SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage: https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Investigation of Deep Learning Model for Vehicle Classification

Ahsiah Ismail[1,*], Amelia Ritahani Ismail[1], Adzreen Nulsyazwan S. Nadzeer[1], Asmarani Ahmad Puzi[1], Suryanti Awang[2], Roziana Ramli[3]

[1] Department of Computer Science, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia (IIUM), 53100 Kuala Lumpur, Malaysia
[2] Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia
[3] Department of Computer and Information Sciences, Northumbria University, Newcastle-upon-Tyne, NE1 8ST, UK

ABSTRACT

The usage of automobiles in cities and metropolitan areas has increased drastically throughout the years and there is a need to monitor the flow of road traffic to improve the traffic congestion and safety. One of the best ways to monitor the traffic is using an artificial intelligence and machine learning. An automatic vehicle tracking system based on artificial intelligence and machine learning can offers capability to analyse the real-time traffic video data for the purpose of traffic surveillance. The computer vision is one of the subsets in machine learning that can train the computer to understand the visual data and perform specific tasks such as object detection and classification. A Vision-based system can be proposed to detect road accidents, predict traffic congestion and further road traffic analytics. This can improve the safety in transportation where it can recognize types of vehicles on the road, detecting road accidents, predicting the traffic congestion and further road traffic analytics. In the context of road traffic monitoring, the parameters of the traffic such as the type and number of vehicles that passes through must be recorded in order to gain valuable insights and make prediction such as the occurrence of traffic congestion. However, this requires reliable informative and accurate data as input for analytics. Therefore, in this research the deep learning model for vehicle classification is investigated to detect, classify types of vehicles and further predictive analytics. The vehicle classification is proposed based on Single Shot Detector (SSD) architecture model. The proposed model is tested on five different classes of vehicles with a total of 1263 images. Experimental results show that SSD model able to achieve 0.721 of precision, 0.741 of recall and 0.731 of F1 Score. Finally, the result show that the SSD model is more accurate among all the models for all the performance measure with the difference of more than 0.052 of precision, 0.706 of recall and 0.05 of F1 Score.

Keywords:
Computer vision; deep learning; object detection; object classification

1. Introduction

The usage of automobiles in cities and metropolitan areas has increased drastically throughout the years and it has become a priority to monitor the flow of road traffic [1]. According to the 2018

* *Corresponding author.*
E-mail address: ahsiah@iium.edu.my

https://doi.org/10.37934/araset.62.2.6676

UN Urbanization Report, 75% extra travel time is spent by people on average due to traffic congestion [2]. As the number of transportation vehicles increases, traffic management has become very tedious due to large-scale traffic conditions [3]. Other than that, the number of fatal accidents that occur annually is staggering and keeps increasing. It is important to have a tools or system such as automatic vehicle tracking system which can analyse real-time traffic video data for the purpose of traffic surveillance [4]. One of the best ways to monitor traffic is using artificial intelligence (AI) system AI system with the vision sensor technology capability to integrate informative data collection and transmission enable an efficient road monitoring such as predicting traffic congestion, detecting road accidents and road traffic analytics [5, 6].

Having a transportation system that implements AI method will also improve safety in transportation. The AI system can recognize vehicles according to their type, detecting road accidents, predicting traffic congestion and road traffic analytics. In the context of road traffic monitoring, the parameters of the traffic such as the type and number of vehicles that passes through must be recorded to collect data and gain valuable insights to make prediction such as the occurrence of traffic congestion. To achieve this, a system that is capable of object recognition must be developed. Therefore, in this research the deep learning model for vehicle classification is proposed to detect and classify the types of vehicles. In this research, the reliability of deep learning model for vehicle classification as a data input for automatic vehicle tracking system is investigated. The vehicle classification information can be further used as the input for analytics model for future prediction. Based on this information, the analytics synthesize, analyze the trends and identify the patterns for prediction, decision making and actions to monitor the traffic and road safety. To test the proposed model, a dataset of vehicles that consists of 1263 images vehicle from five different classes are created. The dataset is created to evaluate the capability of the model to detect and classify types of vehicles. Then, the performance of the deep learning models is compared in terms of the precision, recall, and F1 score.

The rest of this paper is organized as follows: Section 2 presents the related works. Section 3 described the details of the proposed model. The performance of the proposed model is presented in Section 4; followed by the discussion in Section 5. The conclusions and future works are finally made in Sections 6.

2. Literature Review

In this modern world, the usage of vehicles is growing rapidly. According to statistics, it is predicted that the rise in automobile sales will reach up to 71 million worldwide in 2021 [7]. Due to this expeditious increase, analyzing and measure the number of vehicles and its characteristics such as types of vehicles, speed and occupancy the road traffic in real-time is very important for an efficient city traffic management [2]. The focus of this research is to investigate a suitable method that is able to classify types of vehicles for road traffic monitoring. The classification can be used for the detection types of vehicles on a road traffic for a vehicle tracking system and as input for further analytics. This research falls under image classification, where in computer vision, the AI method can be used to detect the instances of the object in digital image. In monitoring traffic, there are two methods can be used which are using the AI and non-AI.

The most utilized non-AI method that used in the monitoring technology is the inductive loop detector which was used since its introduction in the 1960s [8]. When a vehicle passes through, an inductive loop detector which is placed under the roadway captures the changes in the characteristics of the magnetic profile signal such as amplitude, phase, and frequency. The inductance of the loop detector will decrease as a vehicle passes through which then sends a pulse that signifies the presence of a vehicle. Numerous studies have shown that this method has achieved a good performance with accuracy up to 99% in classifying large vehicle. However, the method requires the installation of coil under the roadway. This can be costly and difficult to implement [9].

The implementation of AI for traffic management and monitoring has become more prominent due to its capability. The deep learning method of the AI approach are the most chosen due to their performance especially in more complex environments [10]. The AI-enabled camera that had been introduced by FLIR Systems is able to capture and process data in real-time. The images captured can be further used to classify the types of vehicles using the deep learning method. These data can be used to monitor and control the road traffic [11]. In the context of road traffic monitoring using the deep learning method, the parameters of the traffic such as the types and number of vehicles that passes through must be recorded in order to collect data and gain valuable insights for the prediction such as the occurrence of the traffic congestion. By using data such as such as images and videos with the help of certain attributes, the AI method capable to recognise specified objects in the images. This can be defined as object recognition.

The goal of object recognition is to train a computer to understand the image which is something that is natural to humans. Deep learning is a popular method due to its visual information processing capability. Deep learning requires less manual intervention, no risk of data fitting, has a high fault-tolerant and scalability. Thus, the deep learning is the most suitable for object recognition [5]. Convolutional Neural Network (CNN) is one of the deep learning architectures in which it is the most suitable neural network to be used for the object recognition. In vehicle recognition, the CNN method able to accurately create a non-handcrafted features to represent objects through its deep architecture [12]. This leads to a great success in object recognition or detection [2]. This success stems from the fact that when humans see something, the layers of the brain's neuron detect the features of the object which in the end, the particular object is detected and CNN mimics this way of how mammals, such as humans perceive objects visually [5].

In recent literature, the deep neural network has been rapidly used. Ahmad Bahaa et al. proposed a method of using seismic data to identify and classify vehicles using three different neural networks which are Deep Neural Network, CNN, and Recurrent Neural Network [9]. The data set is collected by placing geophones 15m apart, located nearby the road at Kyushu University. The vehicle is tagged as large, medium and small. The data is then pre-processed by eliminating signal that contained surrounding noise or have overlapping between other vehicles. Noises were also represented in the data as one of the classes. The data is then separated into training, validation and testing set for the purpose of testing the neural network. The data that is collected is in abundance and it is carefully processed. In their work, they are using different neural network in order to find the most suitable model that can be use with seismic data. However, using seismic data for vehicle identification will be difficult considering there will be a lot of interference by noise especially in extremely noisy places such as city. This can affect the accuracy of detection and classification. Although they have conduct practical application of this method and able to obtain high accuracy, it is conducted only inside a university campus which is relatively quiet environment as compared to the city.

Huaizhong *et al.,* [13] utilise deep learning techniques for the purpose. For this purpose, the data of a surveillance video is collected from a fixed Unmanned Aerial Vehicle (UAV). The vehicle recognition and tracking model was developed using Mask R-CNN and then traffic analysis is performed. The usage of vehicle recognition and tracking model for the purpose of traffic monitoring such as vehicle counting, speed estimation and traffic congestion analysis are demonstrated in their work. However, this method of traffic analysis is expensive as it requires surveillance video using UAV. Besides that, it also requires high processing computing capability due to the data is in the form of video, thus can be costly.

Similarly, Michael *et al.,* [14] proposed a method for vehicle detection in aerial images using a neural network called Double Focal Loss Convolutional Network (DFL-CNN). Due to the nature of aerial photos which is very different from the generic photos, the architecture of the neural network consists of a skip-connection from the low layer to the high layer in order to learn the features of the photo. The traditional cross entropy is replaced by focal loss function. To test the method, the images are carefully selected and then manually labelled. The proposed neural network able to achieves a better result compared to other neural networks such as Faster R-CNN in their work. This is due to the high-quality dataset use in the experiment. Although the model outperforms other type of neural networks, the precision of the model only able to obtain 64%, which is only slightly higher than neural networks such as Faster R-CNN.

Based on the review, the deep learning CNN method are chosen for vehicle classification. It is selected due to its stability. The research will focus on the suitable selection of the model in recognizing types of vehicles.

3. Methodology

The vehicle classification consists of two main phases, namely data acquisition and model training and testing. Three deep learning model of CNN namely Single Shot Detector architecture (SSD), EfficientDet and CenterNet architecture is investigated in this research. The model is tested on vehicle dataset with a total number of 1263 images of vehicle from five different classes and their performance are compared.

3.1 Convolutional Neural Network

In Convolutional Neural Network (CNN), several characteristics of images that are taken as input are assigned according to its importance for the purpose of differentiating them from one another. In order to obtain a good prediction, the important characteristics of the images are not eliminated when the CNN reducing the images while processing it. The CNN has three main layer namely convolutional layer, pooling layer and fully-connected layer. The function of Convolutional Layer is to extract the features and it is a crucial part of a CNN due to the fact that almost all the computation process occurs at this layer. While the Pooling Layer is almost the same in terms of its functionality as convolutional layer. Both layers involve the sweeping of filter across the entire image. However, there is no weight for the filter in pooling layer. For the fully connected layer, the feature extracted through the previous layer is used for the task of classification.

3.1.1 Single shot detector architecture

Deep CNNs are utilised in Single Shot Detector (SSD) for the purpose of classifying the object and identifying the location of the object in the image [15]. SSD is a multi-scale sliding window detector. Figure 1 shows the architecture of the SSD model.

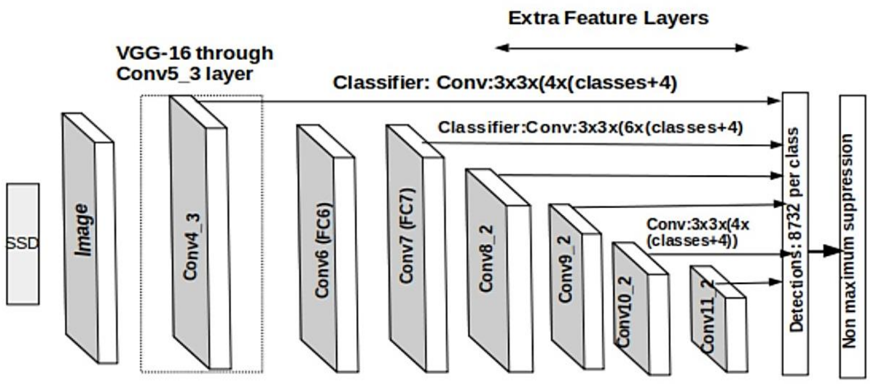

Fig. 1. SSD model architecture [16]

The feature extraction and object detection are the two main section in SSD. The VGG-16, which is a CNN architecture acts as the feature map extractor, then followed by six other auxiliary convolutional layer as shown in Figure 1. For the purpose of object identification, five of the auxiliary layer are used. The main advantages of using SSD are that, it makes more predictions and has a better coverage in several aspects such as the scale, location and aspect ratio of the image [17].

3.1.2 CenterNet architecture

Anchorless object detection architecture is utilised in CenterNet. By having such structure, a much more elegant algorithm that is natural to the CNN flow replaces the Non-Maximum Suppression (NMS) at the post process [18]. The architecture of the CenterNet model is shown in Figure 2.

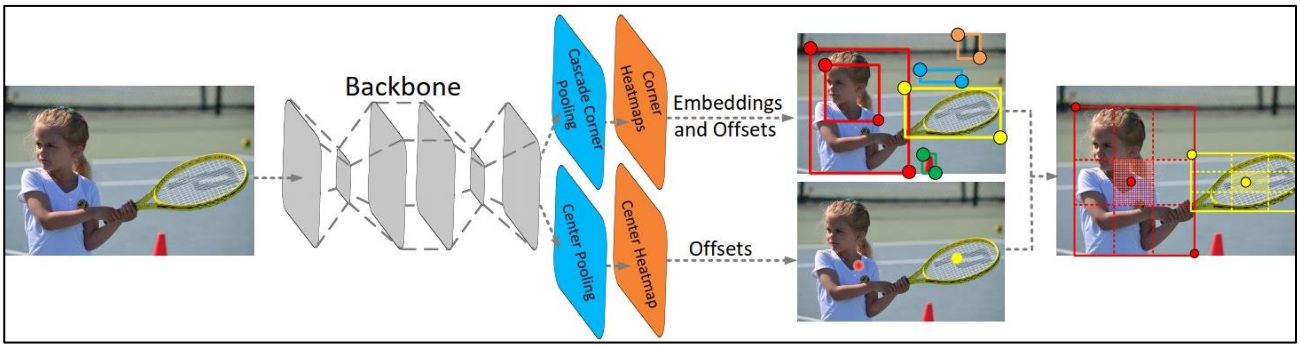

Fig. 2. CenterNet model architecture [19]

To improve the precision and recall, the object is detected as a triplet in CenterNet. The cascade corner pooling and centre pooling is applied by a convolutional backbone network which produces two corner heatmaps and a centre keypoint heatmap as illustrate in Figure 2. A potential bounding box is detected using the detected pair of corners and the final bounding box is determined by using the detected centre keypoint. By using such architecture, CenterNet requires less time for inferencing [18].

3.1.3 EfficientDet *architecture*

The one stage detector paradigm is employed in EfficientDet. Figure 3 shows the architecture of the EfficientDet Model. In EfficientDet model, all the dimensions of the backbone are scaled up to improve accuracy and efficiency.

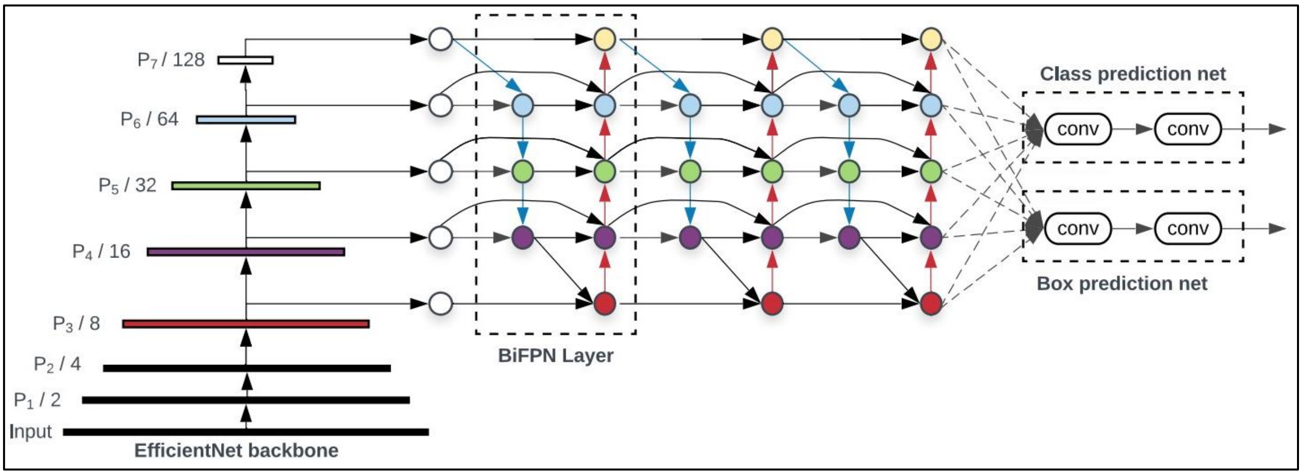

Fig. 3. EfficientDet architecture

As shown in Figure 3, level 3 until 7 features from the EfficientNet backbone is taken by the BiFPN Layer. A top-down and bottom-up bidirectional feature fusion is applied at the BiFPN Layer. The object class and bounding box is produced from the fused features which are fed to the class and box network [20]. EfficientDet is useful for application that requires low latency because it has a significant speedup on GPU and CPU [21].

The flowchart of the experiment conducted on the proposed model that consist of four phases as shown in Figure 4. Each of the phase will be discuss in the following subsections.

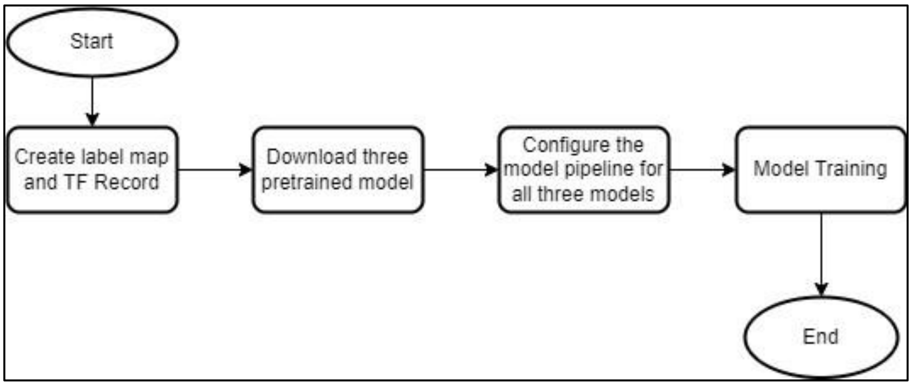

Fig. 4. Experiment flowchart

3.2 Label Map and TF Record

Firstly, a label map in the form of PBTXT file is created. The file contains the name of the classes. After the label map is created, the data for the training and testing sets, which are the images and its label is converted into a sequence of binary code in the form of TF Record file for training and testing purposes. The size of the image, bounding box and label are all converted into binary sequence. Each training and test sets will have their own TF Record file.

3.3 Pretrained Models

The model will be train using Tensorflow. For this purpose, the SSD, CenterNet, and EfficientDet pretrained models are downloaded from the Tensorflow GitHub repository alongside with the model pipeline.

3.4 Model Pipeline Configuration

In the model pipeline, the batch number are set to four and the number of classes are set to five. The file path for label map, training data TF Records and testing data TF Records file are set according to where it is stored in the file directory. All the other parameters in the model pipeline are kept unchanged. The parameters setting value use are listed in Table 1.

Table 1
Parameter value

Model / Parameter	SSD	CenterNet	EfficientDet
Learning rate	0.04	0.01	0.08
Weight	0.0004	N/A	0.00004
Classification weight	1.0	N/A	1.0
Localization weight	1.0	N/A	1.0
Task loss weight	N/A	1.0	N/A
Offset loss weight	N/A	1.0	N/A
Scale loss weight	N/A	0.1	N/A
Object centre loss weight	N/A	1.0	N/A
Matched threshold	0.5	N/A	0.5
Unmatched threshold	0.5	N/A	0.5
Score threshold	0.00000001	N/A	0.00000001
IoU threshold	0.6	N/A	0.5

3.5 Model Training

After the model pipeline of each model is configured, the models are ready to be train. The model will be train for two iterations in order to evaluate their performance towards number of training steps. For the first iteration, the model will be trained with 10,000 training steps. While in the second iteration, the model will be trained with 15,000 training steps.

3.6 Experimental Setup

To demonstrate the reliability of the proposed model, a series of experiments is conducted on a laptop with a specification of 16GB RAM and the NVIDIA Quadro M1200 GPU installed to speed up the model training process. The dataset of vehicle images is used the test the model. To prepare the dataset for training the model, 1263 images of vehicle from five different classes is collected and then manually labelled. Table 2 shows the details characteristics of vehicle dataset images.

In this experiment, 80 percent of the dataset is used as the training set and 20 percent of it will be used as the testing set. The classification performance of all models SSD, CenterNet and EfficientDet are measured in terms of precision, recall and F1 score. Precision is the ratio of true positives and all of the positives predicted, recall is the ratio of true positives and all of the actual positives while F1 Score is a measure of performance that combines precision and recall into single number.

Table 2
Details of characteristics in vehicle dataset

Class	Total
Ambulance	146
Bus	298
Car	306
Motorcycle	236
Truck	277
Total	1263

4. Results

In this research to obtain the most suitable model for the vehicle classification, the SSD, CenterNet and EfficientDet model are tested on vehicle dataset. Despite the SSD model able to achieve the highest F1 Score among the model tested as presented in Table 3 and Table 4, the reasons that reduce the performance for all the model tested are due to the small sample dataset utilized in the experiment. The model of the CNN approach requires a large number of images for training in order to obtain a desired result [22-27]. While in this research, a small number of vehicle dataset is use as a proof-of-concept to classify the types of vehicles. To ensure the reliability of the model, instead of analyzing the performance only against types of vehicles, we also evaluated on the number of training steps. There are two training steps tested in the experiment which are 10,000 and 15,000 training steps as presented in Table 3 and Table 4. Their performance is evaluated and compared. All the performance measure of the SSD method increased with the increment number of training steps as presented in Table 3 and Table 4. However, as the number of training steps increase, the computation time will also increase. The SSD model took approximately 8 hours to train for 10,000 steps meanwhile it took approximately 12 hours for the SSD model to be trained for 15,000 steps.

Table 3
10,000 training steps evaluation

Model	Precision	Recall	F1 score
SSD	0.614	0.686	0.648
CenterNet	0.588	0.712	0.626
EfficientDet	0.616	0.626	0.644

Table 4
15,000 training steps evaluation result

Model	Precision	Recall	F1 Score
SSD	0.721	0.741	0.731
CenterNet	0.582	0.706	0.638
EfficientDet	0.658	0.705	0.681

For 10,000 training steps, the EfficientDet model able to obtain the highest precision, the CenterNet model is the highest recall and the SSD model able to obtain the highest F1 Score. While at 15,000 training steps, the SSD model shows the most significant improvement. The SSD model achieved the highest precision, recall, and F1 score.

5. Discussion

Overall, among all the models, the SSD outperform others in detecting and classifying the vehicle images. The SSD method exceeds 0.648 F1 Score for all training step evaluation. Thus, the SSD model is used to further evaluate its performance in detecting the vehicle images. In this research, the model is tested on two different scenarios which are the vehicle in a close proximity and the vehicle which is far away. Figure 5 and Figure 6 shows the images in these two scenarios.

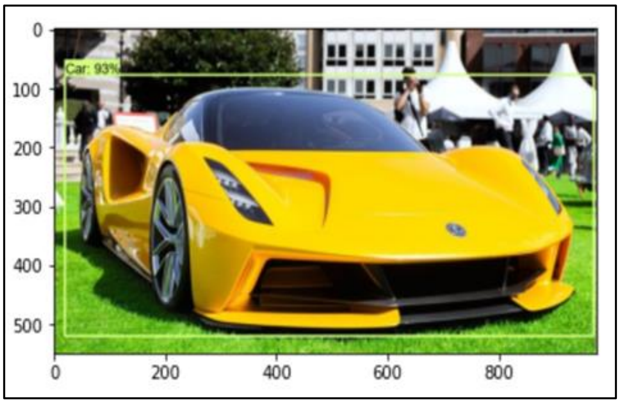

Fig. 5. Vehicle in close proximity

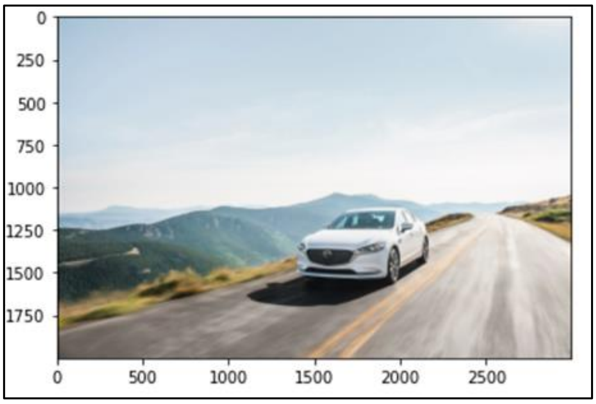

Fig. 6. Far away vehicle

From the experiment conducted, it is observed that when the vehicle image is in close proximity, the SSD model able to detect the types of vehicles. However, when the vehicle image is far away, the performance of the model is decrease. The is due to the reason that the SSD model unable to detect the small objects [28].

6. Conclusions and Future Works

This paper proposed a deep learning model for vehicle classification in traffic monitoring. The result shows that the SSD model able to effectively detect and classify the types of vehicles. The SSD model outperformed others in terms of all performance measure of precision, recall and F1 Score. Though the proposed model is superior to other two models and able to achieves higher for all the performance measure of precision, recall and F1 score in the dataset tested, the model was shown to be less effective in detecting vehicle that are far away. In the future, we will concentrate on improving the detection on the vehicle that are far away and consider on increasing the diversity of vehicle images dataset.

Acknowledgement
This research was funded by UMP-IIUM Sustainable Research Collaboration 2022 grant (IUMP-SRCG22-015-0015).

References
[1] Abd Wahid, Nurul Bahiyah, Nurnabilah Syahirah Jafri, Muhammad Afiq Mohd Nor, Venusha Segar, Suzita Ramli, Noraine Salleh Hudin, Nor Zila Abd Hamid, Mohd Talib Latif, and Anggi Tias Pratama. "Surfactants and Microbial Aerosol of Urban Particulate Matter (PM10) in Kuala Lumpur City Centre." *Journal of Advanced Research in Micro and Nano Engineering* 16, no. 1 (2024): 61-69. https://doi.org/10.37934/armne.16.1.6169

[2] Zhang, Huaizhong, Mark Liptrott, Nik Bessis, and Jianquan Cheng. "Real-time traffic analysis using deep learning techniques and UAV based video." In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1-5. IEEE, 2019. https://doi.org/10.1109/AVSS.2019.8909879

[3] Priya, S. Sasi, S. Rajarajeshwari, K. Sowmiya, and P. Vinesha. "Road traffic condition monitoring using deep learning." In *2020 International Conference on Inventive Computation Technologies (ICICT)*, pp. 330-335. IEEE, 2020. https://doi.org/10.1109/ICICT48043.2020.9112408
[4] Khairi, Danial Mohd, Mohd Azman Abas, Mohd Farid Muhamad Said, and Wan Saiful-Islam Wan Salim. "Fuel consumption mathematical models for road vehicle–A review." *Progress in Energy and Environment* (2021): 59-71.
[5] Haritha, H., and Senthil Kumar Thangavel. "A modified deep learning architecture for vehicle detection in traffic monitoring system." *International Journal of Computers and Applications* 43, no. 9 (2021): 968-977. https://doi.org/10.1080/1206212X.2019.1662171
[6] Wang, Dan, Terh Jing Khoo, and Zhangfei Kan. "Exploring the application of digital data management approach for facility management in Shanghai's high-rise buildings." *Progress in Energy and Environment* (2020): 1-15.
[7] Urbánová, Mária, Dominika Čeryová, Viktória Benďáková, and Patrícia Husárová. "Electric vehicles from an economic point of view." *Economics and Culture* 20, no. 1 (2023): 102-113. https://doi.org/10.2478/jec-2023-0009
[8] Baker, Francesca. "The technology that could end traffic jams." *BBC. com (2018, December 12)* (2018).
[9] Ahmad, Ahmad Bahaa, and Takeshi Tsuji. "Traffic monitoring system based on deep learning and seismometer data." *Applied Sciences* 11, no. 10 (2021): 4590. https://doi.org/10.3390/app11104590
[10] Mozaffari, Sajjad, Omar Y. Al-Jarrah, Mehrdad Dianati, Paul Jennings, and Alexandros Mouzakitis. "Deep learning-based vehicle behavior prediction for autonomous driving applications: A review." *IEEE Transactions on Intelligent Transportation Systems* 23, no. 1 (2020): 33-47. https://doi.org/10.1109/TITS.2020.3012034
[11] Kulhandjian, Hovannes. *AI-based Pedestrian Detection and Avoidance at Night Using an IR Camera, Radar, and a Video Camera*. No. 22-53. San Jose State University. College of Business. Mineta Transportation Institute, 2022.
[12] Thiiban Muniappan, Thiiban Muniappan, Nor Bakiah Abd Warif, Ahsiah Ismail, and Noor Atikah Mat Abir. "An evaluation of convolutional neural network (cnn) model for copy-move and splicing forgery detection." *International Journal of Intelligent Systems and Applications in Engineering* 11, no. 2 (2023): 730-740.
[13] Zhang, Huaizhong, Mark Liptrott, Nik Bessis, and Jianquan Cheng. "Real-time traffic analysis using deep learning techniques and UAV based video." In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, p. 1-5. IEEE, 2019. https://doi.org/10.1109/AVSS.2019.8909879
[14] Yang, Michael Ying, Wentong Liao, Xinbo Li, and Bodo Rosenhahn. "Deep learning for vehicle detection in aerial images." In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 3079-3083. IEEE, 2018. https://doi.org/10.1109/ICIP.2018.8451454
[15] Wadhwa, Kanishk, and Jay Kumar Behera. "Accurate real-time object detection using SSD." *International Research Journal of Engineering and Technology* 7, no. 5 (2020).
[16] Saji, Ruhin Mary, and N. V. Sobhana. "Real Time Object Detection Using SSD For Bank Security." In *IOP Conference Series: Materials Science and Engineering* 1070, no. 1, p. 012060. IOP Publishing, 2021. https://doi.org/10.1088/1757-899X/1070/1/012060
[17] Kanimozhi, S., G. Gayathri, and T. Mala. "Multiple Real-time object identification using Single shot Multi-Box detection." In *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, pp. 1-5. IEEE, 2019. https://doi.org/10.1109/ICCIDS.2019.8862041
[18] Khanam, Rahima, Muhammad Hussain, Richard Hill, and Paul Allen. "A comprehensive review of convolutional neural networks for defect detection in industrial applications." *IEEE Access* (2024). https://doi.org/10.1109/ACCESS.2024.3425166
[19] Duan, Kaiwen, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, and Qi Tian. "Centernet: Keypoint triplets for object detection." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6569-6578. 2019. https://doi.org/10.1109/ICCV.2019.00667
[20] Tan, Mingxing, Ruoming Pang, and Quoc V. Le. "Efficientdet: Scalable and efficient object detection." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10781-10790. 2020. https://doi.org/10.1109/CVPR42600.2020.01079
[21] Mekhalfi, Mohamed Lamine, Carlo Nicolò, Yakoub Bazi, Mohamad Mahmoud Al Rahhal, Norah A. Alsharif, and Eslam Al Maghayreh. "Contrasting YOLOv5, transformer, and EfficientDet detectors for crop circle detection in desert." *IEEE Geoscience and Remote Sensing Letters* 19 (2021): 1-5. https://doi.org/10.1109/LGRS.2021.3085139
[22] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." *Communications of the ACM* 60, no. 6 (2017): 84-90. https://doi.org/10.1145/3065386
[23] Ismail, Ahsiah, Mohd Yamani Idna Idris, Mohamad Nizam Ayub, and Lip Yee Por. "Vision-based apple classification for smart manufacturing." *Sensors* 18, no. 12 (2018): 4353. https://doi.org/10.3390/s18124353
[24] Ismail, Ahsiah, Mohd Yamani Idna Idris, Mohamad Nizam Ayub, and Lip Yee Por. "Investigation of fusion features for apple classification in smart manufacturing." *Symmetry* 11, no. 10 (2019): 1194. https://doi.org/10.3390/sym11101194

[25] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition." *IEEE transactions on pattern analysis and machine intelligence* 37, no. 9 (2015): 1904-1916. https://doi.org/10.1109/TPAMI.2015.2389824

[26] Xiao, Zhitao, Xinxin Zhang, Lei Geng, Fang Zhang, Jun Wu, and Yanbei Liu. "Research on the method of color fundus image optic cup segmentation based on deep learning." *Symmetry* 11, no. 7 (2019): 933. https://doi.org/10.3390/sym11070933

[27] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In *2009 IEEE conference on computer vision and pattern recognition*, p. 248-255. Ieee, 2009. https://doi.org/10.1109/CVPR.2009.5206848

[28] Yundong, L.I, Dong Han, L.I. Hongguang, Xueyan Zhang, Baochang Zhang, and Xiao Zhifeng. "Multi-block SSD based on small object detection for UAV railway scene surveillance." *Chinese Journal of Aeronautics* 33, no. 6 (2020): 1747-1755. https://doi.org/10.1016/j.cja.2020.02.024