Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Performance Comparison of Deep Neural Networks on Lane Detection for Driving Scene

Chong Cai Xuan[1], Fauzan Ahmad[1,*]

[1] Department of Electronic System Engineering, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur

ARTICLE INFO

Article history:
Received 27 January 2022
Received in revised form 29 June 2022
Accepted 14 July 2022
Available online 26 July 2022

Keywords:
Deep neural network; lane detection; deep learning; autonomous vehicle

ABSTRACT

One of the main challenges in developing autonomous vehicles is lane detection. Various methods have been used for lane detection such as sensor-based, feature-based and model-based. The emergence of deep neural network approaches had shown some promising results in lane detection. In this research, 2 popular deep neural network-based models namely, Efficient Neural Network (ENet) and Efficient Residual Factorized ConvNet (ERFNet) are selected for comparative study. The selected network models were validated with the TuSimple dataset. The raw image from the dataset was pre-processed with 3 methods, they are image resizing, channel-wise normalization and random data augmentation with random image rotation by 3 degrees. Both ENet and ERFNet are trained with 50 epochs and a batch size of 20 mixed-precision are implemented. The performance of trained models are evaluated in terms of accuracy, false positive rate (FP), false negative rate (FN), number of floating-point operations performed (FLOP), parameters count and speed of network models in terms of frame per second (FPS). ENet obtained an accuracy of 95.251% under the TuSimple dataset while ERFNet obtained an accuracy of 96.035%. ERFNet having a higher number of FLOP and parameters than ENet reflects that the ERFNet requires a larger computational cost than ENet. Both ENet and ERFNet have proven to capable to operate in real-time as they were able to run in 82.75 fps and 86.16 fps respectively. Both network models were compared to the state-of-the-art method, but only ERFNet remain competitive with others as it achieves a minimum accuracy of 96%.

1. Introduction

According to the Global Status Report on Road Safety 2018 published by the World Health Organization (WHO), there are 1.15 million lives lost due to road traffic injuries and the number of deaths steadily increased to 1.35 in 2016 [1]. The report also shows that the main cause of death for children and youth population (5-29 years old) is road traffic injuries. The United States National Highway Traffic Safety Administration (NHTSA) also pointed out that from 2005 to 2007, 94% of the causes of car crashes are driver related [2]. Humans are the most uncertain safety element found in a car, a common driver is most likely to have dangerous driving practices such as drunken driving, overspeeding, distracted while driving and drowsy driving. Such bad practice of driving will

* *Corresponding author.*
E-mail address: fauzan.kl@utm.my

cause the driver to fail at maintaining the car in the correct lane and keep the proper distance from other vehicles when cruising down the road. A vehicle collision is very likely to happen if a car is unexpectedly steering away from the road lane as a result of unattended lane detection.

The large number of deaths caused by road traffic injuries cannot be ignored. Car manufacturers had developed passive safety and active safety technology to increase the survival rate of car accidents. Passive safety usually takes place after the collision occurs, it helps to minimize the damage dealt with the driver, passengers and pedestrians such as seat belts and airbags. Active safety refers to the car crash prevention technology by assist the driver in steering or controlling the vehicle or warning the driver of a car crash, such as the Anti-lock Braking System (ABS) and adaptive cruise control (ACC). ABS helps to reduce the chance of wheels lock-up during braking and maintains the tractive contact between the wheels and road surface [3]. ACC is cruise control that helps drivers in the task of longitudinal control of the vehicle by slowing down and speeding up automatically [4].

Another important automotive safety system is the advanced driver-assistance systems (ADAS) and most of the time the terms “ADAS” and “active safety” are used interchangeably. ADAS takes advantage of the sensors and cameras on cars to enable the possibilities of an automated technology. Through a safe human-machine interface, ADAS will warn the driver. For instance, the lane departure warning system mentioned above will send a signal to alert the driver to drive in the middle of the lane. Other ADAS safety systems also include blind-spot assist [5], automatic parking [6], and adaptive cruise control [4], lane departure warning system [7] and lane-keeping assist system [8]. Lane detection plays an important role in the lane departure warning system and lane keep assist. In the ADAS system, the lane departure warning system will monitor the lane marking in front of the vehicle by using lane detection and warn the driver if the vehicle drifts out of its respective lane. The lane-keeping assist system will take control of the vehicle from the driver to automatically steering or braking to keep the vehicle in lane. A few premium class model vehicles are equipped with these systems, such as Chevrolet Malibu Hybrid [9], Genesis G70 [10] and Tesla model 3 [11].

Lane detection is the technique of locating the lane markers on the road and presenting the result to an intelligent system. The harsh condition on the road may affect the accuracy of the lane detection, such as fog, rain, shadow and sun glare. Worn out lane marks on the road also would hinder the lane detection system. One of the lane detection techniques is using a spinning multi-channel LiDAR sensor which is installed on top of the vehicle to scan its surroundings [12]. The detected lane lines are used to generate digital maps with the help of GPS sensors. The LiDAR-based lane detection works very well in low light and detecting multiple lanes but the LiDAR sensor and high precision GPS sensor are known to be expensive. Other than that, model-based lane detection techniques such as Hough Transform [13] and Kalman Filter [14]. These model-based techniques perform well in locating the lane from clear lane lines, roads that have unclear lane marks reduce its accuracy. These model-based techniques usually are more complex and require hand-crafted processes and also consume high processing time.

The advancement of the deep neural network allows a simpler, faster and more efficient approach to lane detection without having any hand-crafted process. Most deep learning methods can learn from a large training dataset unsupervised by a human. The examples of the deep learning methods are Convolutional Neural Networks (CNN) [15], Mask Region-Based Convolutional Neural Networks (R-CNN) [16], Fully Convolutional Neural Network (FCN) [17], Efficient Neural Network (ENet) [18] and Efficient Residual Factorized ConvNet (ERFNet) [19]. There are many choices for developing a lane detection algorithm or technique and critical implementations have not been discussed and studied. These factors play an important role in improving lane detection performance.

Lane detection is the key factor in developing autonomous vehicles. The challenges arise with the high cost of the sensor-based method and the low accuracy of the model-based method under bad weather. The deep neural network method seems to be very promising in tackling this task. In this research, 2 popular deep neural networks namely Efficient Neural Network (ENet) [18] and Efficient Residual Factorized ConvNet (ERFNet) [19] are selected for performance comparison. Both ENet and ERFNet will be studied in this research to attempt to solve 2 problems. First, the performance of deep neural networks namely ENet and ERFNet in the lane detection task. Second, the robustness of ENet and ERFNet in performing lane detection under various road conditions.

2. Methodology

Two deep neural networks are selected and compared to evaluate their performance on lane detection namely ENet and ERFNet. TuSimple dataset [20] is used for training and validating both selected deep neural networks. The raw dataset from TuSimple is preprocessed before using it as training images for ENet and ERFNet. The preprocessing methods are image resizing, channel-wise data normalization and random image rotation by 3 degrees. All the processed training data was fed to the network model to learn the images of road lanes. Once the training process is complete, the network model is then benchmarked with the test dataset provided by TuSimple. This benchmark result should provide an insight into the performance of the network model. If the performance of the network model does not satisfied at least 80 % accuracy, the parameters of the network model will be adjusted. The training process and benchmarking is then repeated with another compared network model. The research flow of this research is shown in the Figure 1.

2.1 Dataset

The learning dataset used for this research was provided by TuSimple Lane Detection Challenge [20]. The TuSimple dataset consists of 3626 video clips with 3626 annotated frames for the training data and 2782 video clips for the testing data. The captured video clips from the dataset contain different traffic conditions with a maximum 4 lane highway road under good and medium weather conditions. All the video clips are captured by cameras that are mounted on a vehicle dashboard. The labels of the dataset are given in the form of polylines that mark the lane line in the images and the labels are recorded in the JSON format file.

2.2 Data Pre-processing

2.2.1 Image resize

The dimension of the original image from the TuSimple dataset is 1280 × 720. The images were resized to a smaller size of 640 × 360. This is to reduce the computational resources required in the model training process. If the training image is too large, the memory of the graphics processing unit (GPU) would eventually run out as that is where the training process takes place.

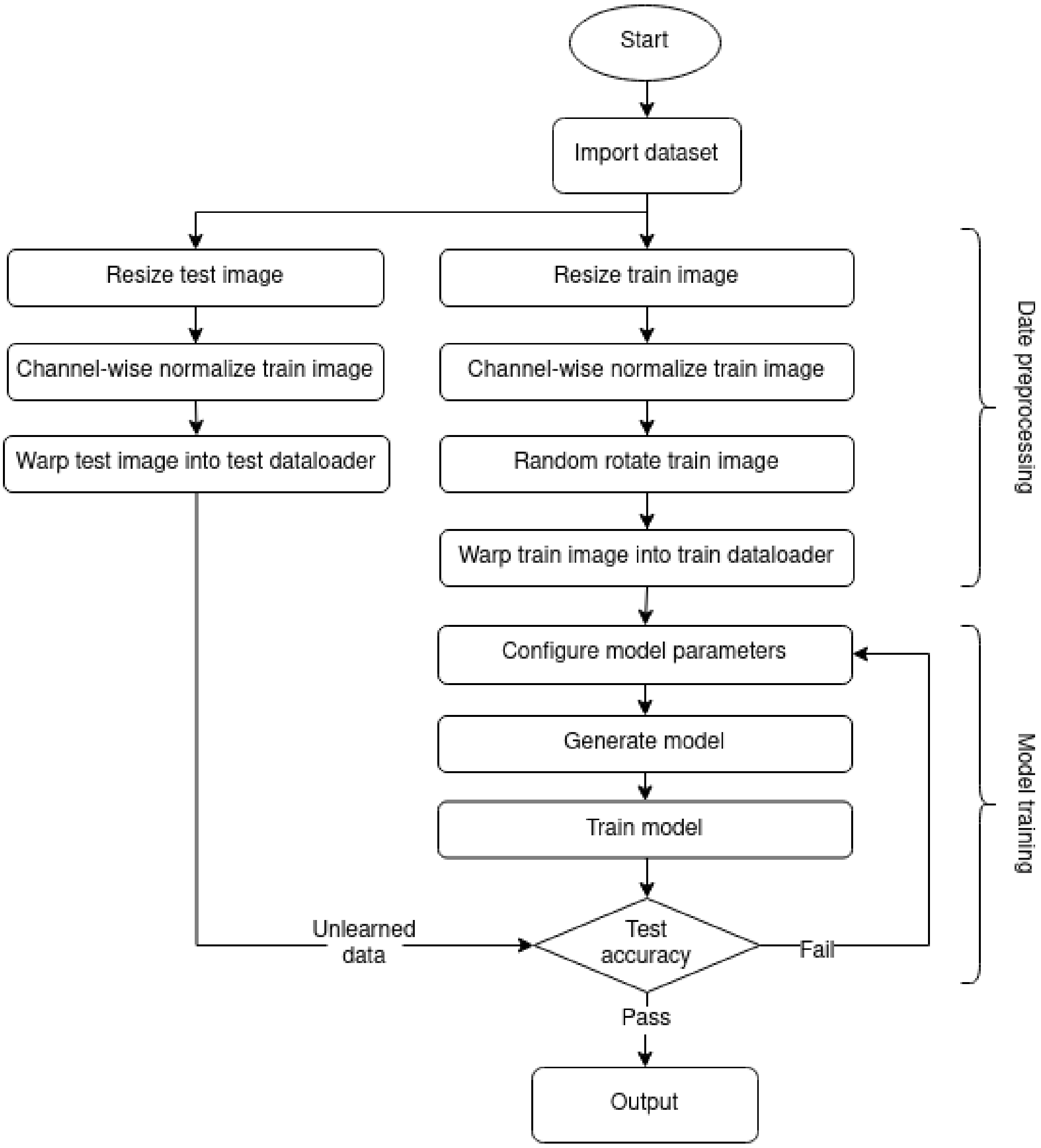


Fig. 1. Research flowchart

2.2.2 Channel-wise normalization

A channel-wise data normalization is carried out to improve the cohesion of the dataset and thus reduce the inconsistency of the dataset. Pontalba *et al*. [21] also proved that normalization does aid in deep learning classification. TuSimple dataset image consists of 3 channels as they represent the red, green and blue channels that carry the colour information of the images.

2.2.3 Randomized data augmentations

A randomly selected transformation will be applied while importing the training dataset to the model. The transformation is performing a small rotation of 3 degrees to a randomly selected training image. As the selection of rotating images is random and will not be the same for every epoch of training, it exposes the model with slightly different images which help it generalize the training data better.

3. Results

The benchmark results of both ENet and ERFNet with TuSimple dataset are presented. The training loss of both network models is also compared. The performance of both network models are evaluated in terms of number of floating-point operation performed (FLOP), number of parameters and speed of network models in frame per seconds (fps). Both ENet and ERFNet were tested with TuSimple and carry out on Google Colaboratory platform.

3.1 Performance Benchmark Results

The performance benchmark result that evaluates ENet and ERFNet in terms of speed, the number of floating-point operation performed (FLOP) and the number of parameters is shown in Table 1. The benchmark test is carried out on NVIDIA Tesla T4 GPU that provided on the Google Colaboratory cloud computing platform. According to Table 1, ERFNet having a slightly higher frame per seconds (fps) than ENet by 3 fps. This indicates that ERFNet computes slightly faster than ENet in the tested GPU. Both neural networks are considered to have high frame rates that enable real-time applications as the minimum speed to achieve is 30 fps. This also ensure the practical uses of such encoder-decoder deep neural networks architecture. But when come to the numbers floating-point operation and parameters, ENet has a smaller number compared to ERFNet. ERFNet that having a larger number of FLOP and parameters requires stronger hardware to handles the computation. This also shows that ERFNet will be harder to fit into an embedded processor for real-time lane detection as a very strong processor is required, whereas ENet having a higher chance as its computational cost is lower.

Table 1
Performance benchmark result of ENet and ERFNet

Network Model	Image Resolution	FPS	FLOP (G)	Parameters (M)
ENet	640 × 360	82.75	4.25	0.95
ERFNet	640 × 360	86.16	26.32	2.66

3.2 Training Loss

The training loss of both ENet and ERFNet with the number of iterations were recorded. The graph of training loss against the number of iterations of both network models is shown in Figure 1. The recorded training loss for ENet at 8191st iterations is 0.04266; ERFNet at 8191 iterations is 0.05606. Figure 1 also shows the trend of training loss of ENet is lower than ERFNet. The training loss reflects hows well the network model is learning from the training data. Generally, the lower the loss of the network model, the better the network model. But when comparing the results from Table 2, ERFNet is performing better than ENet in terms of accuracy. ENet that having low training loss and low accuracy is reflecting that the network model is slightly over-fitting. Over-fitting causes the network model to start learning the noise within the network model resulting in decreased performance. The over-fitting may indirectly be caused by the fixed number of training epoch. The over-fitting issue can be reduced by training with less epoch or early stopping of the training before over-fitting occur.

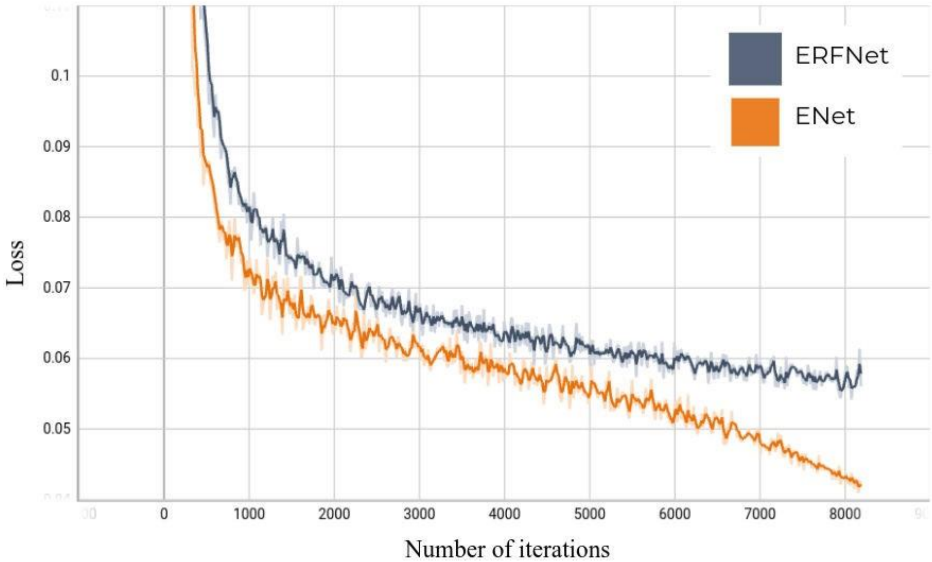


Fig.1. Graph of training loss against the number of iterations of ENet and ERFNet

3.3 Test Results with TuSimple Dataset

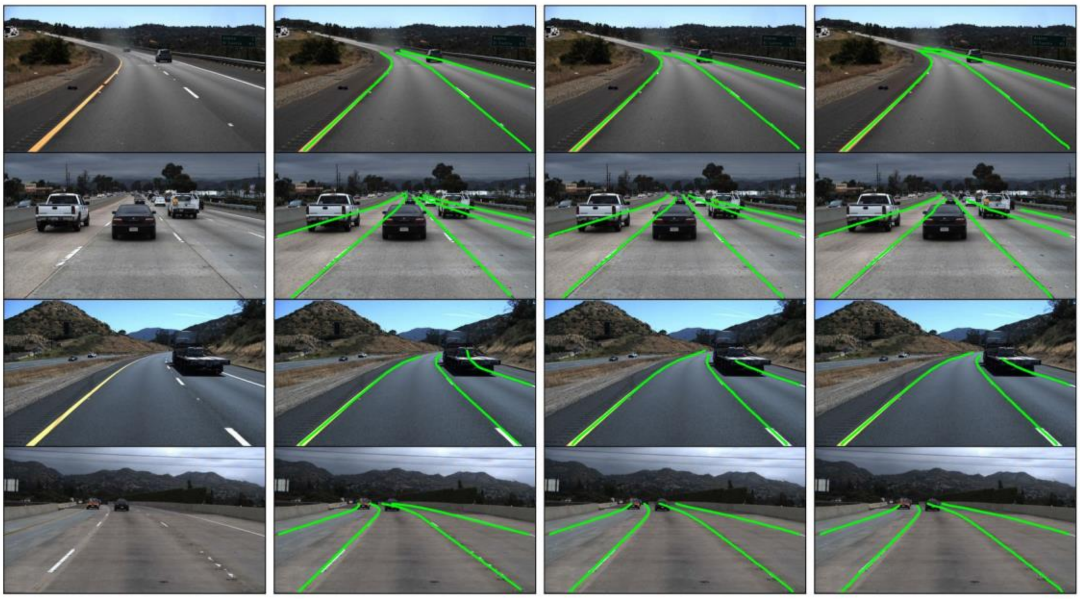
The benchmark results of ENet and ERFNet on the TuSimple Dataset is shown in Table 2. According to Table 2, ERFNet has a higher accuracy of 96.045 % while ENet having a lower accuracy of 95.251 %. This shows that the ERFNet performs better than ENet in predicting the lane on TuSimple test images. Other than accuracy, ERFNet also has a false positive rate of 7.341 % and false negative rate of 5.571 % whereas ENet has a false positive rate of 6.101 % and false negative rate of 3.601 %. These results also reflect that ERFNet's predictions having less error when compared to Enet. In other words, ERFNet is proven to perform better than ENet in accurately predicting the lanes on a driving scene of the TuSimple dataset.

Table 2
TuSimple dataset test results of ENet and ERFNet

Network Model	Accuracy (%)	FP (%)	FN (%)
ENet	95.251	7.341	5.571
ERFNet	96.035	6.101	3.601

Various example of the prediction of ENet and ERFNet on TusSimple dataset is shown in Figure 2. An analysis of quality of the lane detection is carry out. The results in Figure 2 demonstrate both ENet and ERFNet able to detect the road lane in front of the vehicle accurately. But in the distant scene, ENet tends to give a coarser prediction. Both network models demonstrate their ability in detecting lane line in driving scene, ERFNet is performing better than ENet in accurately detecting lane lines.

The performance of ENet and ERFNet on TuSimple test data is compared with other methods and the comparison is show in Table 3. The comparison shown that neither ERFNet nor ENet are not the best method performed in TuSimple test data set. Both ERFNet and ENet having a quite large value of FP and FN which indicates that more works are require in reducing the false prediction. ERFNet able to stay competitive with other methods to have at least 96% accuracy.



(a) Raw Image (b) Ground Truth (c) ENet (d) ERFNet

Fig. 2. Example of predicted lane marks on TuSimple dataset by ENet and ERFNet

Table 3
Comparison of different methods on TuSimple test set (did not sorted in any particular order)

Network model	Accuracy	FP	FN
EL-GAN [41]	96.39	**0.0412**	0.0336
SCNN [40]	96.53	0.0617	**0.0180**
PolyLaneNet [51]	93.36	0.0942	0.0933
FastDraw [52]	95.2	0.076	0.045
CondLaneNet-S [54]	95.48	2.18	3.80
CondLaneNet-M [54]	95.37	2.20	3.82
CondLaneNet-L [54]	**96.54**	2.01	3.50
LaneNet [51]	96.4	0.0780	0.0244
ENet [18]	95.251	7.341	5.571
ERFNet [19]	96.035	6.101	3.601

4. Conclusions

In conclusion, the encoder and decoder architecture of ENet and ERFNet is proven able to perform lane detection tasks to the degree of achieving at least 95 % in the TuSimple dataset. From the validation results from the TuSimple dataset benchmark, ENet obtained an accuracy of 95.251 %, FP of 7.341 % and FN of 5.571 % while ERFNet obtained an accuracy of 96.035 %, FP of 6.101 % and FN 3.601 %. The results show that ERFNet is performing better than ENet in terms of accuracy in detecting lane lines in the TuSimple test set. The performance benchmarks on both network models also found out that ERFNet having a higher number of FLOP and parameters than ENet, while ERFNet performs slight faster than ENet in terms of fps on tested GPU. This proves that ERFNet requires a large computational power than ENet in performing the lane detection task. Both ERFNet and ENet expected to have the capability to run in real-time as both having a speed that higher than 30 fps. ERFNet able to stay competitive with state-of-the-art methods by maintaining at least 96 % accuracy in lane detection on the TuSimple dataset while ENet is struggling competitive with slightly lower accuracy.

Acknowledgement
This research was not funded by any grant.

References

[1] W. Organization, Global Status Report on Road Safety 2018. Geneva: World Health Organization, 2019.
[2] Mathematical Analysis Division of the National Center for Statistics and Analysis, NHTSA, "Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey", NHTSA's National Center for Statistics and Analysis, Washington, DC, 2015.
[3] A. Aksjonov, V. Vodovozov, K. Augsburg and E. Petlenkov, "Design of Regenerative Anti-Lock Braking System Controller for 4 In-Wheel-Motor Drive Electric Vehicle with Road Surface Estimation", International Journal of Automotive Technology, vol. 19, no. 4, pp. 727-742, 2018. Available: 10.1007/s12239-018-0070-8. https://doi.org/10.1007/s12239-018-0070-8
[4] G. Marsden, M. McDonald and M. Brackstone, "Towards an understanding of adaptive cruise control", Transportation Research Part C: Emerging Technologies, vol. 9, no. 1, pp. 33-51, 2001. Available: 10.1016/s0968-090x(00)00022-x. https://doi.org/10.1016/S0968-090X(00)00022-X
[5] S. Raj, R. Ezhilarasie and A. Umamakeswari, "Zigbee-Based Collision Avoidance System in Blind Spot and Heavy Traffic using Ultrasonic Sensor", Indian Journal of Science and Technology, vol. 9, no. 48, 2016. Available: 10.17485/ijst/2016/v9i48/108015. https://doi.org/10.17485/ijst/2016/v9i48/108015
[6] S. Ma, H. Jiang, M. Han, J. Xie and C. Li, "Research on Automatic Parking Systems Based on Parking Scene Recognition", IEEE Access, vol. 5, pp. 21901- 21917, 2017. Available: 10.1109/access.2017.2760201. https://doi.org/10.1109/ACCESS.2017.2760201
[7] S. Narote, P. Bhujbal, A. Narote and D. Dhane, "A review of recent advances in lane detection and departure warning system", Pattern Recognition, vol. 73, pp. 216-234, 2018. Available: 10.1016/j.patcog.2017.08.014. https://doi.org/10.1016/j.patcog.2017.08.014
[8] A. Mammeri, G. Lu and A. Boukerche, "Design of lane keeping assist system for autonomous vehicles", 2015 7th International Conference on New Technologies, Mobility and Security (NTMS), 2015. Available: 10.1109/ntms.2015.7266483 [Accessed 24 December 2020]. https://doi.org/10.1109/NTMS.2015.7266483
[9] "Chevrolet Malibu Hybrid - 2019", media.gm.com, 2020. [Online]. Available: https://media.chevrolet.com/media/us/en/chevrolet/vehicles/malibu-hybrid/ 2019.html. [Accessed: 24- Dec- 2020].
[10] "Genesis G70", Genesis.com, 2020. [Online]. Available: https://www.genesis.com/ca/en/luxury-sedan-genesis-g70-safety.html. [Accessed: 24- Dec- 2020].
[11] T. T. Team, "More Advanced Safety for Tesla Owners," Tesla, Inc, 03-May- 2019. [Online]. Available: https://www.tesla.com/blog/more-advanced-safety- tesla-owners. [Accessed: 24-Dec-2020].
[12] J. Jung and S. Bae, "Real-Time Road Lane Detection in Urban Areas Using LiDAR Data", Electronics, vol. 7, no. 11, p. 276, 2018. https://doi.org/10.3390/electronics7110276
[13] N. Aminuddin, M. Mat Ibrahim, N. Mohd Ali, S. Ahmad Radzi, W. Mohd Saad and A. Darsono, "A NEW APPROACH TO HIGHWAY LANE DETECTION BY USING HOUGH TRANSFORM TECHNIQUE", Journal of Information and Communication Technology, 2017. Available: 10.32890/jict2017.16.2.8231.
[14] J. Xiao, L. Luo, Y. Yao, W. Zou and R. Klette, "Lane Detection Based on Road Module and Extended Kalman Filter", Image and Video Technology, pp. 382- 395, 2018. Available: 10.1007/978-3-319-75786-5_31 [Accessed 15 January 2021].https://doi.org/10.1007/978-3-319-75786-5_31
[15] Wang, W., Lin, H. & Wang, J. CNN based lane detection with instance segmentation in edge-cloud computing. J Cloud Comp 9, 27 (2020). https://doi.org/10.1186/s13677-020-00172-z
[16] B. Liu, H. Liu, and J. Yuan, "Lane Line Detection based on Mask R-CNN," Proceedings of the 3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019), 2019. https://doi.org/10.2991/icmeit-19.2019.111
[17] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," arXiv:1411.4038 [cs], Mar. 2015, Accessed: Jan. 16, 2021. [Online]. Available: http://arxiv.org/abs/1411.4038. https://doi.org/10.1109/CVPR.2015.7298965
[18] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation," arXiv:1606.02147 [cs], Jun. 2016, Accessed: Jun. 04, 2021. [Online]. Available: http://arxiv.org/abs/1606.02147

[19] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, "ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation," IEEE Trans. Intell. Transport. Syst., vol. 19, no. 1, pp. 263-272, Jan. 2018, doi: 10.1109/TITS.2017.2750080. https://doi.org/10.1109/TITS.2017.2750080
[20] "TuSimple Lane Detection Challenge," GitHub.https://github.com/TuSimple/tusimple- benchmark (accessed Jan. 17, 2021).
[21] J. T. Pontalba, T. Gwynne-Timothy, E. David, K. Jakate, D. Androutsos, and A. Khademi, "Assessing the Impact of Color Normalization in Convolutional Neural Network-Based Nuclei Segmentation Frameworks," Front. Bioeng. Biotechnol., vol. 7, p. 300, Nov. 2019, https://doi.org/10.3389/fbioe.2019.00300