Jurnal Kejuruteraan 34(5) 2022: 843-850
https://doi.org/10.17576/jkukm-2022-34(5)-11

Incremental Learning of Deep Neural Network for Robust Vehicle Classification

Ahmad Mimi Nathiratul Athriyah*, Abdul Kadir Muhammad Amir, Hasan F. M. Zaki, Zainal Abidin Zulkifli & Abdul Rahman Hasbullah

Department of Mechatronic Engineering, International Islamic University Malaysia Selangor, Malaysia
Centre for Unmanned Technologies, International Islamic University Malaysia Selangor, Malaysia

**Corresponding author: nathiratul95@gmail.com*

Received 21 September 2021, Received in revised form 03 December 2021
Accepted 31 December 2021, Available online 30 September 2022

ABSTRACT

Existing single-lane free flow (SLFF) tolling systems either heavily rely on contact-based treadle sensor to detect the number of vehicle wheels or manual operator to classify vehicles. While the former is susceptible to high maintenance cost due to wear and tear, the latter is prone to human error. This paper proposes a vision-based solution to SLFF vehicle classification by adapting a state-of-the-art object detection model as a backbone of the proposed framework and an incremental training scheme to train our VehicleDetNet in a continual manner to cater the challenging problem of continuous growing dataset in real-world environment. It involved four experiment set-ups where the first stage involved CUTe datasets. VehicleDetNet is utilized for the framework of vehicle detection, and it presents an anchorless network which enable the elimination of the bounding boxes of candidates' anchors. The classification of vehicles is performed by detecting the vehicle's location and inferring the vehicle's class. We augment the model with a wheel detector and enumerator to add more robustness, showing improved performance. The proposed method was evaluated on live dataset collected from the Gombak toll plaza at Kuala Lumpur-Karak Expressway. The results show that within two months of observation, the mean accuracy increases from 87.3 % to 99.07 %, which shows the efficacy of our proposed method.

Keywords— Single-lane free flow (SLFF); automated vehicle detection and classification (AVC)

INTRODUCTION

Vehicle classification for toll collection in Malaysia relies on conventional methods such as human observation, sensor-based detection and classification based on electronic devices. One of the significant problems with the manual toll collection system is that it is highly prone to human errors during observation of vehicle classes due to fatigue, distractions and unobjective human intuition (Awang et al. 2018). For every vehicle that passes the lane, their type of classes will be observed and manually evaluated by the operator in the toll booth which is a labor extensive task given a prolong period.

In addition, the existing treadle-based sensor available at the Kuala Lumpur – Karak Expressway as depicted in Figure 1, is only capable of being evaluated based on the operator's observation by tracking the number of passing vehicles' wheels. However, this sensor can only be implemented after the payment at the toll booth where the toll gate will open to pass the vehicle. The sensor cannot be implemented before the toll booth due to restriction in the space because every vehicle has a different length and size. Additionally, this sensor cannot differentiate between vehicle classes that share similar wheel counting and axle profiles e.g. between Class 1 (vehicles with 2 axles and 3 or 4 wheels) and Class 4 (taxis) (PLUS Malaysia Berhad). Moreover, the treadle-based sensor incurs high maintenance costs due to mechanical wear and tear.

FIGURE 1. The current treadle sensor used at KL-Karak Expressway

In introducing MLFF, the first step or transition step is to approach it using the Single Lane Free Flow (SLFF) (Ministry of Work Malaysia, 2019). To realise the concept of SLFF and MLFF in Malaysia, accurate identification, detection, classification of the vehicle and charging of the vehicle depending on its class are very important. Therefore, in this paper, a robust vehicle classification approach was proposed using the camera and deep learning object detection algorithm.

RELATED WORK

Many vehicle classification methods have been proposed and can be categorized based on the type of sensors used. These sensors include laser sensor, magnetic sensor, piezoelectric sensor and vision sensor.

CONVENTIONAL VEHICLE CLASSIFICATION METHOD

Laser sensors have the highest reliability as they enable retrieving the vehicle's three-dimensional (3D) profile (Ng et al. 2011). A feature characterization experiment conducted for the supervised classification of laser scanner profiles by Sandhawalia et al. (2013) shows that the accuracy of the method is 99.2%. The 3D profile from the vehicle was illustrated as a two-dimensional (2D) image with depth values. Then, the depth values will be compared with the references of depth value specific for each class. The advantage of laser range sensor over the other sensors is that it is less sensitive to deteriorated environment conditions such as light intensity, rain and fog.

On the other hand, laser sensors require a high initiation cost if the system were implemented (Hassan et al. 2005). Nevertheless, this method is only suitable for the vehicle classes differentiated based on the vehicle's size. In the case of vehicle classification for the Malaysian toll system, the reliability of this method becomes degraded as there are different vehicle classes that share the same vehicle size.

The magnetic sensor or inductive sensor can be used to measure the change of magnetic field within the range of 1 micro gauss to 10 gausses when a vehicle is passing the sensor (Chen et al. 2019). The result of the change of the magnetic field can be analyzed and classified into different vehicle types. Due to the high sensitivity, the magnetic sensor needs to be placed further away from any magnetic material, disrupting the accuracy of the magnetic field signal. Signal length is different to other types stated in the paper (Kaewkamnerd et al. 2010). It is also stated that when the same classification is performed on the vehicle with almost identical dimensions such as car, van and pickup trucks, the result becomes as low as 78%.

Placing a piezoelectric strip sensor diagonally on the road can be used as a method to detect the passing vehicle wheels and categorize them according to the vehicle class as mentioned in the paper (Kuo et al. 2011). It can also recognize various vehicle classification, including vehicles with 2 axles and 3- or 4-wheels like car and vehicles with 2 axles and 5- or 6-wheels such as lorry. However, this piezoelectric sensor cannot differentiate the vehicles that have the same number of wheels but in different classes such as cars and taxis.

Another method for vehicle classification system is using a low-cost camera to capture image or image sequences of the vehicle (Ng & Tay, 2012). Based on the captured vehicle image from a continuous video stream, a classification system will analyze and infer the class of vehicle that appear in the image. Hence, the vision-based vehicle classification algorithm is critical in determining the accuracy and performance of the system. The advancement of deep learning in vision applications opens up the possibilities to implement a vision-based method for highly accurate and robust vehicle classification (Gupta et al. 2021).

DEEP LEARNING-BASES VEHICLE CLASSIFICATION

Deep learning is a subset of machine learning inspired by the structure of the human neural system which uses the neural network in analyzing different factors such as recognizing faces (Miller & Brown 2018). Most of the researchers favor to use deep learning over machine learning algorithm although it works well for huge dataset (Aravind 2020).

Traditional approaches for visual vehicles detection and classification have two flaws: one, the region selection strategy based on sliding windows is not specific, the time complexity is large, and the detection windows are redundant; and two, hand-designed features are not very suitable for diverse changes such as R-CNN (Girshick et al. 2014).

Another deep learning based detectors are Single Shot Detectors (Liu et al. 2016) and YOLO (Vikram & S, 2018). Both detectors utilize anchors box. Anchor-based approaches have the advantage of constraining the aspect ratio and preventing abnormal detection shapes due to the prior knowledge provided by anchors. However, the drawbacks of anchors may outweigh the merits. Firstly, hyperparameters like the number, size and aspect ratio of anchor boxes have an impact on detection performance. Secondly, since all anchors contribute to the overall loss during model training and only a few ground truths bounding boxes match positive anchors, the unbalanced samples could make the training process difficult. Thirdly, anchor-based approaches frequently result in overlapping detections. Necessitating the use of non-maximum suppression (NMS) for post-processing can be inaccurate in some cases.

Keypoint-based detectors such as CenterNet and CornerNet solve the problems of the anchor-based methods and these methods are naturally compatible with multi-class classification problems. Keypoint-based detectors perform the detection and classification of objects in single-stage and do not need anchors nor NMS which provide a good trade-off between speed and accuracy (Duan et al. 2019). As these methods are based on convolutional neural network, the architecture of the methods is highly customizable and extensible for various applications.

METHODOLOGY

This section discusses the methodology of our proposed vehicle classification method. This includes the experimental setup of the deep learning model used as a backbone of the overall vehicle classification algorithm, hardware, and the circuit design, as shown below in Figure 2.

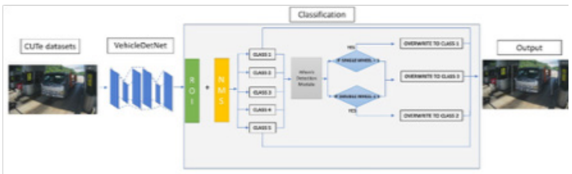


FIGURE 2. Vehicles Classification pipeline

VehicleDetNet

Vehicle classification requires the detection of a wide range of vehicle movement at numerous scales, as well as the exact localization of the vehicle boundary. Edge-keypoint-based detectors, such as CornerNet and ExtremeNet, are the most common anchor-free approaches currently available (Kim et al. 2021). This method is unable to model the total information of the vehicles because it lacks overall perception capabilities and require a post-processing step for grouping the different keypoints which drastically slows down the detection process.

Another type of center-based keypoints detector is CenterNet. This detection design eliminates the requirement for anchors and the computationally intensive NMS to achieve fast and accurate object detection. It is based on the realization that box predictions can be ordered for importance by the location of their centres rather than by their overlap with the object. This knowledge is now applied to other deep learning problems. However, practically throughout the experiment, the model detected more than one bounding box during inference. The example is as shown in Figure 3.

FIGURE 3. Redundant detection occurs in practical situation.

In this paper, we also present an anchorless framework for vehicle classification. This method can eliminate the bounding boxes of candidates' anchors but we proposed to use IoU-based non-maximum suppression (NMS) during the post-processing to get the best bounding box as a result. NMS is further discussed in the classification network.

The overall anchorless architecture is shown in Figure 4. A convolutional backbone network applies two types of pooling: cascaded corner pooling and centre pooling (Duan K. et al. 2019). The cascade corner pooling techniques give corner heatmaps, while the centre pooling will be resulting from centre keypoint heatmaps. A pair of detected corners and a similar embedding will be used to detect a potential bounding box. Next, the detected centre keypoints are used to determine the final bounding boxes and filter out the incorrect bounding boxes. The network models an object as a single point as their own bounding box critical point estimation was used to detect all the centre of the objects with prediction if $Y_{(x,\, y,\, c)} = 1$ for a detected keypoint and $Y_{(x,\, y,\, c)} = 0$ for a background.

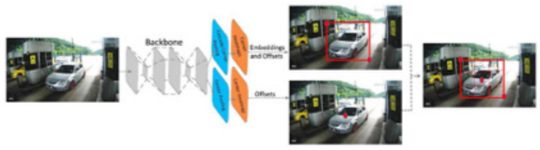


FIGURE 4. Anchorless architecture

ClassesNet

After object detection process, we apply a region of interest (ROI) on the model to focus on the area of on-target vehicle and minimize the background. On each frame of the stream video, multiple background and dynamic objects may appear on the scene and by utilizing ROI it will enable the model to detect the target vehicle as shown below in Figure 5.

FIGURE 5. Before and after ROI applied

As mentioned before, we proposed to utilize non-maximum suppression (NMS) to select the highest confident bounding box as a result. NMS was applied in the case where multiple bounding boxes overlap for the same object.

Each bounding box contain 5 predicted elements which are x, y, w, h and a box confidence score. The coordinates represent the center of the box, relative to the grid cell location. These coordinates are normalized to fall between zero and one. The (w, h) box dimensions are also normalized to [0,1], relative to the image size. The confidence score indicates how probable the bounding box is to include an object and how precise it is. The confidence in both the classification and the localization is measured by the class confidence score for each prediction box. If no object exists in that cell, the confidence score should be zero. Otherwise, we want the confidence score to equal the intersection over union (IoU) between the predicted box and the ground truth which is given by the equations (1)-(3).

$$box\ confidence\ score = P_r(object) \cdot IoU \tag{1}$$

$$\begin{matrix} conditional\ class \\ probability \end{matrix} = P_r(class_i \mid object) \cdot IoU \tag{2}$$

$$class\ confidence\ score = P_r(class_i) \cdot IoU \tag{3}$$

The Jaccard index, or intersection over union (IoU), is a metric for evaluating the accuracy of an object detector. It calculates the intersection size and divides it by the union size. More generally, IoU is a measure of the overlap between two bounding boxes. The higher the IoU the better the accuracy. An IoU score IoU $\geq$ 0.5 is normally considered as a true positive.

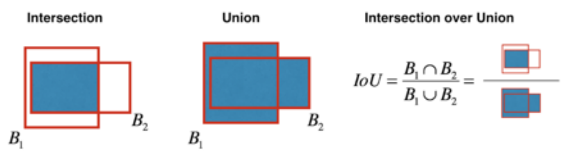


FIGURE 6. Intersection over Union

NMS discards certain choices for different objects that are different detections of the same object. For each object class, remove all bounding boxes where the probability of an object being present is less than a certain threshold (0.4). Bounding boxes with the highest confident score will be filtered out and classify corresponding on the vehicles classes as a result. The result of detection after NMS is shown below.

FIGURE 7. NMS

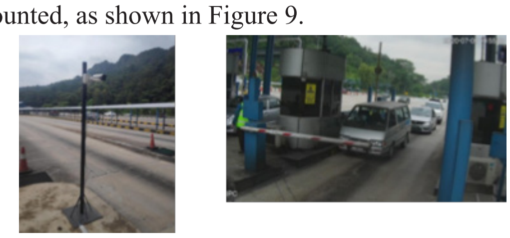
After the detection result and classify the vehicles based on their classes, we did another filtered based on wheels detection module. After the camera has detected the vehicles, the vehicle type classification system will distinguish the classes of vehicles such as class 1 for light vehicles, class 2 for middleweight, class 3 for the heavyweight, class 4 for taxis and class 5 for buses. The algorithm detects and classifies based on the features of the vehicles and the number of wheels or wheels. Then, the toll charge will be deducted based on the type of vehicle classes. For example, all lorries have standard features, but they will be filtered based on the total number of wheels or wheels. If there is more than one single wheel, then class 2 lorries will be overwritten by class 1 lorries. If the double wheel is less than or equal to two, the class 2 lorries will change to class 2 or remain. Meanwhile, for class 3, lorries will be detected based on the number of double wheels, which is more than two axles of double wheels (PLUS Malaysia Berhad).

CUTe DATASETS

In order to perform vehicle classification in Malaysia toll plaza scenario, we have introduced CUTe Vehicles Dataset in this paper. The dataset contains over 12 422 labelled images of 5 classes of vehicles as mentioned before. The images were captured in different view, pose, angle and lighting condition. These images are captured from Gombak and Puchong toll plaza in different lighting conditions such as low light environment, dark in the night, various weather conditions such as daylight, rainy days.

FIGURE 8. Examples of CUTe dataset

In the experiment setup, the datasets were collected from three different angles at two locations to increase the variety of data, thus, increasing the robustness of the algorithm. Cameras are mounted 45 degrees at horizontal axis and we considered two placement of camera such as in Figure 9 and Figure 10 to evaluate the effect of the camera placement. To capture the image of incoming vehicles to the toll booth, an IP camera is mounted, as shown in Figure 9.



(a) (b)

FIGURE 9. (a) The camera position (b) The view from the mounted camera (6m from the toll booth)

Figure 10 shows the current position of the camera. The camera was repositioned due to several reasons. Firstly, there is a roof above the camera to improve the lighting conditions and weather such as rain. Secondly, the camera was able to detect the plate numbers since there was no obstruction from the toll gate. Thirdly, the repositioned camera was able to capture the image of double wheels at the rear axles.

(a) (b)

FIGURE 10. (a) The camera position (b) The view from the mounted camera and current position of the camera (4m from the toll booth)

The different camera distance and position will impact the performance. It will be further discussed in Result and Analysis section.

FIGURE 11. The position of the camera at Puchong Plaza

Before the main setup was deployed at the actual experimental tested at Gombak Toll Plaza, we performed data collection at Toll Plaza Puchong, as shown in Figure 11, for labelling purposes and training our deep learning model. The trained model was then transfer-learned to the new dataset at the Gombak Toll Plaza. It is worth mentioning that the testing datasets only contains the videos from the Gombak Toll Plaza.

CIRCUIT DESIGN

The circuit design of the whole system is shown in Figure 5. The Jetson TX2 Development Board is utilized for the inference of the classification model. An IP camera is connected to perform the object detection process. There are two power supplies used in this system to cater to all the images processing. For NVIDIA Jetson TX2 Development Board, it uses a 19V DC power supply. A 48V DC power supply is used and connected with Gigabit Passive Power over Ethernet (PoE) to power up the IP Camera. The camera uses only an ethernet cable to power up and receive or send the data images. The 5 Ports 10/100Mbps Desktop Network Switch will switch to real-time data input from the toll operator in the toll booth and the image data from the IP camera to the NVIDIA Jetson TX2 Developer Board.

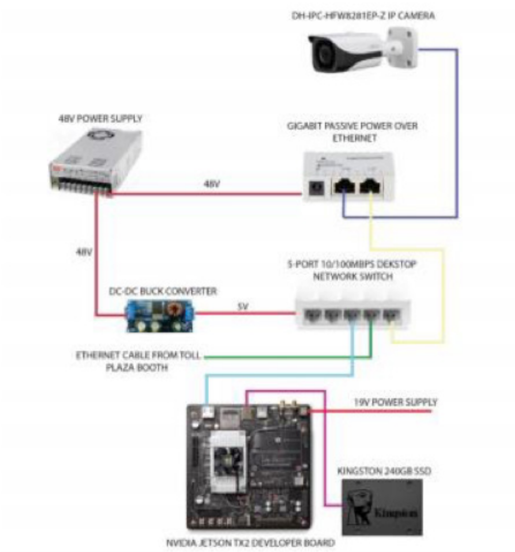


FIGURE 12. Overall Circuit Design to Capture the Videos and Perform the Inference

To augment the learning of the VehicleDetNet towards the real environment dataset, we propose an incremental training which is a method where data accumulate knowledge and learn continuously (Qu et al. 2021). Figures 13 shows the proposed incremental model training pipeline.

FIGURE 13. Incremental model training pipeline.

In Phase 1 of the manual model training, a dataset was collected for the first time at the Puchong Toll Plaza for training the base VehicleDetNet model with a total of 3049 datasets. The ratio for Class 1 and Class 2 are 40 and 7 respectively. Meanwhile Class 3, Class 4 and Class 5 shared the same ratio which is 1. Notice that this dataset is highly imbalance. However, our model can robustly classify the vehicles correctly. This will be further explained in Section 4.

The weekly image data from Toll Plaza Puchong were automatically labelled based on their class types and number of wheels. The training and validation datasets were randomly divided into 80% and 20%, respectively. The training datasets were added and increased from time to time.

The trained models were then ported into the application program interface (API) of our vehicle classification module. In Phase 2, data collection was carried out at Toll Plaza Gombak with a different model training scheme. Specifically, the new dataset was designed to perform the auto labeling process where the images captured by the IP camera automatically generated the label that contained the information of the images. The auto labels at this point are generated by inferring from the trained model in Phase 1. These include the class of the vehicle and the number of wheels.

Next, the predicted label and the ground truth that were extracted from manual operator at the toll booth were compared to evaluate the performance of the trained model. In the case where the compared labels are non-equivalent, the images with the incorrect label will be re-labelled and the pipeline continues with a re-training where the parameters of the model will directly be updated to the new corrected example. We coined this training data augmentation as incremental training strategy. This training strategy can also be thought of as a semi-supervised or weakly-supervised training scheme where the labels for the new data are predicted based on the previously trained model, and the newly generated labels will be used to re-train the model in the next phase (Zhang, 2014). For every week, the data were collected and checked for any error that can occur during the process.

For the training of the VehicleDetNet, we used DLA-34 proposed by Zhao et al. (2021). It is used as a backbone for feature extraction to extract the features of vehicles to be classified. DLA is the images classification network with hierarchical skip connections and dense prediction, which increases the feature map resolution symmetrically.

PERFORMANCE EVALUATION

The performance evaluation and accuracy of the architecture are based on two approaches which are validation set and ground truth. The total number of correct prediction vehicles is divided by the total number of predictions made to calculate the accuracy based on the validation set. However, as the classification algorithm relies on a vehicle detection model, there are "no detection" cases where the prediction did not occur even though the vehicles were present.

Therefore, the second approach aims to compare the classification accuracy. The number of vehicles that were not detected by the algorithm needs to be included in the calculation to achieve minimal error when we deploy in real-time. Specifically, the accuracy of the algorithm is calculated using Eq. 4.

$$Accuracy = \frac{Total\ number\ of\ correct\ prediction}{Total\ prediction + not\ predicted} \quad (4)$$

RESULT AND ANALYSIS

The performance of the proposed model is taken as a mean classification accuracy. Table 1, 2, and 3 show the confusion matrix based on the three different models, which are Incremental_I, Incremental_II and Incremental_III representing the model trained using the incremental training strategy in the first month, second month and third month, respectively. Those models have differences in training datasets and various dataset types based on the location the data were collected, as shown in Table 4. We use confusion matrix to measure the performance of the classification accuracy of the proposed model, from which we calculate the performance metrics of recall, precision, accuracy, and ROC curve. It is worth mentioning that those performance metrics were calculated over the real-time data as testing dataset, as compared to a fixed testing dataset for all experiments, where the ground truth is extracted from the real-time classification made by the human operator at the tool booth. The results are tabulated as shown below in Table 1, 2 and 3.

The evaluation of the performance method by using the second approach was applied because the accuracy of the result by using Vehicle Classification System must achieve the current system used in toll plaza, which is 99.7%. The result of Vehicle Classification is compared with the real-time datasets obtained from the current toll system. In addition, the undetected vehicles could be identified since the toll operator keyed in the vehicle type even though it was not detected by the camera. It can be seen in the "no detection" column in the table below:

TABLE 1. Confusion Matrix for Incremental_I

Confusion Matrix		Actual Class					
		class 1	class 2	class 3	class 4	class 5	Total Vehicles
Predicted	class 1	10567	502	9	34	15	
	class 2	14	1671	181		25	
	class 3	3	29	2186	5	4	
	class 4	9	12	10	475	23	
	class 5	23	36	12	4	452	
	class 6						
	No detection	72	93	33	3	11	
Total of each class:		10688	2343	2431	521	530	16513
Accuracy		98.90%	71.30%	89.90%	91.20%	85.30%	
Error		1.10%	28.70%	10.10%	8.80%	14.70%	

TABLE 2. Confusion Matrix for Incremental_II

Confusion Matrix		Actual Class					
		class 1	class 2	class 3	class 4	class 5	Total vehicles
Predicted	class 1	4058	5	1	4		
	class 2	1	1523	7			
	class 3		15	1430			
	class 4	7			128		
	class 5	3				53	
	class 6						
	No detection	11	16	10			
Total of each class:		4080	1559	1448	132	53	7272
Accuracy		99.50%	97.70%	98.80%	97.00%	100.00%	
Error		0.50%	2.30%	1.20%	3.00%	0.00%	

TABLE 3. Confusion Matrix for Incremental_III

Confusion Matrix		Actual Class					
		class 1	class 2	class 3	class 4	class 5	Total vehicles
Predicted	class 1	3615	25	1	19		
	class 2	3	1440	39			
	class 3		103	1546			
	class 4	5			130		
	class 5	2				60	
	class 6						
	No detection	44	202	28	1	1	
Total of each class:		3669	1770	1614	150	61	7264
Accuracy		98.50%	81.40%	95.80%	86.70%	98.40%	
Error		1.50%	18.60%	4.20%	13.30%	1.60%	

TABLE 4. Number and type of datasets for training models

Models Training	Train	Validation
Incremental I	6889	1722
Incremental II	3049	762
Incremental III	9938	2484

Incremental_I was trained on the major datasets from toll plaza LDP (Puchong Barat). Due to that, the average accuracy is 87.3% which is suboptimal because we directly perform the domain adaptation to the toll plaza LPT (Gombak) without retraining to the specific target dataset. The performance of the trained model in the following month, Incremental_II that contain only 3811 training datasets from plaza toll LPT (Gombak) with 6 meters camera from the vehicles increased to 92.1%. However, due to the lighting condition and bad weather, the accuracy is still low. The performance did not achieve the target as we hypothesize that the number of datasets is not properly scaled. The next model which was trained on the third month, Incremental_III contains all three types of datasets, and the number of datasets increases to 12k. The results are plausible, where the accuracy increased to 99.07% and remain tested the model until the next two month, as shown in Figure 14. It is because the training datasets contain variation and the size of the training dataset is more properly scaled for training.

As we can observe from Table 3, the numbers of no detection are still significant. This is because in real-world implementation, there are lots of variations of data that need to be captured by the trained model. Therefore, we conjecture that the performance can be further enhanced by growing more training datasets in various environment and lighting conditions. Particularly, there is low number of examples of Class 4. As we can see, the accuracy for all models is not satisfactory for Class 4. There are multiple reasons that lead to this phenomenon. There are cars which were registered as taxis but do not have any taxi features. Thus, the trained model classified those examples as Class 1 while the toll operator classified as Class 4. The Class 4 type vehicle have three significant features, which differ from the Class 1 type vehicle. The differences are the physical appearance and colour, "Taxi" sign and white plate number.

In the first position of the camera, the detection for Class 4 type vehicle solely depended on their colour and appearance as the camera position was too far from the vehicle. The sign of "Taxi" on top of the car and the colour of the plate number becomes clear in the camera's current position.

In addition, the numbers of accuracy for Class 2 in all the above confusion matrix tables also resulted in lower accuracy. This is because, in some situations, the model was not able to detect the wheels of vehicles due to occlusion or abrupt lighting condition.

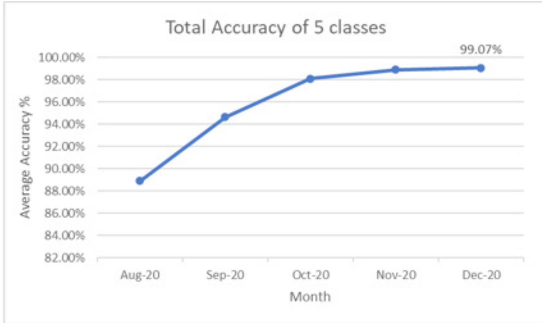


FIGURE 14. Average accuracy for five classes

TABLE 5. Summary performance of 3 Models based on validation evaluation

Based on Validation datasets		Precision	Recall (TPR)	Type I error (FPR)	Type II error (FNR)	Accuracy	Specificity (TNR)
Macro average metric	Incremental_I	82.1	61.4	8	38.6	88.5	97.8
	Incremental_II	99.6	98.9	0.3	1.1	99.7	100
	Incremental_III	99.7	99.5	0.9	0.4	99.8	99.9
Micro average metric	Incremental_I	71.5	70.9	7.1	29.1		92.9
	Incremental_II	99.3	99.3	0.2	0.7		99.8
	Incremental_III	99.6	99.6	0.08	0.3		99.9

To address the issue of class imbalance in our testing datasets, the evaluation is divided into two sections: Macro average and Micro Average Metric as shown in Table 6. The usage of average in this research due to imbalance class which is the ratio is 4:3:1:1:1 for respective Class 1 until Class 5. From this analysis, we can observe that the average Macro value is lower than the average value of Micro for recall on Incremental_I. It means that the model performs poorly on the classes with smaller number of examples.

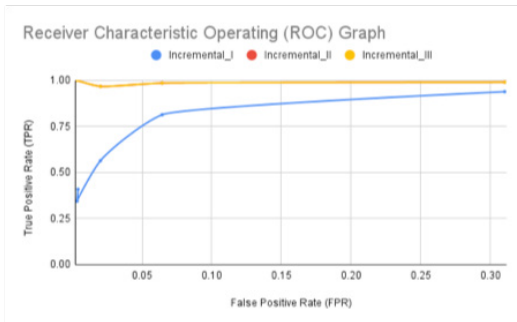


FIGURE 15. ROC Graph

ROC Curves summarized the interaction of the true positive rate for the predictive model with different probability threshold. The real positive rate is calculated by dividing the number of genuine positive elements into real positive and false negative elements. It describes how good the framework can be when positive results are predicted.

It can be observed in Figure 15, the ROC curve for Incremental_II and Incremental_III have better performance compared to Incremental_I as the curves closer to 1.

CONCLUSION

This paper devised an anchorless vehicle detection and classification algorithm coined as VehicleDetNet that utilizes a state-of-the-art DLA-34 model as backbone. To cater the challenging problem of continuous growing dataset in real-world environment, we have proposed an incremental training scheme to train our VehicleDetNet in a continual manner. The proposed vehicle detection and classification algorithm has been evaluated on a new and realistic CUTe Vehicles Dataset which was gathered from two real highway tolls in Malaysia.

From our experimental results, it is shown that our model can achieve 99.07% over three months of evaluation. Another additional advantage of this research is the implementation of the system required a low-cost maintenance. In future work, the performance of the technique could be improved by implementing and designing the IoT-Based for real-time monitoring and controlling the system.

ACKNOWLEGDMENT

We would like to extend our appreciation to Collaborative Research in Engineering, Science and Technology (CREST) for sponsorship (CREST18-001-0001), Delloyd Industry Sdn.Bhd and Ministry of Higher Education of Malaysia (FRGS19-159-0768) for the guidance and support. We also thank NVIDIA for donating the GPU used for this research.

DECLARATION OF COMPETING INTEREST

None

REFERENCES

Arinaldi, A., Pradana, J. A., & Gurusinga, A. A. 2018. Detection and classification of vehicles for traffic video analytics. *Procedia Computer Science* 144: 259–268. https://doi.org/10.1016/j.procs.2018.10.527

Awang, S., & Nik Azmi, N. M. A. 2018. Automated toll collection system based on vehicle type classification using Sparse-Filtered Convolutional Neural Networks with Layer- Skipping Strategy (SF- CNNLS) automated toll collection system based on vehicle type classification using sparse-filtere. *Journal of Physics*, 1–6.

Bornschein, J., Visin, F., & Osindero, S. 2020. Small data, big decisions: Model selection in the small-data regime. ArXiv.

Chen, X., Kong, X., Xu, M., Sandrasegaran, K., & Zheng, J. 2019. Road vehicle detection and classification using magnetic field measurement. *IEEE Access* 7(c): 52622–52633. https://doi.org/10.1109/ACCESS.2019.2908006

Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., & Tian, Q. 2019. CenterNet: Keypoint triplets for object detection. Proceedings of the IEEE International Conference on Computer Vision, 2019-October, 6568–6577. https://doi.org/10.1109/ICCV.2019.00667

Eom, H., Figueiredo, R., Cai, H., Zhang, Y., & Huang, G. 2015. MALMOS: Machine learning-based mobile offloading scheduler with online training. *Proceedings - 2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, MobileCloud 2015*, 51–60. https://doi.org/10.1109/MobileCloud.2015.19

Girshick, R., Donahue, J., Darrell, T., & Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 580–587. https://doi.org/10.1109/CVPR.2014.81

Gupta, A., Anpalagan, A., Guan, L., & Khwaja, A. S. 2021. Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. *Array* 10(September 2020): 100057. https://doi.org/10.1016/j.array.2021.100057

Hussain, K. F., & Moussa, G. S. 2005. Automatic vehicle classification system using range sensor. *International Conference on Information Technology: Coding and Computing*, 1–6.

Idris, R. 2019. The number of vehicles for the festive season increased to 20 percent. *Sinar Harian*, 1.

Junyeon Hwang, Kunsoo Huh, and Donghwi Lee. 2019. Vision-based vehicle detection and tracking algorithm design. *Optical Engineering* 48(12): 127201 (1 December 2009). https://doi.org/10.1117/1.3269685

Kaewkamnerd, S., Chinrungrueng, J., Pongthornseri, R., & Dumnin, S. 2010. Vehicle classification based on magnetic sensor signal. *2010 IEEE International Conference on Information and Automation*, ICIA 2010, (June): 935–939. https://doi.org/10.1109/ICINFA.2010.5512140

Kim, J., Nam, Y., Kang, M., Kim, K., Hong, J., Lee, S., & Kim, D. N. 2021) Adversarial Defect detection in semiconductor manufacturing process. *IEEE Transactions on Semiconductor Manufacturing*.

Kuo, C.-W., & Tang, M.-L. 2011. Survey and empirical evaluation of nonhomogeneous arrival process models with taxi data. *Journal of Advanced Transportation* 47(June): 512–525. https://doi.org/10.1002/atr

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. 2016. SSD: Single shot multibox detector. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-46448-0_2

Ministry of Works Malaysia, *Malaysia ITS Blueprint: Multilane Free Flow (MLFF)* (Kuala Lumpur: Ministry of Works, Malaysia, 2019),41.

Ng, J. Y., & Tay, Y. H. 2012. Image-based Vehicle Classification System. 1–11. http://arxiv.org/abs/1204.2114

PLUS Malaysia Berhad. (Vehicle Classes.) Accessed November 30,2021, https://www.plus.com.my/index.php?option=com_content&view=article&id=42&Itemid=148&lang=en.

Qu, H., Rahmani, H., Xu, L., Williams, B., & Liu, J. 2021. *Recent Advances of Continual Learning in Computer Vision: An Overview*. http://arxiv.org/abs/2109.11369

Salleh, I. I., Mohd Yusoff, K. U.-A., & Abdul Aziz, B. Z. 2006. Electronic Toll Collection (Etc) Systems Development in Malaysia. PIARC International Seminar on Intelligent

Sandhawalia, H., Rodriguez-Serrano, J. A., Poirier, H., & Csurka, G. 2013. Vehicle type classification from laser scanner profiles: A benchmark of feature descriptors. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, (Itsc), 517–522. https://doi.org/10.1109/ITSC.2013.6728283

Vikram, M. M., & S, A. V. 2018. *Project Report A YOLO Based Approach for Traffic Sign Detection*

Zhang, Z. 2014. Weak Perspective Projection. Computer Vision, 877–883. https://doi.org/10.1007/978-0-387-31439-6_115

Zhao, K., & Yan, W. Q. 2021. Fruit detection from digital images using CenterNet. *Geometry and Vision* 1386: 313

Zhou, X., Wang, D., & Krähenbühl, P. 2019. Objects as points. http://arxiv.org/abs/1904.078