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

Monocular Distance Estimation-based Approach using Deep Artificial Neural Network

Siti Nur Atiqah Halimi[1], Mohd Azizi Abdul Rahman[1,*], Mohd Hatta Mohammed Ariff[1], Nurulakmar Abu Husain[1], Wira Jazair Yahya[1], Khairil Anwar Abu Kassim[2], Mohd Azman Abas[3], Syed Zaini Putra Syed Yusoff[4]

[1] Malaysia Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
[2] ASEAN NCAP Operational Unit, Malaysian Institute of Road Safety Research, Kajang, Selangor, Malaysia
[3] School of Mechanical Engineering Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
[4] Techcapital Resources Sdn. Bhd., Universiti Putra Malaysia, Serdang, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 23 March 2023
Received in revised form 16 July 2023
Accepted 23 July 2023
Available online 30 August 2023

Keywords:
Autonomous Emergency Steering; Autonomous Emergency Braking; distance estimation; monocular vision; Deep Learning

ABSTRACT

Those in authority are evaluating the test evaluation for threat assessments currently in place. Since people often depend on their feelings and moods, this may create inequality. Therefore, this study suggested applying deep learning for Autonomous Emergency Steering (AES) and Autonomous Emergency Braking (AEB) assessments in the safety rating protocol. The suggested method for the test in situation-based threat assessments is a monocular distance estimation-based approach. The camera's objective is to make it simple to conduct assessments using only an onboard dash camera. This study proposes a method based on a monocular distance estimation-based approach for test methodology in the situational-based threat assessments using deep learning for the AES system to complement the AEB system for active safety features. Then, the accuracy of the distance estimation models has validated with the ground truth distances from the KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) dataset. Thus, the output of this study can contribute to the methodological base for further understanding of drivers the following behaviour with a long-term goal of reducing rear-end collisions.

1. Introduction

This research work appears to be part of the recently released National Automotive Program (NAP 2020) [1]. The NAP 2020 is a policy that promotes investment, technological improvement, and overall sustainable growth. It promotes new growth sectors by incorporating future development technologies such as Industrial Revolution 4.0 (IR4.0), Mobility as a Service (MaaS) and Next Generation Vehicle (NxGV). On the other side, this research is incorporated primarily in the Southeast Asian New Car Assessment Programme (ASEAN NCAP) Roadmap 2021-2025 under the Title 'Safety Assist' technical development effort in Southeast Asian nations, particularly Malaysia [2]. The ASEAN

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

https://doi.org/10.37934/araset.32.1.107119

NCAP, for example, focuses solely on AEB technology [3], which is a function that warns drivers of impending collisions and assists them in using the vehicle's full capability. The AES system is one of the active safety features that help with evasive steering. When a probable collision is detected, unlike AEB, the AES system will automatically steer to assist in avoiding an accident. Later, combining AEB and AES will improve intelligent mobility applications with at least Level 4 vehicle automation, which is high automation [4].

Furthermore, the AES system might be integrated into an Advanced Driver Assistance System (ADAS) for automation in the future of driverless cars. Many scholars, such as [5-8], have discussed AES vehicles' general design and viability. However, only a few AES and AEB intervention systems are now available, which might be why the lack of precise analyses and practical solutions, particularly from automakers. As a result of being benchmarked to the European New Car Assessment Programme (EURO NCAP), new evaluations for the ASEAN NCAP safety standards may be established and unified. This study provides a framework for situational-based evaluations and testing methods for AES demand to supplement AEB for prospective inclusion in the safety automobile rating project.

The assessment and those in authority are evaluating test evaluation for threat assessments that are currently in place. Since people often depend on their feelings and moods, this may create inequality. Therefore, this study suggested applying deep learning to assess the assessments for AES and AEB in the safety rating test protocol. The method is proposed for test methodology in situation-based threat assessments using a monocular distance estimation-based approach. The distance is calculated between the dash camera mounted in a car and the car in front of it using distance estimation. The camera's objective is to make it simple to conduct assessments and tests using only a dashcam.

The testing strategy will evaluate collision-avoidance scenarios with essential metrics like ideal braking distances to verify the methodology for active safety features. Consequently, following the NAP 2020 [1], particularly concerning Next-Generation Vehicles, the results of the approach used for assessment and test evaluation can be included in the safety vehicle rating criteria. As a result, the results of this study can help provide the methodological framework for a deeper comprehension of driver behaviours with the long-term purpose of reducing rear-end collision rates.

The remaining section of the article is organised as follows: In Section 2, the KITTI dataset, the Deep Artificial Neural Network (Deep ANN) architecture, and the distance estimation method used in this study are explained. Table 1 also provides the list of abbreviations used in the paper. Section 3 examines the outcomes achieved for each task and the proposed approach's overall performance. Lastly, Section 4 concludes with a statement of conclusion and further research.

Table 1
List of abbreviations used in the paper

Acronym	Explanation
2D	Two Dimensional
3D	Three Dimensional
ADAS	Advanced Driver Assistance Systems
AEB	Autonomous Emergency Steering
AES	Autonomous Emergency Braking
ASEAN NCAP	Southeast Asian New Car Assessment Program
CPU	Central Processing Unit
CSV	Comma-Separated Values File format
Deep ANN	Deep Artificial Neural Network
EURO NCAP	European New Car Assessment Programme
Faster R-CNN	Faster Region-based Convolutional Neural Network
GPS	Global Positioning System
IMU	Inertial Measurement Unit
IR 4.0	Industrial Revolution 4.0
KITTI	Karlsruhe Institute of Technology and Toyota Technological Institute dataset
m	Meters
MaaS	Mobility as a Service
MAE	Mean Average Error
MSE	Mean Squared Error
NAP 2020	National Automotive Program 2020
NxGV	Next Generation Vehicle
RAID	Redundant Array of Independent Disks
ReLU	Rectified Linear Unit
RMSE	Root Mean Squared Error
txt	Text file format
Xloc	3D object location for the x-axis in camera coordinates in meters
Xmax	Maximum x-coordinate value of an object's 2D bounding box in an image
Xmin	Minimum x-Coordinate Value of an object's 2D bounding box in an image
Yloc	3D object location for y-axis in camera coordinates in meters
Ymax	Maximum y-Coordinate Value of an object's 2D bounding box in an image
Ymin	Minimum y-Coordinate Value of an object's 2D bounding box in an image
Zloc	3D object location for z-axis in camera coordinates in meters

2. Methodology

There are currently two different approaches to estimating visual distance: monocular vision and binocular vision. Binocular vision uses two cameras positioned at different angles to capture pictures or movies from both views. Using a combination of stereo matching and camera calibration [9], disparity maps [10] are used to calculate distances, depth information [11] is used to estimate distances, and Faster R-CNN [12] is used to identify obstacle items before calculating distances. While binocular approaches get some relevant 3D information, [13] claims that they are frequently delayed in imaging due to the computation of position deviation. Contrarily, monocular visual distance estimate has benefits such as speed, simplicity, and affordability. As a result, the proposed method in this study for estimating vehicle distance is based on monocular vision and employs a deep learning-based strategy. By reconstructing 3D structures from a few geometry-restricted pictures, geometry-based approaches [14-16] detect depth. In order to achieve competitive accuracy, deep-learning-based systems [17-19] estimate the distance from monocular vision.

Deep ANN algorithms are used as the deep learning approach for the proposed method. The workflow used in this study is described in Figure 1. The KITTI dataset is employed to train and test the proposed method. Four different architectures with varying numbers of hidden layers are

designed to select the optimum architecture for the proposed method. Then, the outputs of all four algorithms are evaluated and compared.

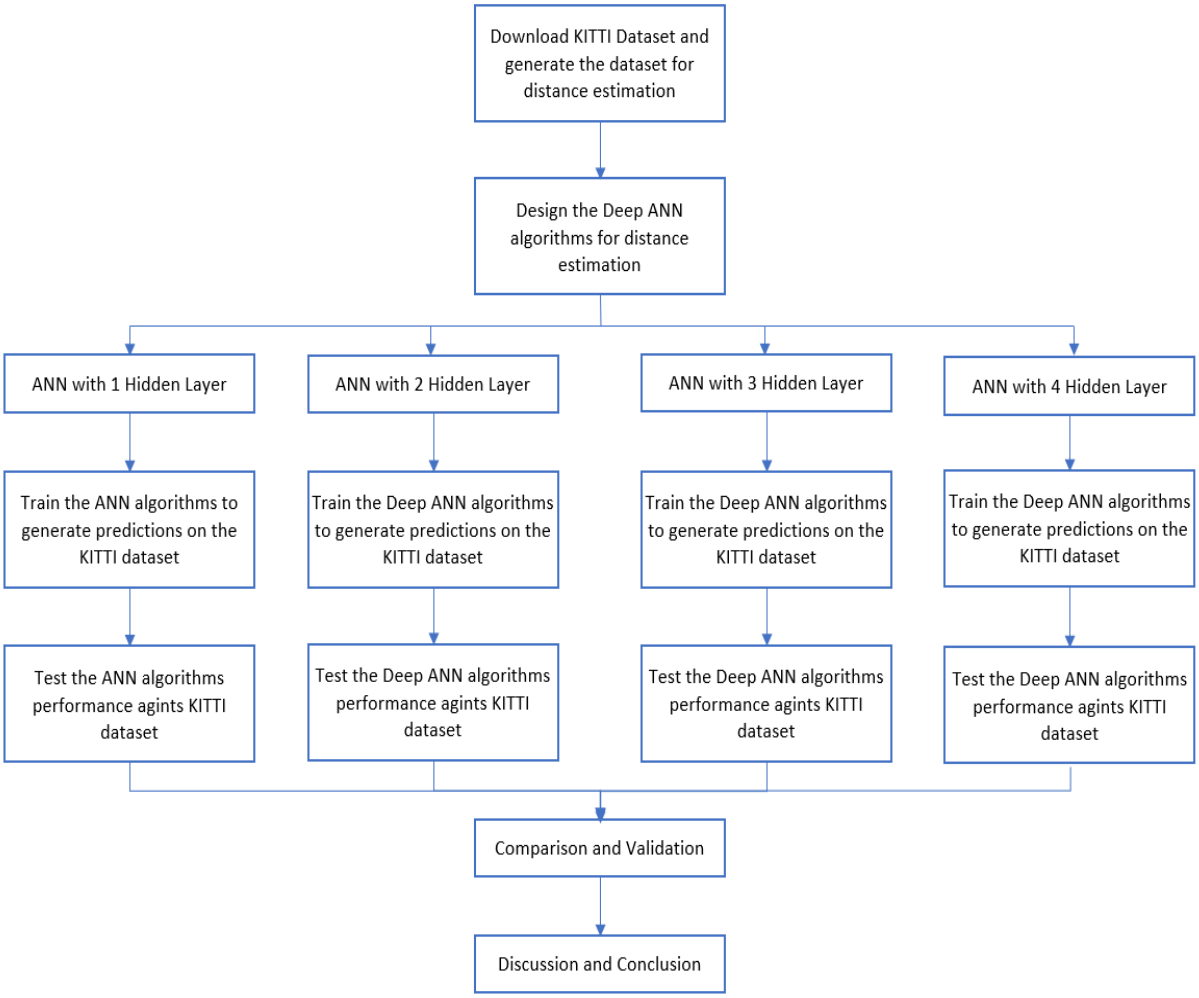

Fig. 1. Workflow for distance estimation using Artificial Neural Networks

2.1 KITTI Dataset

The most widely used dataset for mobile robots and autonomous driving is KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) [20]. A Volkswagen Passat B6, which has been retrofitted with pedal and steering wheel actuators, was used to record the KITTI dataset. An eight-core Intel i7 computer running Ubuntu Linux and a real-time database are used to capture the data. It also has a RAID system and eight CPU cores. Table 2 lists the sensors that were utilised to create the KITTI dataset.

Table 2
List of sensors used in the KITTI dataset

Sensors	Model	Quantity
Inertial Navigation System (GPS/IMU)	OXTS RT 3003	1
Lasers DEEP-ANNer	Velodyne HDL-64E	1
Grayscale cameras, 1.4 Megapixels	Point Grey Flea 2 (FL2-14S3M-C)	2
Color cameras, 1.4 Megapixels	Point Grey Flea 2 (FL2-14S3C-C)	2

Test images, train images, and train annotations are all included in the KITTI dataset's raw data for distance estimation. The train annotations should be converted from txt to CSV format. The train annotations in a CSV file comprise 51,865 data, whereas the images from the KITTI dataset have 7,481 images. Table 3 shows several examples of train annotation in CSV files.

Table 3
A few examples of train annotation in CSV File

Class	Truncated	Occluded	Observation Angle	Box				Dimensions			Location		
				Xmin	Ymin	Xmax	Ymax	Height	Width	Length	Xloc	Yloc	Zloc
Pedestrian	0	0	-0.2	712.4	143	810.73	307.92	1.89	0.48	1.2	1.84	1.47	8.41
Truck	0	0	-1.57	599.41	156.4	629.75	189.25	2.85	2.63	12.34	0.47	1.49	69.44
Car	0	0	1.85	387.63	181.54	423.81	203.12	1.67	1.87	3.69	-16.53	2.39	58.49
Cyclist	0	3	-1.65	676.6	163.95	688.98	193.93	1.86	0.6	2.02	4.59	1.32	45.84
DontCare	-1	-1	-10	503.89	169.71	590.61	190.13	-1	-1	-1	-1000	-1000	-1000

'Car', 'Van', 'Truck', 'Pedestrian', 'Person sitting', 'Cyclist', 'Tram', 'Misc', and 'DontCare' are the nine types of items that are mentioned in the column for Class. The next column is labelled "Truncated," and it contains float numbers ranging from 0 (not truncated) to 1 (truncated), where truncated denotes an item that has left the borders of the picture. The integer numbers 0 through 3 in the Occluded column indicate the degree of occlusion: 0 indicates complete visibility, 1 indicates some occlusion, 2 indicates substantial occlusion, and three indicates unknown. The observation angle of the object, which ranges from $-\pi$ to π, is shown in the Observation Angle column. The Box column displays the left, top, right, and bottom pixel coordinates of an object's 2D bounding box in an image. The 3D object dimensions were displayed in the Dimensions column, including height, width, and length in metres. The 3D object location is shown in the final column for Location as x, y, and z in camera coordinates in meters.

However, just the Box and Location columns are required to create a dataset for depth estimation. The new dataset is developed to focus on AEB and AES evaluation. Thus, only the class of 'Car' is utilised to divide the dataset into train and test. Specifically, 70% of the train and 30% of the test datasets are divided. Thus, out of 51,865 data points, only 28,742 have the class' car.' The class 'Car' data comprise 20,080 data for the training dataset and 8,662 data for the test dataset.

2.2 Deep ANN

The choice of hidden layers is a particularly challenging issue since there is a risk of overfitting and underfitting, negatively impacting the network's efficiency and time complexity [21]. These conditions can happen when there are too many hidden layers. The overtraining of the network begins as a result of the overfitting condition when the number of hidden layers is excessively high relative to the task's difficulty. It harms the network's time complexity and often happens when its performance is closely matched to the test data [22]. Underfitting conditions develop when the network's hidden layer count is fewer than the problem's complexity [23]. Such issues are hardly ever handled by the network because it is sometimes referred to as undertraining. The network's effectiveness is severely impacted by it. The network's temporal complexity becomes extremely low and yields ineffective results in such a situation. Fundamentally, the primary goals of this study are

to determine the number of hidden layers present in the network and the impact of those hidden layers on the network.

The dataset in this work is trained and tested using the proposed distance estimation algorithms with Deep ANN architecture, as shown in Figure 2. The input variables are included within the Input Layer. The Input Layer is occasionally referred to as the Visible Layer. Four inputs, representing values from the Box column, serve as the input variables for the proposed method (Xmin, Ymin, Xmax, Ymax). Meanwhile, the Output Layer is a collection of neurons that generate the output variables. The value of the distance from the camera (Zloc) is the output variable produced by the Deep ANN architecture in this study. The Hidden Layers are the neuron layers between the input and output layers. The Hidden Layers enable a neural network's function to split into particular data processing.

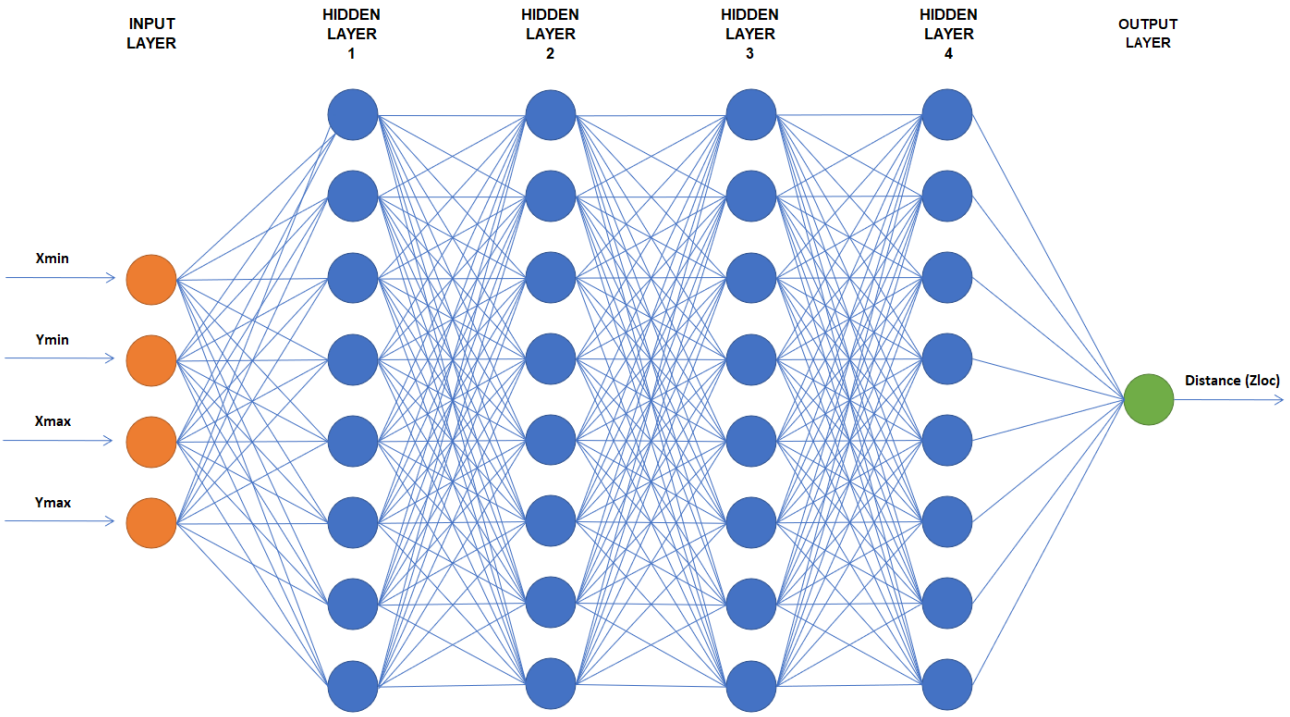

Fig. 2. The Deep ANN architecture

Figure 3 illustrates the structure of the neuron. As indicated, every neuron will receive inputs consisting of weights and features. According to [24], if a neuron is in the network's first layer, its features will be the "real features" of the dataset. Otherwise, if it is not in the first layer, its features will be the output of other neurons from previous layers. Each ANN layer will be assigned a unique weight. Each neuron must perform a specific nonlinear computation to calculate multiple complicated functions. Here, activation functions are put into action. After computing the Linear component, the neuron transmits this value to an activation function. The activation function performs additional computations on the received value before outputting it. One of the tuning parameters is the number of hidden layers and the number of neurons included inside each of those layers.

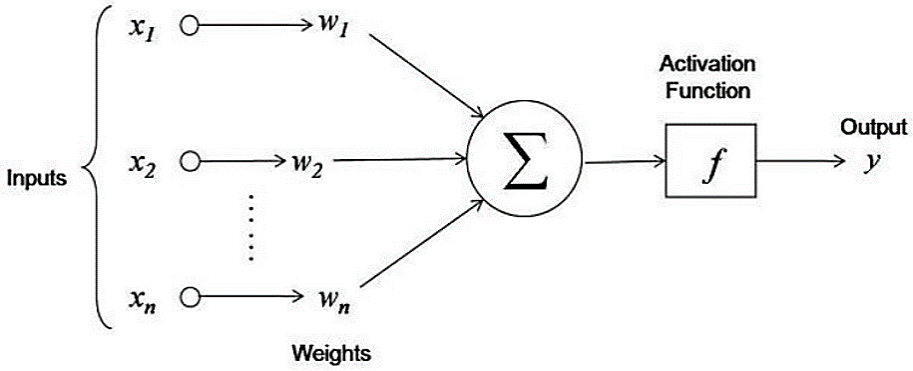

Fig. 3. Neuron structure in ANN architecture [25]

Additionally, activation functions were applied to boost the neural network's nonlinearity. Finding the optimum network design and configuration is difficult since there is no single clear-cut option. Instead, a few different Hidden Layers comparisons are made to determine which network is the best. The outcome showed that the most effective model had four hidden layers, ReLU activations, a batch size of 2048, and 50 epochs.

2.3 Distance Estimation

Assuming the camera is as good as a theoretical pinhole camera, this research finds that it can roughly calculate the distance to an item of a certain width from the camera. This sub-section aims to learn the formula for calculating the Z value. Figure 4 illustrates the possibilities to determine how to convert a point in the two-dimensional plane (u, v) to a three-dimensional (X, Y, Z) coordinate system. This transformation is described by a mathematical model that can be represented as an Eq. (1) [26].

$$p = K[t] \times Q \tag{1}$$

where p is a projected $[u, v, 1]$ point and K represents the camera's inherent features, often the focal length and the main point or optical centre (u_o, v_0). The description of a camera often lists these characteristics. The transition from an external world point to an internal camera viewpoint is described by the parameters $[R|t]$, which are extrinsic to the system. The Euclidean coordinate system states that Q is a 3D point with coordinates $[X, Y, Z, 1]$. Meanwhile, s is a scale illustrating the pixel scaling concerning focal length shift to get the intrinsic attributes. As a result, Eq. (1) may be simplified to Eq. (2) and Eq. (3).

$$u = \frac{1}{s_x} f \frac{X}{Y} + u_0 \tag{2}$$

$$v = \frac{1}{s_y} f \frac{Y}{Z} + v_0 \tag{3}$$

For Eq. (2) and Eq. (3), the values of X and Y may be derived using Eq. (4) and Eq. (5) if the scale parameter is considered to be 1. As a result, the Z parameter's value can be calculated by using Eq. (6).

$$X = \frac{(u - u_0) \times Z}{f_x} \tag{4}$$

$$Y = \frac{(v - v_0) \times Z}{f_y} \tag{5}$$

$$Z = depth\ (v, u) \tag{6}$$

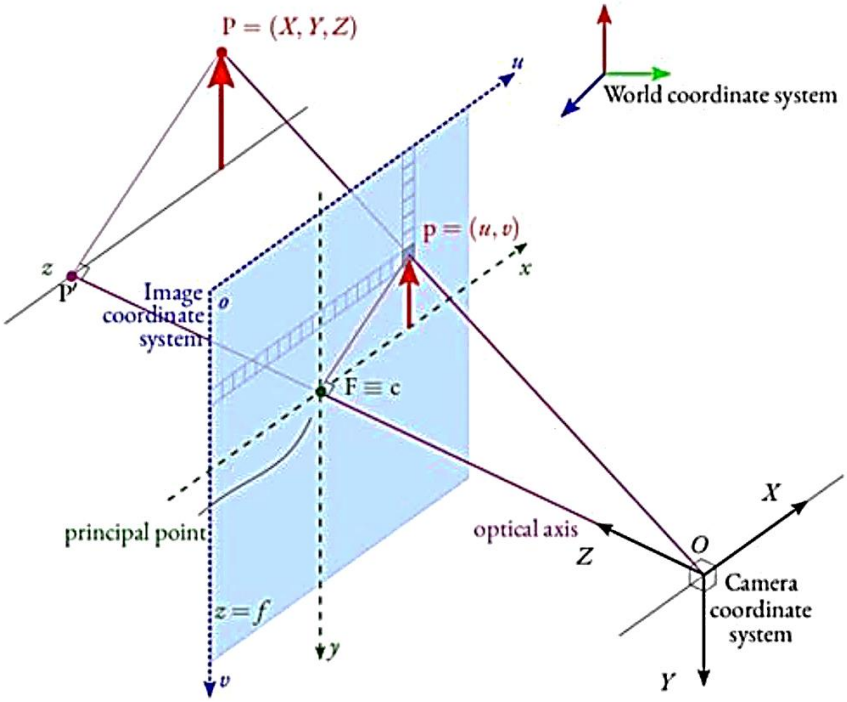

Fig. 4. The model for pinhole camera [27]

3. Results and Discussion

This section reports the result of the proposed method's performance. The KITTI dataset is used to validate the performance of the proposed distance estimation algorithm using four differences in Deep ANN architecture. The first Deep ANN architecture has 1 Hidden Layer, while the other Deep ANN architectures have 2 Hidden Layers, 3 Hidden Layers and 4 Hidden Layers, respectively. The value of Hidden Layers of 1 to 4 is used to prevent the overtraining condition. When the number of hidden layers is excessively high relative to the task's difficulty, a negative impact on the network's time complexity often happens when the network's performance is closely matched to the test data. It is because the input layer has only four input variables.

Meanwhile, to prevent undertraining conditions, every Hidden Layer consists of 8 neurons. The best distance estimation algorithm is validated with the number of Hidden Layers in Deep ANN architecture. It is compared in terms of prediction distance in distance estimation and the value of validation metrics.

3.1 Evaluation of Distance Estimation

The prediction distance with all four proposed distance estimation algorithms is trained and tested with the KITTI dataset. After those algorithms are trained, the prediction distance is tested. The prediction visualiser is displayed as shown in Figure 5. The red lettering indicates the actual and prediction distance, while the green box is for vehicle detection.

Fig. 5. The prediction distance for distance estimation using DEEP-ANN algorithms

Table 4 shows a sampling of the distance estimation algorithms' outcomes. The results demonstrated that the proposed distance estimation method with four hidden layers of Deep ANN has the lowest measured error. The measurement error decreases as the number of Hidden Layers in the Deep ANN architecture increases. However, in Deep Ann architecture with 4 Hidden Layers, the measured error increases as the distance increases. Due to the focus of this study on emergency AEB and AES assessment, the accuracy of distance estimation for the nearest distance is adequate.

Table 4
The distance estimation results

Actual Distance (m)	1 Hidden Layer		2 Hidden Layers		3 Hidden Layers		4 Hidden Layers	
	Distance Estimation (m)	Measured Error (%)	Distance Estimation (m)	Measured Error (%)	Distance Estimation (m)	Measured Error (%)	Distance Estimation (m)	Measured Error (%)
8	4.4960	43.8003	4.9989	37.5134	8.3322	4.1520	8.0331	0.4143
18	27.1379	50.7663	19.4321	7.9560	16.4499	8.6117	17.9334	0.3702
28	34.9400	24.7859	32.5001	16.0718	29.8253	6.5190	27.8579	0.5074
38	34.9521	8.0208	39.0573	2.7823	39.5712	4.1348	38.7283	1.9165
48	39.6806	17.3320	43.2663	9.8620	46.1205	3.9155	47.5800	0.8751
58	45.8229	20.9951	52.6804	9.1717	57.2699	1.2589	58.1776	0.3062
68	45.6679	32.8413	53.6581	21.0910	60.1889	11.4870	63.1792	7.0894
78	45.6276	41.5031	53.7096	31.1415	60.5548	22.3657	63.4979	18.5925

3.2 Evaluation of Validation Metrics

Many regression models rely on distance metrics to determine the convergence to the best result. Even the definition of a "best" result needs to be explained quantitatively by some metric. Usually, the metrics used are the Mean Average Error (MAE), the Mean Squared Error (MSE) or the Root Mean Squared Error (RMSE).

The proposed algorithms are validated with compared the MAE value, MSE value and RMSE value. Table 5 shows that the value of MAE for the Deep ANN architecture with 4 Hidden Layers is the lowest compared with 1 Hidden Layer, 2 Hidden Layers and 3 Hidden Layers by 56.56%, 22.41% and 12.04%, respectively. For the value of MSE, the Deep ANN architecture with 4 Hidden Layers also has the lowest value compared with 1 Hidden Layer, 2 Hidden Layers and 3 Hidden Layers by 79.51%, 47.52% and 20.60%, respectively. Furthermore, lastly, the Deep ANN architecture with 4 Hidden Layers for the RMSE value also has the lowest value compared with 1 Hidden Layer, 2 Hidden Layers and 3 Hidden Layers by 54.73%, 27.56% and 10.89%, respectively.

Table 5
Validation metrics

Validation Metrics	1 Hidden Layer	2 Hidden Layers	3 Hidden Layers	4 Hidden Layers
MAE	5.8896	3.2976	2.9089	2.5587
MSE	61.3991	23.9724	15.8441	12.5806
RMSE	7.8357	4.8962	3.9805	3.5469

Figure 6 shows the graph of the relationship between the Hidden Layers and the validation metrics. In conclusion, as the Hidden Layers increase, the value of validation metrics decreases. Thus, the Deep ANN architecture with 4 Hidden Layers is better compared to fewer Hidden Layers in prediction error and validation metrics.

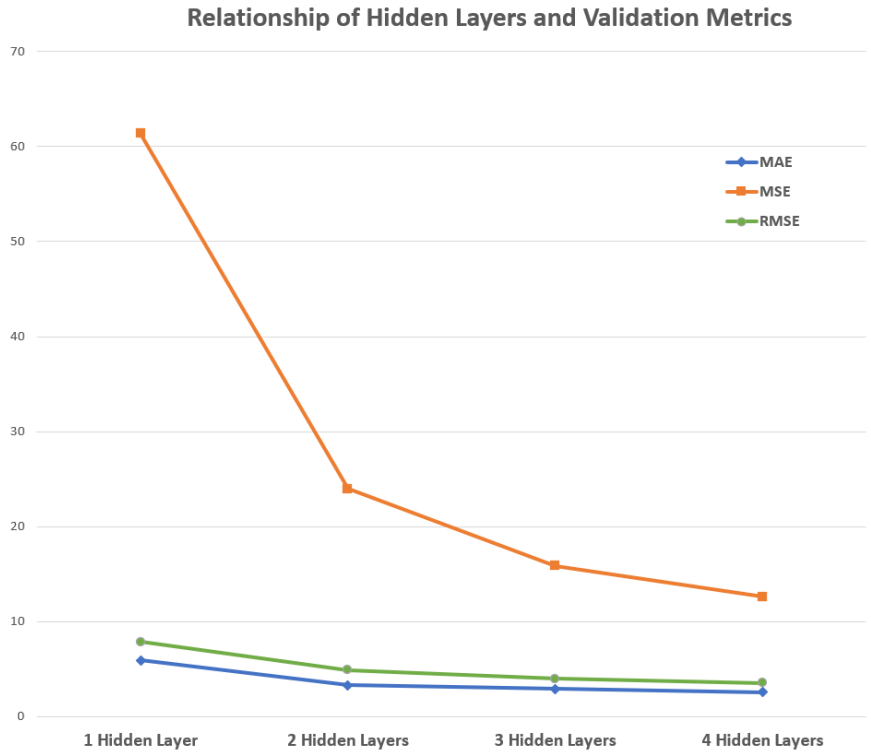

Fig. 6. The relationship between Hidden Layers and Validation Metrics

The Loss History graph for all algorithms is also generated, as shown in Figure 7. The Loss History graph for all Deep ANN architectures is a good fit learning curve. A training and testing loss identifies a good fit learning curve that decreases to the point of stability with a minimal gap between the two final loss values [28]. However, this study focuses on the prediction value's accuracy. Thus, the Deep ANN with 4 Hidden Layers is better than the Deep ANN architecture with fewer Hidden Layers.

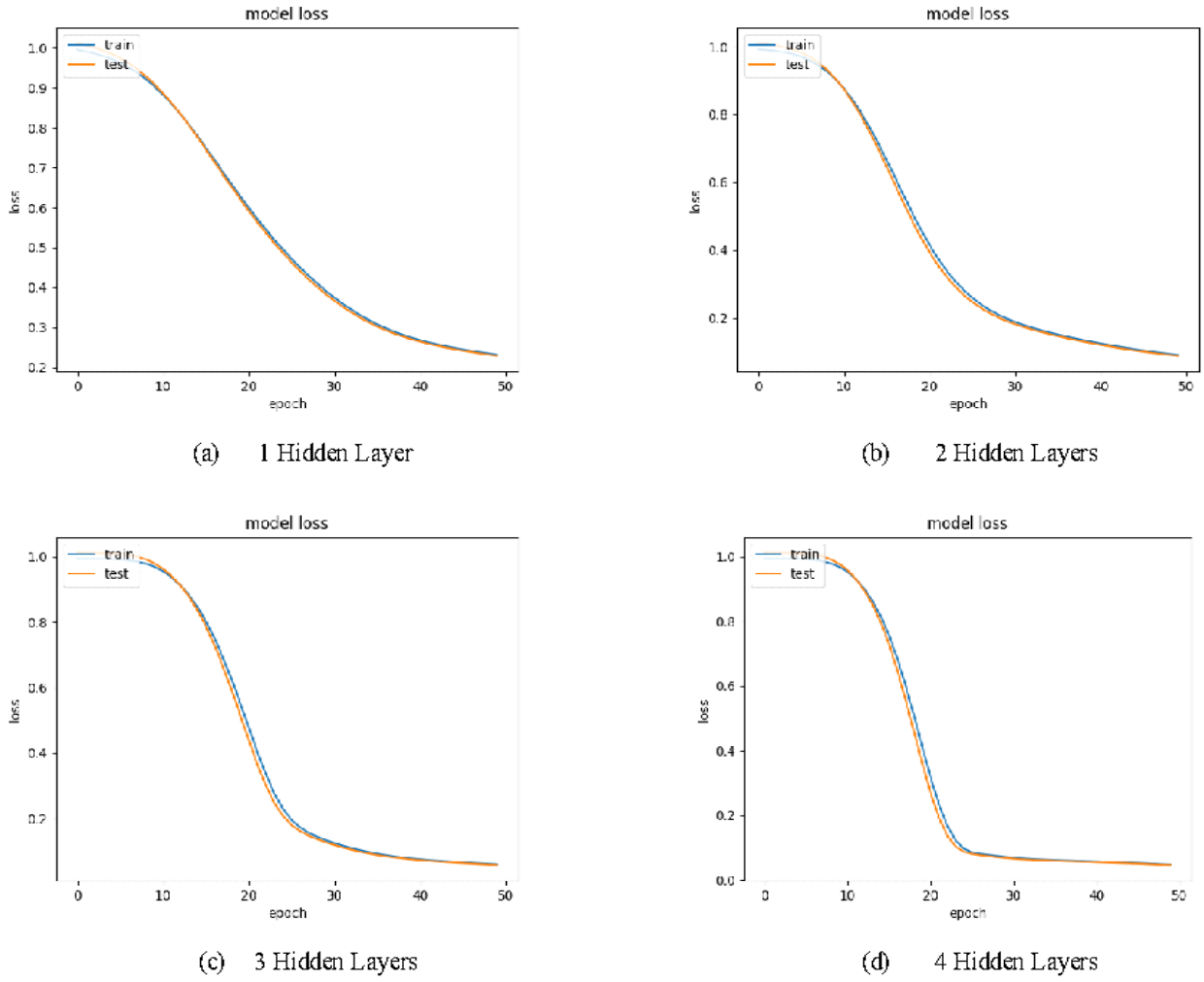

Fig. 7. The loss history graph

The primary goals of this study are to determine the number of hidden layers present in the network and the impact of those hidden layers on the networks. The results demonstrated that the proposed distance estimation method with four hidden layers of Deep ANN has the lowest measured error. The measurement error decreases as the number of Hidden Layers in the Deep ANN architecture increases. However, in Deep Ann architecture with 4 Hidden Layers, the measured error increases as the distance increases. Due to the focus of this study on the AEB and AES assessment, the accuracy of distance estimation for the nearest distance is adequate. In addition, as the Hidden Layers increase, the value of validation metrics such as MAE, MSE and RMSE decreases. Thus, the Deep ANN architecture with 4 Hidden Layers outperforms fewer Hidden Layers in prediction error and validation metrics.

5. Conclusions

This study proposed a test methodology using deep learning to assess the assessments for AES and AEB in the safety rating test protocol. The suggested method for test methodology in situation-based threat assessments is based on a monocular distance estimation-based approach. The monocular visual distance estimate has benefits such as speed, simplicity, and affordability compared to binocular vision. Deep ANN algorithms use as the deep learning approach for the proposed

method. The KITTI dataset is employed to train and test the proposed method. Four different architectures with varying numbers of hidden layers are designed to select the optimum architecture for the proposed method. The results show that the Deep ANN architecture with 4 Hidden Layers outperforms fewer Hidden Layers in prediction error and validation metrics. Therefore, this study suggested applying the proposed method to assess the assessments for AES and AEB in the safety rating test protocol. Besides that, the proposed method will evaluate collision-avoidance scenarios with essential metrics like ideal braking distances to verify the methodology for active safety features. As a result, this study can help provide the methodological framework for a deeper comprehension of driver behaviour with the long-term purpose of reducing rear-end collision rates.

Acknowledgement
This research was funded by the Ministry of Education (MOE) through Fundamental Research Grant Scheme (FRGS/1/2021/TK0/UTM/02/42) and partially supported by the ASEAN NCAP Collaborative Hollistic Research (ANCHOR) Grant No. A3-21.

References

[1] Ministry of International Trade and Industry. "National Automotive Policy 2020." (2020)
[2] ASEAN NCAP "Assessment Protocol Safety Assist." (2019)
[3] Yang, Lan, Yipeng Yang, Guoyuan Wu, Xiangmo Zhao, Shan Fang, Xishun Liao, Runmin Wang, and Mengxiao Zhang. "A systematic review of autonomous emergency braking system: impact factor, technology, and performance evaluation." *Journal of advanced transportation* 2022 (2022). https://doi.org/10.1155/2022/1188089
[4] Favarò, Francesca, Sky Eurich, and Nazanin Nader. "Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations." *Accident Analysis & Prevention* 110 (2018): 136-148. https://doi.org/10.1016/j.aap.2017.11.001
[5] Jeong, Eunbi, and Cheol Oh. "Evaluating the effectiveness of active vehicle safety systems." *Accident Analysis & Prevention* 100 (2017): 85-96. https://doi.org/10.1016/j.aap.2017.01.015
[6] Liu, Runqiao, Minxiang Wei, and Wanzhong Zhao. "Trajectory tracking control of four wheel steering under high speed emergency obstacle avoidance." *International Journal of Vehicle Design* 77, no. 1-2 (2018): 1-21. https://doi.org/10.1504/IJVD.2018.098265
[7] Nilsson, Jonas. *Computational verification methods for automotive safety systems*. Chalmers Tekniska Hogskola (Sweden), 2014.
[8] Hamid, Umar Zakir Abdul, Mohd Hatta Mohammed Ariff, Hairi Zamzuri, Yuichi Saito, Muhammad Aizzat Zakaria, Mohd Azizi Abdul Rahman, and Pongsathorn Raksincharoensak. "Piecewise trajectory replanner for highway collision avoidance systems with safe-distance based threat assessment strategy and nonlinear model predictive control." *Journal of Intelligent & Robotic Systems* 90 (2018): 363-385. https://doi.org/10.1007/s10846-017-0665-8
[9] Sun, Xiyan, Yingzhou Jiang, Yuanfa Ji, Wentao Fu, Suqing Yan, Qidong Chen, Baoguo Yu, and Xingli Gan. "Distance measurement system based on binocular stereo vision." In *IOP Conference Series: Earth and Environmental Science*, vol. 252, no. 5, p. 052051. IOP Publishing, 2019. https://doi.org/10.1088/1755-1315/252/5/052051
[10] Huang, Le, Gongping Wu, Jiayang Liu, Song Yang, Qi Cao, Wa Ding, and Wenjie Tang. "Obstacle distance measurement based on binocular vision for high-voltage transmission lines using a cable inspection robot." *Science Progress* 103, no. 3 (2020): 0036850420936910. https://doi.org/10.1177/0036850420936910
[11] Zhang, Jiaxu, Shaolin Hu, and Haoqiang Shi. "Deep learning based object distance measurement method for binocular stereo vision blind area." *International Journal of Advanced Computer Science and Applications* 9, no. 9 (2018). https://doi.org/10.14569/IJACSA.2018.090977
[12] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 28 (2015).
[13] Xie, Fangyuan, Pengshuai Yin, Zehui Ke, Ruizhou Sun, Ke Ding, Jian Chen, and Qingyao Wu. "Vehicle Segmentation with Coarse Distance Estimation Based on Monocular Vision." In *2021 IEEE International Conference on e-Business Engineering (ICEBE)*, pp. 16-20. IEEE, 2021. https://doi.org/10.1109/ICEBE52470.2021.00019
[14] Szeliski, Richard, and Sing Bing Kang. "Shape ambiguities in structure from motion." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, no. 5 (1997): 506-512. https://doi.org/10.1109/34.589211
[15] Mancini, Francesco, Marco Dubbini, Mario Gattelli, Francesco Stecchi, Stefano Fabbri, and Giovanni Gabbianelli. "Using unmanned aerial vehicles (UAV) for high-resolution reconstruction of topography: The structure from

motion approach on coastal environments." *Remote sensing* 5, no. 12 (2013): 6880-6898. https://doi.org/10.3390/rs5126880

[16] Ramírez-Hernández, Luis R., Julio C. Rodríguez-Quinoñez, Moises J. Castro-Toscano, Daniel Hernández-Balbuena, Wendy Flores-Fuentes, Raúl Rascón-Carmona, Lars Lindner, and Oleg Sergiyenko. "Improve three-dimensional point localization accuracy in stereo vision systems using a novel camera calibration method." *International Journal of Advanced Robotic Systems* 17, no. 1 (2020): 1729881419896717. https://doi.org/10.1177/1729881419896717

[17] Wang, Chaoyang, José Miguel Buenaposada, Rui Zhu, and Simon Lucey. "Learning depth from monocular videos using direct methods." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2022-2030. 2018. https://doi.org/10.1109/CVPR.2018.00216

[18] Li, Ruihao, Sen Wang, Zhiqiang Long, and Dongbing Gu. "Undeepvo: Monocular visual odometry through unsupervised deep learning." In *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 7286-7291. IEEE, 2018. https://doi.org/10.1109/ICRA.2018.8461251

[19] Bian, Jiawang, Zhichao Li, Naiyan Wang, Huangying Zhan, Chunhua Shen, Ming-Ming Cheng, and Ian Reid. "Unsupervised scale-consistent depth and ego-motion learning from monocular video." *Advances in neural information processing systems* 32 (2019).

[20] Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite." In *2012 IEEE conference on computer vision and pattern recognition*, pp. 3354-3361. IEEE, 2012. https://doi.org/10.1109/CVPR.2012.6248074

[21] Panchal, Foram S., and Mahesh Panchal. "Review on methods of selecting number of hidden nodes in artificial neural network." *International Journal of Computer Science and Mobile Computing* 3, no. 11 (2014): 455-464.

[22] Gupta, Tarun Kumar, and Khalid Raza. "Optimizing deep feedforward neural network architecture: A tabu search based approach." *Neural Processing Letters* 51 (2020): 2855-2870. https://doi.org/10.1007/s11063-020-10234-7

[23] Zhang, Quan-shi, and Song-Chun Zhu. "Visual interpretability for deep learning: a survey." *Frontiers of Information Technology & Electronic Engineering* 19, no. 1 (2018): 27-39. https://doi.org/10.1631/FITEE.1700808

[24] Narasimhan, K. Adith. "Neural Networks." *Medium* (2021).

[25] Shamoon, Siddiqui. "How would we find a better activation function than ReLU?". *Medium* (2019).

[26] Saleh, Hadi, Shadi Saleh, Nathan Teyou Toure, and Wolfram Hardt. "Robust Collision Warning System based on Multi Objects Distance Estimation." In *2021 IEEE Concurrent Processes Architectures and Embedded Systems Virtual Conference (COPA)*, pp. 1-6. IEEE, 2021. https://doi.org/10.1109/COPA51043.2021.9541452

[27] Arabi, Saeed, Anuj Sharma, Michelle Reyes, Cara Hamann, and Corinne Peek-Asa. "Farm vehicle following distance estimation using deep learning and monocular camera images." *Sensors* 22, no. 7 (2022): 2736. https://doi.org/10.3390/s22072736

[28] Anzanello, Michel Jose, and Flavio Sanson Fogliatto. "Learning curve models and applications: Literature review and research directions." *International Journal of Industrial Ergonomics* 41, no. 5 (2011): 573-583. https://doi.org/10.1016/j.ergon.2011.05.001