed: 11-Feb-2022].

23 JavaTpoint, "Simple Linear Regression in Machine Learning," 5-Oct-2018. [Online]. Available: https://www.javatpoint.com/simple-linear-regression-in-machine-learning. [Accessed: 11-Feb-2022].

24 Beny Maulana Achsan, "Support Vector Machine: Regression," 10-Dec-2019. [Online]. Available: https://medium.com/it-paragon/support-vector-machine-regression-cf65348b6345. [Accessed: 11-Feb-2022].

25 Ashwin Raj, "Unlocking the True Power of Support Vector Regression," 3-Oct-2020. [Online]. Available: https://towardsdatascience.com/unlocking-the-true-power-of-support-vector-regression-847fd123a4a0. [Accessed: 13-Feb-2022].

26 Ashwin Prasad, "Regression Trees | Decision Tree for Regression | Machine Learning," 9-Aug-2021. [Online]. Available: https://medium.com/analytics-vidhya/regression-trees-decision-tree-for-regression-machine-learning-e4d7525d8047. [Accessed: 13-Feb-2022].

27 Achraf KHAZRI, "Faster RCNN Object detection," 9-Sep-2019. [Online]. Available: https://towardsdatascience.com/faster-rcnn-object-detection-f865e5ed7fc4. [Accessed: 13-Feb-2022].

28 S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

29 J. Cao, C. Song, S. Song, S. Peng, D. Wang, Y. Shao, and F. Xiao, "Front vehicle detection algorithm for smart car based on improved ssd model," Sensors, vol. 20, no. 16, p. 4646, 2020.

30 Jianxia Song,Princess Guzman, "How single-shot detector (SSD) works?," 9-Sep-2018. [Online]. Available: https://developers.arcgis.com/python/guide/how-ssd-works/. [Accessed: 13-Feb-2022].

31 Teoh, Soo Siang, and Thomas Bräunl. "Symmetry-based monocular vehicle detection system." Machine Vision and Applications 23, no. 5 (2012): 831-842.

32 Xiaoyong, Wang, Wang Bo, and Song Lu. "Real-time on-road vehicle detection algorithm based on monocular vision." In Proceedings of 2012 2nd International Conference on Computer Science and Network Technology, pp. 772-776. IEEE, 2012.

33 He, Peng, and Feng Gao. "Study on detection of preceding vehicles based on computer vision." In Applied Mechanics and Materials, vol. 333, pp. 912-915. Trans Tech Publications Ltd, 2013.

34 Shiru, Qu, and Li Xu. "Research on multi-feature front vehicle detection algorithm based on video image." In 2016 Chinese Control and Decision Conference (CCDC), pp. 3831-3835. IEEE, 2016.

35 Lixuan Du, Rongyu Zhang and Xiaotian Wang, "Overview of two-stage object detection algorithms." In Proceedings of the 5th International Conference on Intelligent Computing and Signal Processing (ICSP) 2020,20-22 March 2020.

36 Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587. 2014.

37 S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." in IEEE Transactions on

Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

38 Carranza-García, Manuel, Jesús Torres-Mateo, Pedro Lara-Benítez, and Jorge García-Gutiérrez. "On the Performance of One-Stage and Two-Stage Object Detectors in Autonomous Vehicles Using Camera Data." Remote Sensing 13, no. 1 (2021): 89.

39 Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788. 2016.

40 Redmon, Joseph, and Ali Farhadi. "YOLO9000: better, faster, stronger." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263-7271. 2017.

41 W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single Shot MULTIBOX DETECTOR.” Computer Vision – ECCV 2016, pp. 21–37, 2016.

42 Hao Gao, “Understand Single Shot MultiBox Detector (SSD) and Implement It in Pytorch,” 6 June 2018. [Online]. Available: https://medium.com/@smallfishbigsea/understand-ssd-and-implement-your-own-caa3232cd6ad. [Accessed 30 December 2021].

43 S. Jain, "Introduction to Genetic Algorithm & their application in data science," 31-July-2017. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/07/introduction-to-genetic-algorithm. [Accessed 1-Feb-2022].

44 Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft coco: Common objects in context." In European conference on computer vision, pp. 740-755. Springer, Cham, 2014.

45 Tzutalin. LabelImg. Git code (2015). https://github.com/tzutalin/labelImg.

46 Tanu N Prabhu. "Cheat-sheet for Google Colab." 2 May 2020. [Online]. Available: https://towardsdatascience.com/cheat-sheet-for-google-colab-63853778c093. [Accessed 8 January 2022].

47 Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke,Yuan Yu, and Xiaoqiang Zheng. "TensorFlow: Large-scale machine learning on heterogeneous systems." Software available from tensorflow.org. 2015.