INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 17 No. 1 (2025) 416-437
https://publisher.uthm.edu.my/ojs/index.php/ijie

Distance Estimation Using Deep Learning Approaches for Rear-end Collision Avoidance Alerts

Siti Nur Atiqah Halimi[1], Mohd Azizi Abdul Rahman[1]*, Mohamad Danish Hakim A Rashid[1], Mohd Hatta Mohammed Ariff[1], Nurulakmar Abu Husain[1], Hairi Zamzuri[2], Zulhaidi Mohd Jawi[3]

[1] *Malaysia Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur 54100, MALAYSIA*
[2] *Emoovit Technology Sdn Bhd, Futurise Center, Persiaran Apec, Cyberjaya 63000, MALAYSIA*
[3] *ASEAN NCAP Operational Unit, Malaysian Institute of Road Safety Research, Kajang 43000, MALAYSIA*

*Corresponding Author: azizi.kl@utm.my
DOI: https://doi.org/10.30880/ijie.2025.17.01.034

Article Info

Received: 13 September 2024
Accepted: 27 November 2024
Available online: 30 April 2025

Keywords

Artificial intelligence, deep learning, distance estimation, rear-end collision, autonomous emergency braking, autonomous emergency steering

Abstract

Autonomous Emergency Braking (AEB) and Autonomous Emergency Steering (AES) are part of the advanced driver assistance system (ADAS) equipped in intelligent vehicles. AEB is a system that warns drivers of potential collisions and assists them in utilizing the vehicle's maximum capabilities. AES is an active safety system that aids in evasive steering. If it detects a potential collision, unlike AEB, the AES system will autonomously adjust the steering to prevent it. The challenges for AEB and AES include determining how much space is required to avoid an accident while turning or braking and how much distance is required to avoid an impact when braking and turning simultaneously. Considering such inquiries, it is necessary to devise a system to estimate the distance between the vehicles. Therefore, this paper proposes a Monocular Vision Distance Estimation (MVDE) method employing deep learning techniques for accurately calculating the distance between vehicles, particularly for use in AEB and AES systems. The MVDE technique uses monocular vision, emphasizing object detection and distance estimation. In contrast to complex depth estimation techniques, the proposed method employs a Single Shot Detector (SSD) with MobileNet architecture for object recognition and Deep Artificial Neural Networks (Deep ANN) for accurate distance estimation. Using a real-world dataset collected in Cyberjaya, Malaysia, this study rigorously assesses the performance of this method. Results indicate that the MVDE method with four hidden layers in Deep ANN outperforms earlier techniques, with a maximum measured error of 4m to actual distances. In addition, it is competitive with RADAR-based systems and offers a cost-effective alternative for widespread adoption. These findings support the potential of MVDE for augmenting vehicle safety, shaping future automotive standards, and facilitating the widespread implementation of AEB and AES systems.

1. Introduction

The advanced driver assistance system, often known as ADAS, is an integral component of the whole system found in intelligent navigation vehicles [1]. When combined, these systems help create a more secure environment for drivers and passengers. A Lane Detection, Lane-Keeping System, Lane Departure Warning System, Blind Spot Assist, Autonomous Emergency Braking (AEB), and Autonomous Emergency Steering (AES) are some ADAS modules that may be installed in a vehicle [2]. Table 1 captures key characteristics and common issues related to the mentioned ADAS modules. This paper focuses on AEB and AES.

Table 1 *The characteristics and drawbacks of ADAS modules*

Technology	Characteristics	Drawbacks
Lane Detection [3]	Detects lane markings on the road using cameras and sensors, and alerts drivers if they are drifting out of their lane.	Can struggle in poor weather conditions or when lane markings are faded or unclear.
Lane-Keeping System [4]	Actively helps the vehicle stay within its lane by providing steering inputs or adjusting brakes when the car starts to drift.	Over-reliance may result in inattentiveness and can disengage unexpectedly if lane markings are not clear.
Lane Departure Warning [5]	Warns the driver when the vehicle unintentionally crosses lane markings without using the turn signal.	False alarms in cases of lane changes without signalling; may not function effectively in areas with poor road markings or on curves.
Blind Spot Assist [6]	Uses radar or sensors to detect vehicles in the driver's blind spots and provides visual or audible alerts to avoid unsafe lane changes.	May fail to detect fast-moving vehicles or motorcycles; driver reliance may reduce proper mirror use.
Autonomous Emergency Braking (AEB) [7]	Automatically applies the brakes if it detects an imminent collision with another vehicle, pedestrian, or obstacle.	False positives may result in unnecessary or abrupt braking; systems may not react quickly enough in certain situations like high-speed or complex traffic scenarios.
Autonomous Emergency Steering (AES) [8]	Steers the vehicle automatically to avoid collisions, often used in conjunction with AEB for more precise evasive action.	It has a strong reliance on sensor accuracy, may not always determine the safest route for evasive actions, and may malfunction in unpredictable or complicated traffic situations.

The AEB [9] system is a cutting-edge safety feature that uses sensors to scan the road in front of it and spot potential collision threats. AEB systems warn the driver when an accident is imminent and, if necessary, automatically apply the brakes to either halt the collision or mitigate its damage. The AEB system [10] can significantly lower the likelihood of accidents by analysing the speed, distance, and trajectory of objects on the route to increase vehicle safety. AEB has certain limitations, though, and its effectiveness depends on various factors.

Meanwhile, the AES system [11] is one of the active safety systems that can assist with evasive steering. AES supplements AEB systems by analysing sensor data to detect possible collision hazards. The driver is given a warning, and if it becomes required, the AES will automatically provide torque to the steering system to aid the driver in steering away from the obstruction and avoid colliding with it. It aims to improve collision avoidance by actively assisting the driver's steering inputs [12]. It is essential to remember that AES, like AEB, has restrictions that must be considered. The efficiency of AES can be influenced by several variables, including sensor performance, object recognition and tracking precision, ambient circumstances, and the system's capacity to give timely and suitable steering assistance [13]. AES could be restricted at more incredible speeds or in specific driving situations, such as inclement weather or slippery roads.

The challenges for AEB and AES include determining how much space is required to avoid an accident while turning or braking and how much distance is required to avoid an impact when braking and turning simultaneously. In light of such inquiries, it is necessary to devise a system to estimate the distance between the vehicles. This study centres on developing a Monocular Vision Distance Estimation (MVDE) technique utilizing a deep learning algorithm. The primary objective is to accurately estimate the distance between vehicles, specifically for the functionality of AES and AEB systems, all facilitated through a single camera. The rationale behind utilizing a camera as the sole sensor is to streamline the process by incorporating deep learning directly within the camera, thus minimizing the computational demands. The proposed algorithm is intended to precisely detect vehicles on the road and calculate the distance between the front vehicle and the camera while simultaneously detecting vehicles. This approach aims to take advantage of the capabilities of deep learning to accomplish these two objectives.

2. Distance Estimation

2.1 Traditional Methods

Before the application of deep learning in distance estimation techniques, one of the approaches for estimating the distance between objects using a geometrical approach was the traditional method. The factors required to determine the distance are the height of the camera relative to the ground and the camera's angle [14]. Most traditional distance estimating techniques depend on scene observations in space or time [15]. Active and passive approaches may be used to classify traditional techniques. Active methods compute the scene's depth by interacting with the objects and surroundings. Other active methods, such as light-based distance estimation, use light illumination to determine the distance to various objects. Other active approaches include ultrasound. These techniques measure the time it takes for an emitted pulse to reach an image sensor based on the known speed of the wave. Passive techniques use the optical properties of collected pictures. By using computational image processing, these techniques extract depth information. There are two basic techniques under the category of passive methods: multi-view depth estimation, such as stereo depth estimation, and monocular depth estimation.

Traditional depth estimation techniques concentrate primarily on multi-view geometry. This study does not provide a comprehensive assessment of these techniques. However, it is essential to note that multi-view classical approaches have several drawbacks, including computational complexity and accompanying high energy demands. Current research utilizes deep-learning techniques to generate more precise findings with fewer computational and energy requirements [16]. Deep learning-based methodologies and the availability of large-scale datasets have drastically altered monocular depth estimation techniques.

2.2 Deep Learning Methods

There are two known approaches employing deep learning to estimate distance. The first one uses a depth map as ground truth and is known as depth estimation [17-20], while the second is known as object distance estimation [21-24], which estimates an item's distance by determining its size.

2.2.1 Depth Estimation

Deep learning-based depth estimation algorithms, on the other hand, evaluate the depth of whole scenes while requiring exact distance measurement between vehicles. Depth estimation approaches include MonoDepth [18, 19], DenseDepth [18], and DepthNet [17].

Using just one image, Masoumian et al. [17] proposed a deep learning framework comprising two separate Deep Neural Networks for depth estimation and object detection. Using the You Only Look Once (YOLOv5) network, the objects in the scene can be detected and localized. In parallel, an autoencoder network determines the relative distances, allowing the estimated depth image to be produced. The suggested framework has much potential, with an accuracy of 96% and an RMSE of 0.203.

Adz-Dzikri et al. [18] suggested a system that detects objects using deep learning based on depth estimation, as shown in Fig. 1. The pre-trained models of Monodepth and DenseDepth were utilized to test and validate the accuracy of the depth prediction approach employed for distance estimation. To successfully perform object detection, both Single Shot Detector Localization (SSD) and Deep Learning were used. The measuring reel tape meters were used as the definitive source of information. The authors concluded that depth prediction has certain flaws. Specifically, it does not function well when there is a considerable shadow, or the lighting conditions nearby are inconsistent.

Ma et al. [31] suggested a novel framework that integrates object identification and distance estimates to recognize objects from photos taken with the mono-camera. They retrain both YOLOv3 and BTS to assess how well the suggested framework performs against the standard set by the KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) [25]. On the KITTI benchmark, they conducted testing. According to the findings, mono-cameras may sometimes replace LIDAR as the principal sensor.

However, these methods pose risks associated with generalization, which can become problematic when the driving scenarios observed in video recordings during training and testing differ. Moreover, deep learning-based depth estimation techniques are computationally costly, and employing them in conjunction with object detection Deep Neural Networks in real-time applications such as autonomous driving may not be accessible.

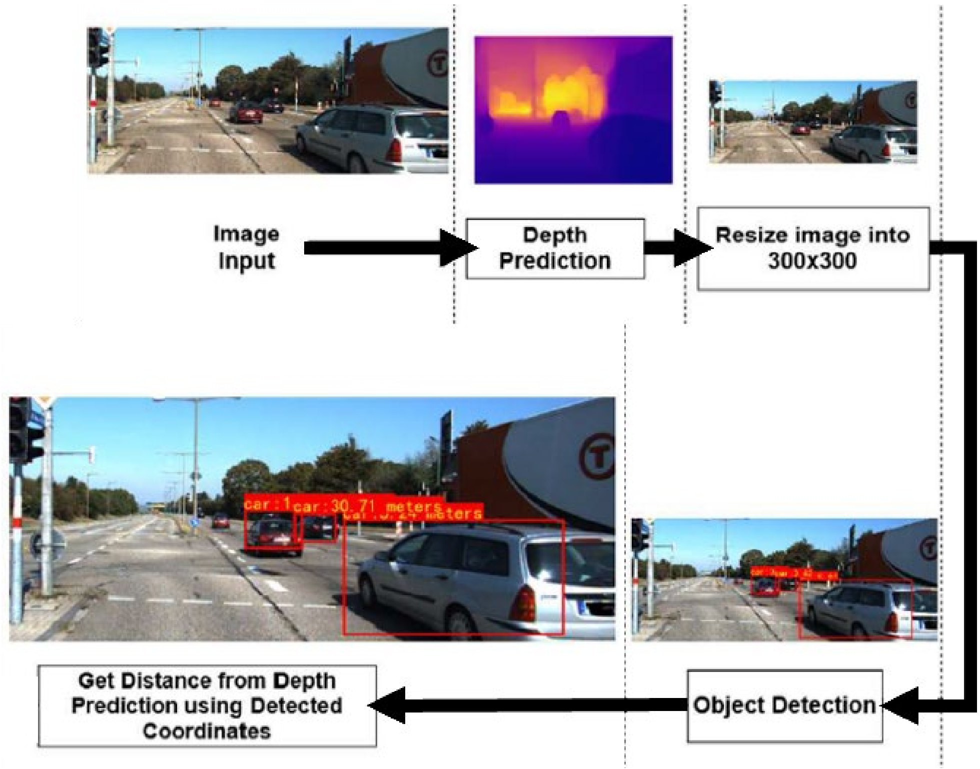

Fig. 1 *Framework of depth prediction and object detection by Adz-Dzikri et al. [18]*

2.2.2 Object Distance Estimation

Another method based on vision is employed for distance measurement, which involves extracting input features from the bounding box of the detected object captured by the camera [22]. For example, DistNet [23] represents a recent effort to develop a network dedicated to distance estimation. Instead of learning image features for distance estimation, the authors opted for a CNN-based model (YOLO) to predict bounding boxes.

To simultaneously identify, classify, and range on-road obstacles on a monocular picture taken by a low-cost image sensor, Parmar et al. [21] introduced a single-integrated deep-learning architecture called DeepRange. The average error in range estimation on the KITTI test dataset is less than one meter. The authors argue that future work should consider datasets of on-road situations in various lighting and weather circumstances to create a highly reliable object identification and range system.

An Artificial Neural Network (ANN) was suggested by Karthika et al. [22] to create an efficient and straightforward distance calculation technique for preceding cars using camera pictures. The proposed framework consists of YOLO and ANN algorithms, as shown in Fig. 2. The suggested method is tested and trained on the KITTI dataset. Compared to Bao's method [26], experiments show that the method has enhanced the system's resilience and decreased the error rate over long distances with a superior R2 score of 0.992. This study assessed the suggested approach on roadways with linear structures. Analysis of the ramp's curving surface and rough road conditions is still pending.

By evaluating the road gradients of both the ego car and the target vehicle and using a 2D object distance detection DeepNet, Back et al. [19] suggested an inter-vehicle distance estimate framework that can consider the slope variations of a road advance. Compared to deep learning-based depth estimation algorithms DORN and Monodepth2, the authors' numerical studies show that the suggested approach significantly reduces time complexity and distance estimate error.

Vajgl et al. [23] have shown that while there have been attempts to alter YOLOv3 to include monocular absolute distance measurement, the architecture of YOLO has not yet been fully merged with the distance estimation capability. The researchers confirmed that distance estimation enhanced bounding box estimate and improved bounding box detection accuracy compared to YOLOv3 standards. The KITTI dataset was used as a benchmark, and the mean absolute and relative distance errors were 2.5 meters and 11%, respectively. These results support the claim that solutions that fully integrate distance estimation into YOLO's functionality can achieve a significantly better result than those that build distance estimation on top of the model's output.

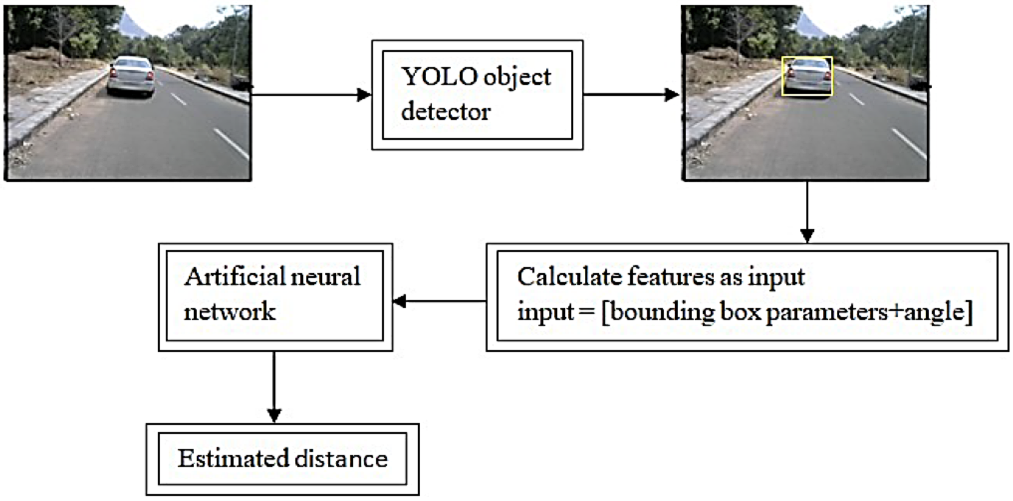

Fig. 2 *Framework of YOLO and ANN used in distance estimation by Karthika et al. [22]*

A deep-learning-based strategy was suggested by Arabi et al. [24] in 2022 as a way to extract the depth information of objects in an image captured by a low-resolution monocular camera. Using DeepStream, vehicle identification and tracking were performed, and bounding box data were collected. In this study, the linear regression model, the pinhole model, and the ANN were all direct distance estimation techniques that were considered and compared. The height of identified objects in a picture was calculated using the pinhole camera model and connected to the object's actual distance from the camera. The calibration stage of the distance estimation procedure entailed comparing the distances calculated from the pinhole model with those of the RTK GPS as the ground truth to determine the camera's focal length. The ANN model, which had the lowest mean error and standard deviation among the three models stated, was the best distance estimator.

The ANN [22, 24] is the most complex and has the biggest variation in object distance estimation because it creates a nonlinear mapping between each detection parameter and estimates the output. The most difficult component of ANN is determining a suitable configuration and architecture.

Object distance estimation employs more straightforward algorithms and is less computationally intensive than depth estimation. Nevertheless, both techniques share a common drawback - their precision is compromised under specific weather conditions, such as intense rain, dense fog, or heavy snowfall. These scenarios of diminished visibility can significantly impact the accuracy of distance estimation.

A comparison table between conventional distance estimation, deep learning-based distance estimation, and deep learning-based depth estimation, along with the advantages and drawbacks of each method is shown in Table 2. This table outlines the trade-offs between traditional and deep learning-based methods for distance estimation. While deep learning techniques offer more accuracy and efficiency in some cases, they come with computational and visibility-related challenges.

3. Monocular Vision Distance Estimation (MVDE) Method

This study proposes a novel approach based on deep learning and monocular vision for vehicle distance estimation. We chose a vision sensor due to its capacity for complex processing when combined with advanced software [27]. This choice enables us to accurately calculate the distance between vehicles, a critical aspect of vehicle safety systems such as AEB and AES.

Our approach holds significant promise for enhancing road safety while maintaining cost-effectiveness. By leveraging the power of deep learning and monocular vision, we aim to provide a robust and accessible solution for vehicle distance estimation. This research aligns with the broader goal of improving automotive safety standards and making advanced safety features more widely available to the public. Utilizing a vision sensor offers the unique advantage of object segmentation and targeted object detection. This capability is particularly advantageous for real-time applications where precision is crucial. In contrast, deep learning-based depth estimation methods can pose challenges when integrated with real-time object recognition due to their high processing costs.

Hence, our study proposes object distance estimation as an alternative method. As depicted in Fig. 3, the framework centres on Deep ANN for distance estimation. Concurrently, the SSD with MobileNet architecture is employed for object detection within the MVDE method. This combination ensures efficient and accurate distance estimation while enhancing object recognition in real-world scenarios, a critical aspect of ADAS and safety protocols.

Table 2 *Comparison between distance estimation methods*

Method	Core Characteristics	Advantages	Drawbacks
Conventional Distance Estimation	Utilizes geometric approaches (height, angle, etc.), active methods (light, ultrasound), or passive techniques (multi-view, stereo, monocular).	• Proven and simple algorithms • Does not require large datasets • Effective in controlled environments	• High computational complexity for multi-view methods • Susceptible to errors in poor lighting or weather conditions • High energy demands for complex scenarios
Deep Learning-Based Distance Estimation	Uses neural networks (e.g., YOLO, DistNet) to detect objects and estimate distances by analyzing bounding boxes or object sizes in the image.	• Less computationally intensive than depth estimation • Accurate for object detection and distance prediction • Suitable for real-time applications	• Limited generalization across diverse scenarios • Prone to inaccuracies in low visibility (rain, fog, etc.) • Requires large labelled datasets for training
Deep Learning-Based Depth Estimation	Estimates the depth of entire scenes using deep learning models like MonoDepth, DenseDepth, and DepthNet, often in conjunction with object detection (e.g., YOLO).	• High precision in distance and depth estimation • Can work with monocular images (single camera) • Can replace expensive sensors like LiDAR in some cases	• Computationally expensive • Requires significant GPU resources for real-time applications • Errors under varying lighting conditions (shadows, inconsistent lighting)

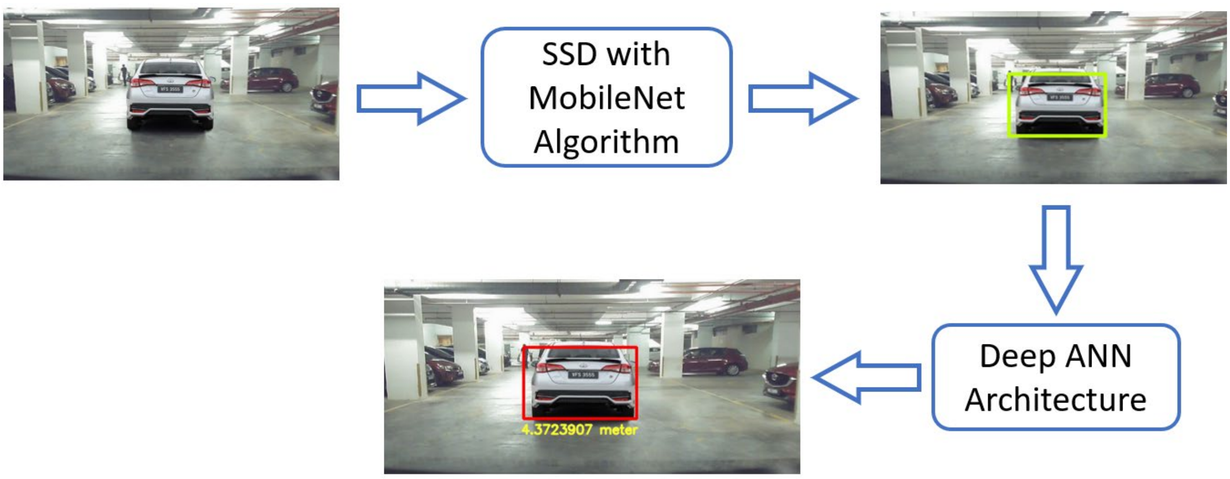

Fig. 3 *Framework of MVDE method*

3.1 SSD with MobileNet Architecture

This study adopts the SSD method with MobileNet architecture as the chosen Deep Learning architecture for object detection within the MVDE method. This selection is driven by the objective of estimating the distance between vehicles, particularly for implementation in AES and AEB systems. The SSD model [28], initially proposed by Liu et al. [29], takes inspiration from two key sources: the anchor box concept from the Faster R-CNN detection model and the regression concepts found in the YOLO algorithm. As depicted in Fig. 4, the SSD model effectively utilizes both bottom and high-level feature maps for object detection, contributing to its capability for multi-scale object detection. This approach aligns with the requirements of the MVDE method and its focus on accurately detecting and estimating distances between vehicles in real-world scenarios.

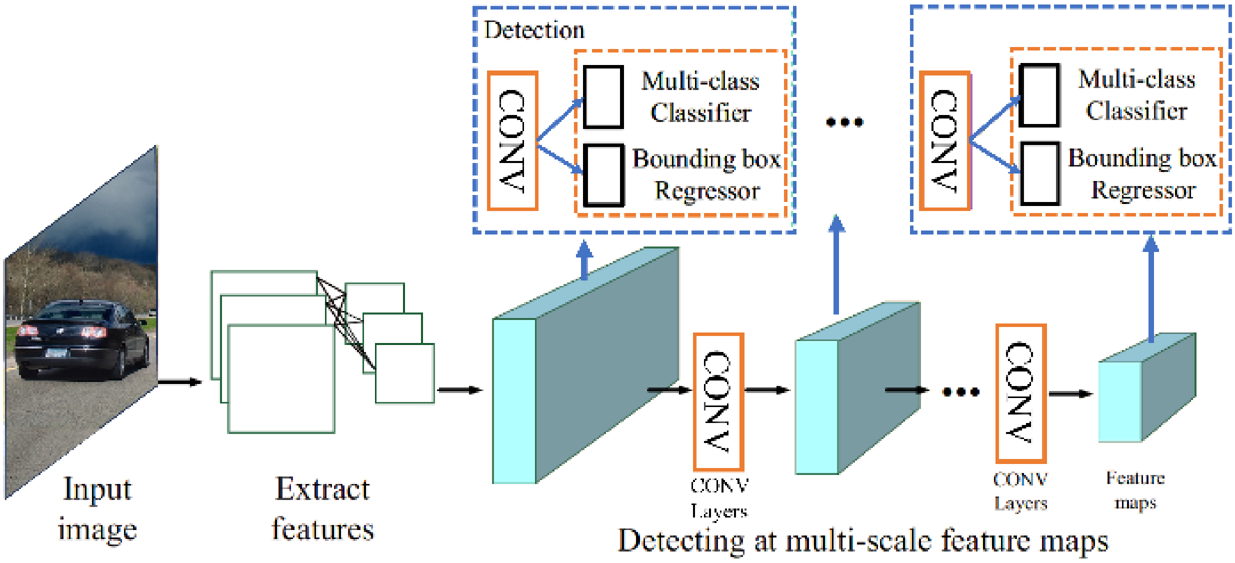

Fig. 4 *SSD architecture*

Furthermore, this study used the MobileNet architecture of Howard et al. [30] as extract features. MobileNet aims to create a lightweight deep CNN that shrinks the model's size and speeds up computation. It is built on the foundation of depthwise separable convolutions. The core of network design's separable convolution can effectively reduce the number of parameters and computation at the expense of lesser performance [31]. The diagram in Fig. 5 illustrates the MobileNet architecture. Consequently, MobileNet can be implemented for several recognition problems, such as landmark recognition, fine-grain classification, face attributes, and object detection. Besides that, MobileNet is a small and efficient CNN model suitable for installation in real-time applications [32].

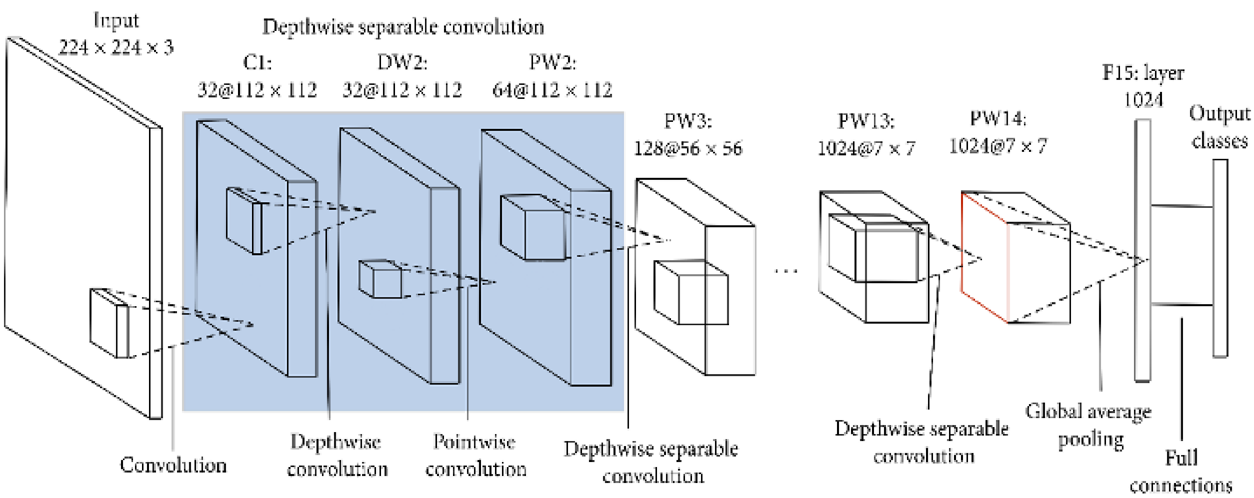

Fig. 5 *MobileNet architecture*

The Linux platform was the foundational framework for developing the MVDE method throughout this study. TensorFlow, a versatile framework for machine learning, was utilized to execute the object detection task. During the study's training and evaluation sections, the SSD with MobileNet architecture was utilized within this framework. In addition, OpenCV (Open-Source Computer Vision Library) was heavily utilized to support the COCO (Common Objects in Context) dataset for object detection and classification.

It is essential to emphasize that this study intentionally restricted its object detection scope to the 'Car' class. This strategic decision was made the study's primary emphasis on the functionalities of AES and AEB systems. By implementing this limitation, the study was able to focus more intently on the specific elements essential to the efficacy of these systems in real-world applications. This methodological approach improves the accuracy and applicability of the research's findings to the targeted safety systems.

3.2 Deep ANN Algorithm

This work uses the KITTI dataset to train and evaluate the object distance estimation method using Deep ANN. The rear-end accident involving two vehicles is the primary subject of this investigation. Because of this, just the class 'Car' is used to separate the dataset into the train and test datasets. In particular, the train and test data sets are split into 70% and 30%, respectively. There are only 28,742 data points with the class's 'car' out of 51,865 total. The class 'Car' data was split into 8,662 data for the test dataset and 20,080 for the training dataset.

Since there is a chance of overfitting and underfitting, which would harm the network's efficiency and time complexity, choosing hidden layers is particularly challenging [33]. When there are too many hidden layers, several circumstances may occur. When the number of hidden layers is vast in comparison to the complexity of the task, the overtraining of the network results from the overfitting condition. When the network's performance closely resembles the test data, it often affects the time complexity of the network [34]. When the number of hidden layers in the network is less than the complexity of the problem, underfitting circumstances arise [35]. Because it is frequently said to be undertrained, the network seldom tackles such challenges. It has a significant negative influence on the network's effectiveness. When faced with such a circumstance, the temporal complexity of the network drops to a deficient level, producing unproductive outcomes.

The proposed distance estimation techniques with Deep ANN architecture, as illustrated in Fig. 6, are trained and evaluated using the modified KITTI dataset. The Input Layer contains the input variables. The proposed method takes Xmin, Ymin, Xmax, and Ymax as input variables. These variables indicate the minimum and maximum values of the object's coordinates inside the bounding box data. The output variables are produced by a group of neurons in the output layer. This study's output variable generated by the Deep ANN architecture measures the distance from the camera (Zloc).

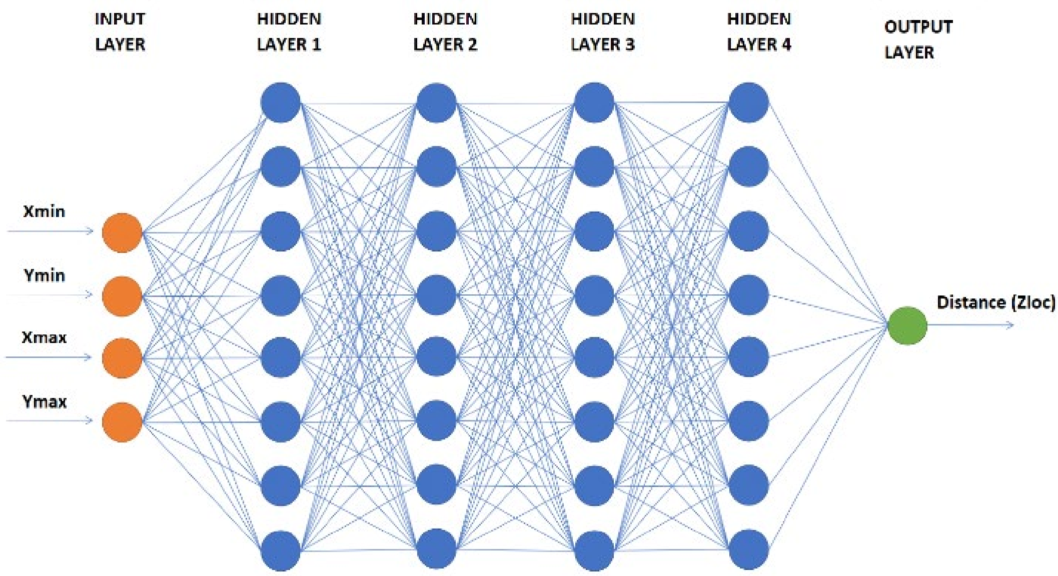

Fig. 6 *The deep ANN architecture*

The neural network architecture comprises intermediary layers known as Hidden Layers. These layers play a pivotal role in breaking down and processing data within the neural network. Determining the ideal network design and configuration can be challenging, as no one-size-fits-all solution exists. Instead, researchers often experiment with various configurations of Hidden Layers to identify the most efficient neural network model.

In this study, several Hidden Layer configurations were tested and compared to ascertain which one performed optimally. The results of these comparisons indicated that the most effective model consisted of four hidden layers, ReLU (Rectified Linear Unit) activation functions, a batch size of 2048, and 50 training epochs. This particular configuration demonstrated superior performance in the context of the study's objectives.

Int. Journal of Integrated Engineering Vol. 17 No. 1 (2025) p. 416-437

3.3 Distance Estimation

In this section, assuming the camera operates as effectively as a theoretical pinhole camera, the study aims to establish a method for calculating the Z value, representing the distance to an object of a known width from the camera. Fig. 7 visually illustrates converting a point in the two-dimensional plane (u, v) to a three-dimensional (X, Y, Z) coordinate system.

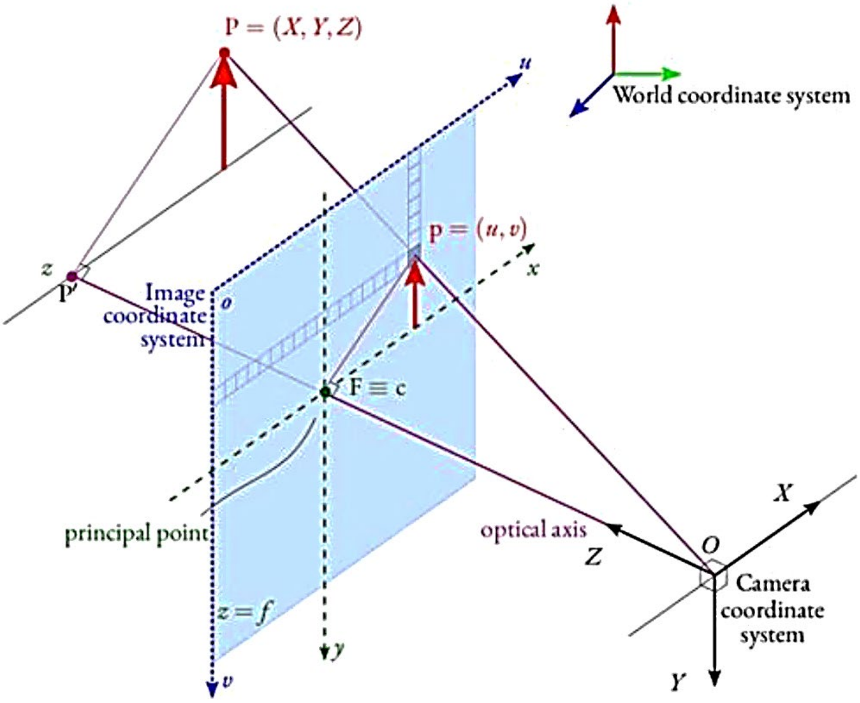

Fig. 7 *The model for pinhole camera*

This transformation is expressed through a mathematical model, as described by equation (1) [36].

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

Here, p signifies a projected point in the form $[u, v, 1]$, where K denotes the camera's intrinsic parameters, typically including the focal length and the principal point (u_o, v_0). These camera characteristics are typically specified in the camera's documentation. The conversion from an external world point to the internal camera viewpoint is detailed by the parameters $[R|t]$, which are extrinsic to the camera system. The Euclidean coordinate system, Q represents a 3D point with coordinates $[X, Y, Z, 1]$. Additionally, the variable s serves as a scaling factor that accounts for pixel scaling relative to the focal length shift, thereby encapsulating the intrinsic attributes of the camera. Consequently, equation (1) can be simplified to equation (2) and equation (3).

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

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

Equations (2) and (3) allow for the determination of X and Y values provided that the scale parameter 's' is assumed to be 1. Consequently, the Z parameter, representing the object's depth, can be computed using equation (6).

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

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

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

These equations collectively transform 2D image coordinates (u,v) into 3D world coordinates (X,Y,Z), a fundamental process in computer vision and 3D scene reconstruction. The values of Z can be computed once the other parameters and image coordinates are known, facilitating the estimation of distances in a three-dimensional space.

Initially collected for autonomous driving research, the KITTI dataset is implemented in real scenarios by training and testing computer vision models, especially Deep ANN. During training, models learn from the diverse images and sensor data set in the KITTI dataset, capturing real-world driving scenarios. In a real scenario, when the trained model encounters new images, it uses its learned knowledge to recognize objects and features. Deep ANN leverages techniques such as monocular depth estimation to translate KITTI data into distance. These models infer depth information from 2D images, utilizing the principles of stereo vision to estimate distances between objects in the scene. The ANN algorithms learn to associate pixel coordinates with corresponding distances, enabling accurate depth perception and distance estimation in real-world applications, such as autonomous driving.

In the specific context of enhancing Deep ANN algorithms trained on the KITTI dataset, these transformation equations play a pivotal role. The need for modifications arises from ensuring accurate distance estimation in real-world scenarios, particularly when dealing with images captured in dynamic environments. Adapting the ANN algorithms using these equations becomes crucial for several reasons.

Firstly, the KITTI dataset, often used for training deep learning models in autonomous driving scenarios, may not perfectly represent the diversity of real-world environments. Modifying the ANN algorithms helps bridge the gap between the dataset and real-world scenarios, allowing the models to better generalize and accurately estimate distances in dynamic and varied surroundings.

Secondly, the equations provide a mechanism to convert 2D pixel coordinates from images into meaningful 3D spatial information. This transformation is essential for understanding the depth and positioning of objects within a scene, a critical aspect for applications such as autonomous driving, where accurate distance estimation is fundamental.

By adapting the ANN algorithms using these transformation equations, the models gain the capability to translate pixel coordinates into precise 3D spatial coordinates, contributing to improved accuracy in estimating distances between objects. This adaptation enhances the model's capacity to navigate and interact with the real world, making it more robust and reliable in dynamic and challenging environments commonly encountered in autonomous driving scenarios. The fine-tuning facilitated by these equations is instrumental in aligning model predictions with the complexities of real-world environments, ensuring safer and more effective performance in practical applications.

3.4 Comparison with Existing Distance Estimation Methods

The proposed MVDE method is especially well-suited for real-time applications like AEB and AES as it has a number of benefits over existing methods. The MVDE employs a more straightforward and effective combination of the SSD with MobileNet for object detection and Deep ANN for distance estimation, leading to reduced computing costs, in contrast to traditional depth estimation methods that depend on complex models such as MonoDepth or LIDAR. Furthermore, compared to sensor-heavy techniques, the MVDE method is more affordable and easily accessible because it just needs a monocular camera. It is perfect for real-time vehicle safety systems because of its efficiency and smaller error margin, which provide quicker and more dependable performance in urgent circumstances.

Table 3 shows a comparison table between the proposed MVDE method and the existing deep learning-based distance estimation methods discussed in section 2, highlighting the advantages of the proposed method. This table demonstrates that the MVDE method is more accurate, computationally efficient, and cost-effective than the existing methods discussed in section 2, making it better suited for real-time distance estimation in autonomous emergency systems like AEB and AES.

Table 3 *The comparison between the proposed MVDE and the existing distance estimation methods*

Feature	Proposed MVDE Method	Existing Methods (from Section 2)	Advantages of MVDE
Object Detection Architecture	SSD with MobileNet	YOLOv5, Monodepth, DenseDepth, YOLOv3, BTS	Lightweight and efficient for real-time applications; MobileNet reduces computational cost.
Distance Estimation Approach	Deep ANN with four hidden layers	Depth Estimation (MonoDepth, DenseDepth), Object Distance Estimation (YOLO-based)	More accurate with less maximum measured error; optimized for specific real-time safety applications.
Computational Efficiency	High efficiency with reduced processing cost	High computational complexity for depth estimation	More computationally efficient, making it suitable for real-time integration in ADAS systems.
Sensor Dependency	Monocular camera only	LIDAR or other sensor-based methods integrated with deep learning	More cost-effective due to reliance on a single, affordable vision sensor (monocular camera).
Real-time Feasibility	Well-suited for real-time distance estimation in AEB and AES systems	Limited real-time functionality due to high processing requirements	More practical for real-world applications like rear-end collision avoidance with fast processing and accuracy.

4. Data Collection

To evaluate the validity of the suggested methodology, this research conducted a comparative analysis using both the observed distances and the distances obtained by RADAR as the reference standard. The experimental context for this evaluation was in Cyberjaya, Malaysia. In this context, the experiment included the assessment of both physical distances and distances obtained by RADAR technology. The measurement was conducted using laser measuring equipment, similar to the one seen in Fig. 8, in combination with RADAR technology.

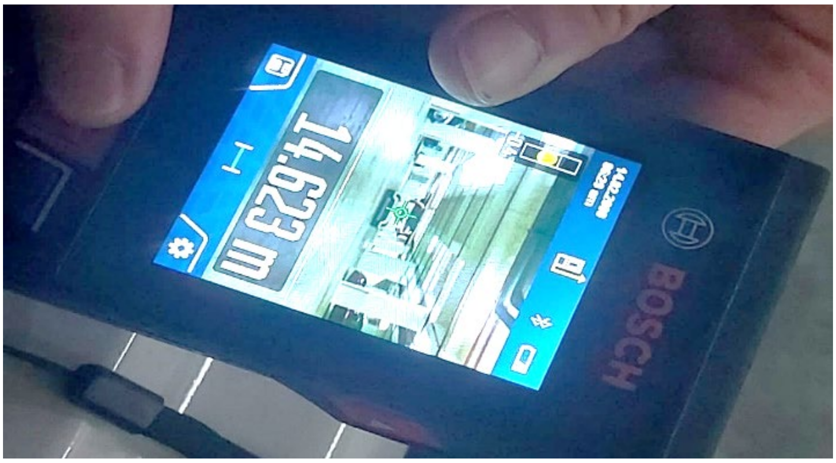

Fig. 8 *Laser measuring device*

In this experiment, vehicles of sedan size were used as the target vehicles. The selection of this kind of vehicle guaranteed a uniform and consistent method for evaluating distances. Incorporating certain-sized sedans improved the accuracy and relevance of the results, consequently strengthening the entire study's conclusions.

Before starting the data collection phase, a critical preparation measure was the installation of both the camera and RADAR components onto the host vehicle. The rigorous setup procedure is shown in Fig. 9, illustrating these crucial components' arrangement and placement. Significantly, the camera was deliberately placed at the highest point of the windshield, strategically positioned next to the rearview mirror. The positioning of the camera not only served to optimize the field of vision but also guaranteed an unobstructed viewpoint to facilitate the collection of data. In contrast, the placement of the RADAR module was carefully positioned in the forefront of the vehicle. This location was selected to use the RADAR's capability to detect and collect information on the vehicle near the environment. The strategic arrangement of these elements serves as a crucial first step, playing a pivotal role in establishing the groundwork for subsequent data-gathering endeavours.

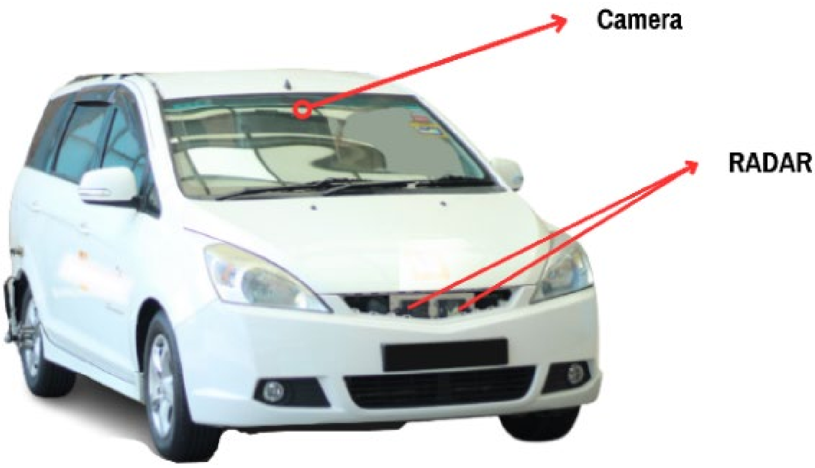

Fig. 9 *The position of the camera and RADAR at the host vehicle*

The camera used in this study was the STURDeCAM20 type. The device had a resolution of 1928 x 1088 pixels, with each pixel being 3µm × 3µm. The steady frame rate of 30 frames per second permitted real-time data recording. The camera exhibited a variable range of focus distances, extending from an ideal distance of 1 meter to infinity, facilitating diverse data collecting.

In contrast, the RADAR module used in the study was the Continental ARS 408-21. The RADAR system operated at a carrier frequency of 77 GHz and exhibited a noteworthy range capacity, encompassing distances ranging from a minimum of 0.20 meters to a significant 250 meters. The RADAR has a high angular resolution of 1.6°, enabling it to distinguish and analyse small-scale features throughout its scanning process accurately. Furthermore, the RADAR's ability to accurately capture spatial differences was emphasized by its distance resolution of 1.79 meters. The RADAR's ability to estimate distances with a precision of ±0.40 meters demonstrated accuracy, which was crucial for achieving the study goals. Table 4 summarizes the key attributes of the camera and RADAR components used in this study.

Table 4 *The characteristics of camera and RADAR*

Camera Characteristics [37]	
Model	STURDeCAM20
Pixel Size	3µm x 3µm
Resolution in Pixel (h x v)	1928 x 1088
Frame Rate	30 frames per second
Focus Distance	1m to infinity
RADAR Characteristics [38]	
Model	Continental ARS 408-21
Carrier Frequency	77 GHz
Range	0.20m to 250m
Angular Resolution	1.6°
Distance Resolution	1.79m
Distance Precision	±0.40m

Int. Journal of Integrated Engineering Vol. 17 No. 1 (2025) p. 416-437

Next, the distance between the primary vehicle and the secondary vehicle was modified to five meters using laser measuring instruments, as seen in Fig. 10. In the present context, the data was collected. Following the visual representation presented in Fig. 11, it subsequently increases the spacing between cars to distances of 10, 15, and 20 meters after the first measurement of 5 meters. Data was acquired at every distance. The collected data includes images of the vehicle of interest and the measured distance determined by the RADAR system. The study was conducted using three sedan-sized vehicles as test subjects.

Fig. 10 *The experiment setup between the host vehicle and the target vehicle*

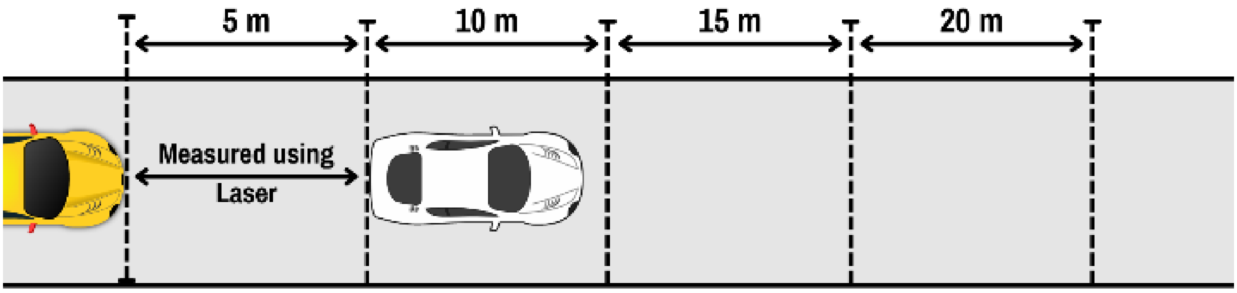

Fig. 11 *Diagram for the experiment setup*

The designated vehicle's visual representation was obtained using a solitary camera affixed to the primary vehicle. One frame was chosen from a sequence of 10 successive frames based on its representative features, which were comparable to each other. This selection was made in order to optimize computational efficiency. The photos underwent preprocessing using the MVDE approach. The bounding box around the target vehicle was determined using the SSD with MobileNet architecture. Afterward, the Deep ANN technique was used to calculate the distance between the camera and the target vehicle. This measured distance was achieved using the bounding box information obtained from the object recognition process. Then, the outcome of the MVDE method was compared with the measured distance between RADAR and the actual distance.

In this experiment, due to the limited available space and the necessity to adhere to stringent testing protocols within the host vehicle, only four specific distances – five, ten, fifteen, and twenty meters – were incorporated. Moreover, safety considerations were paramount to selecting three sedan-sized vehicles as target objects. It is worth noting that this deliberate selection facilitated the optimal execution of the experiment while ensuring safety. However, it is essential to acknowledge that, should future authorization be granted for further experimentation, a broader range of vehicles could be engaged. This expansion in vehicle types would significantly augment the validation process and offer a more comprehensive understanding of the appropriateness and effectiveness of the MVDE method within real-world scenarios. Expanding the scope of vehicles would provide an enriched dataset and a more robust evaluation of the MVDE method's performance, thus reinforcing its applicability and relevance in practical settings.

5. Result and Discussion

There are several subsections to be discussed in this section. All subsections will be elaborated in detail as they are compilations of the results for developing the proposed method. Also, the test results comparison is discussed in this chapter.

5.1 Deep ANN Algorithms

The performance of the proposed Deep ANN algorithm is validated using the KITTI dataset, employing five different numbers of hidden layers in the Deep ANN architecture. The first architecture includes one hidden layer, while subsequent architectures feature two, three, four, and five hidden layers, respectively. This range from 1 to 5 hidden layers is selected to mitigate the risk of overtraining. Excessive hidden layers relative to the task's complexity can negatively impact the network's time complexity when its performance closely aligns with the test data. This is particularly pertinent given that the input layer comprises only four input variables. Moreover, each hidden layer is equipped with 8 neurons to prevent undertraining. The optimal distance estimation algorithm is determined by comparing the performance across varying numbers of hidden layers in the Deep ANN architecture, evaluating prediction accuracy and validation metrics.

The KITTI dataset was utilized for training and validating distance prediction using each proposed distance estimation algorithm. After training, the distance prediction capability of these algorithms was evaluated. Fig. 12 illustrates the visualization of the predictions. Actual and predicted distances are denoted by red text in this visualization, while the green boundary box represents vehicle detection. It is essential to observe that the proposed method is designed to detect only a single object at a time, with a bias for centrally located objects.

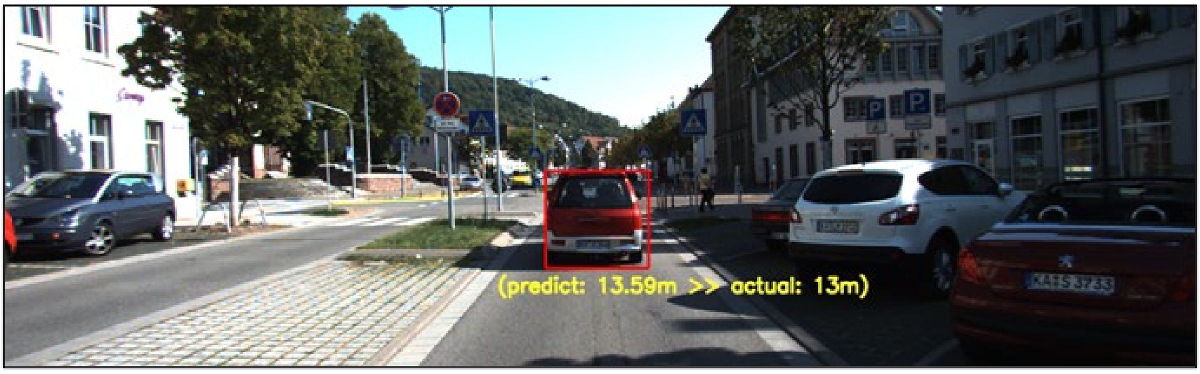

Fig. 12 *Result of proposed deep ANN algorithm*

In addition, it should be noted that the training images within the KITTI dataset, specifically those corresponding to the 'car' class, have certain limitations about the representation of car dimensions. According to the study, the proposed method produced superior results when applied to sedan-sized vehicles. Due to the training phase only training the class 'car', it was determined that this method could not reliably identify additional classes, such as 'truck,' 'van,' and 'motorcycle.'

Fig. 13 shows a sampling of the distance estimation algorithms' outcomes. The value is selected based on the lowest measured error value. The study's results indicated that the proposed distance estimation method exhibited the lowest measured error, particularly when incorporating four hidden layers in the Deep ANN architecture. Moreover, an increase in the number of hidden layers within the Deep ANN architecture corresponded with a reduction in measurement error. However, their measured error values were relatively similar when comparing the four and five hidden layers, indicating that the marginal benefit of additional hidden layers diminished in some instances.

However, a distinct pattern emerged for distances exceeding 55 meters within the proposed Deep ANN architectures. Specifically, as the evaluated distance increased, the measured error also increased. This phenomenon can be attributed to a limitation inherent in the training dataset, particularly concerning scenarios involving greater distances. Consequently, the algorithm's predictive capability becomes constrained when addressing situations lying beyond the effective range covered by the dataset.

Table 5 compares the measured error between the proposed method and the methodology introduced by Karthika et al. [22]. The proposed method's use of Deep ANN with four hidden layers demonstrates a significant benefit compared to the technique suggested by Karthika et al. [22]. The approach that has been developed consistently exhibits improved performance over a range of distances, hence emphasizing its efficacy in improving the accuracy of distance measurement.

Int. Journal of Integrated Engineering Vol. 17 No. 1 (2025) p. 416-437

Despite this, it is interesting to note that the approach proposed by Karthika et al. demonstrates a more advantageous result, particularly when considering 60 meters. The observed difference in outcomes at this specific distance necessitates more investigation and has the potential to provide insights into the different features of both approaches. Given the study's central focus on AEB and AES systems, it is noteworthy to consider the crucial significance of accurate distance estimation for close distances. The study primarily emphasizes collision avoidance and safety improvement in circumstances involving very near proximity. Consequently, the accuracy of estimating distances at shorter ranges is of particular importance.

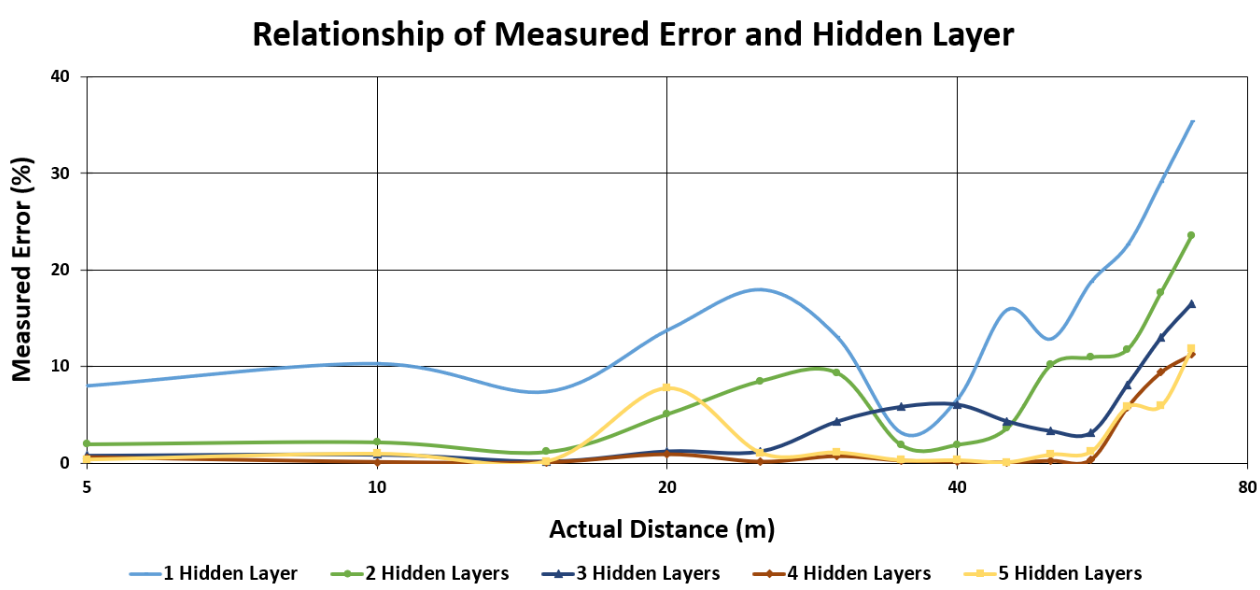

Fig. 13 *Relationship of measured error and hidden layer*

Table 5 *Comparison of distance estimation results from the method of [22] and the proposed method with 4 hidden layers*

Actual Distance (m)	Estimated Distance of Proposed Method 4 Hidden Layers (m)	Measured Error (%)	Estimated Distance from Other Method [22] (m)	Measured Error (%)
10	10.01	**0.09**	10.21	**2.10**
20	19.82	**0.92**	20.12	**0.60**
30	29.79	**0.70**	30.53	**1.76**
40	40.05	**0.13**	40.9	**2.25**
50	49.89	**0.22**	50.42	**0.84**
60	56.53	**5.78**	60.61	**1.01**

5.2 Validation of MVDE Method

Fig. 14, Fig. 15, and Fig. 16 show the results obtained using the MVDE approach, illustrating the outcomes for three separate sedan-sized automobiles denoted as Car A, Car B, and Car C, correspondingly. The results presented in this study provide valuable insights into the performance of the MVDE approach in different settings. Each picture consists of a sequence of images, with the image in the top left representing a target distance of 5 meters, followed by the image in the top right displaying 10 meters. The pictures in the diagram's lower left and lower right sections are associated with distances of 15 meters and 20 meters, respectively. Significantly, these photos jointly provide a complete perspective on the estimated outputs of the MVDE approach over various distances.

In order to enhance understanding, the visual components included within the images are encoded with informative properties. The yellow text in the images represents the distance estimates obtained from the proposed MVDE approach, which is a reliable measure of the system's prediction precision. Furthermore, the visual representation utilizes red bounding boxes to define and identify occurrences of effective vehicle detection.

The image presentation, shown in Fig. 14, Fig. 15, and Fig. 16, serves as a demonstrative exhibition of the performance and effectiveness of the MVDE approach in estimating distances for automobiles of sedan size. The visual representation of estimated results at different distances enhances the flexibility and practicality of the procedure in real-life situations.

The line graph shown in Fig. 17 shows the relationship between the distances created by RADAR and the corresponding absolute distances. The presented visual representation provides valuable information about the consistency and alignment between the two distance measurements in various settings. Upon careful analysis of the graph, it becomes apparent that the distances obtained using RADAR often align with the actual distances. The alignment described is consistently seen in most cases, indicating a dependable estimation accomplished by using RADAR technology.

Nevertheless, it is worth noting that a more detailed analysis may be conducted concerning the RADAR distance measurement, specifically at the 15-meter point. An appreciable absolute error distance of 1.4 meters is recorded at this distance. The measured gap may be attributable to the inherent limits in accuracy associated with RADAR distance measurements. Moreover, it is essential to recognize that multiple circumstances may impact the precision of RADAR distance measurements. The issues mentioned above include electrical noise, the phenomena of multipath propagation, and the impact of reflections from nearby objects. As a result, while the RADAR technology can provide relatively accurate estimations of distances, its accuracy may be influenced by external factors that must be considered in practical scenarios.

Fig. 14 *Result of proposed MVDE method for Car A*

Fig. 15 *Result of proposed MVDE method for Car B*

Fig. 16 *Result of proposed MVDE method for Car C*

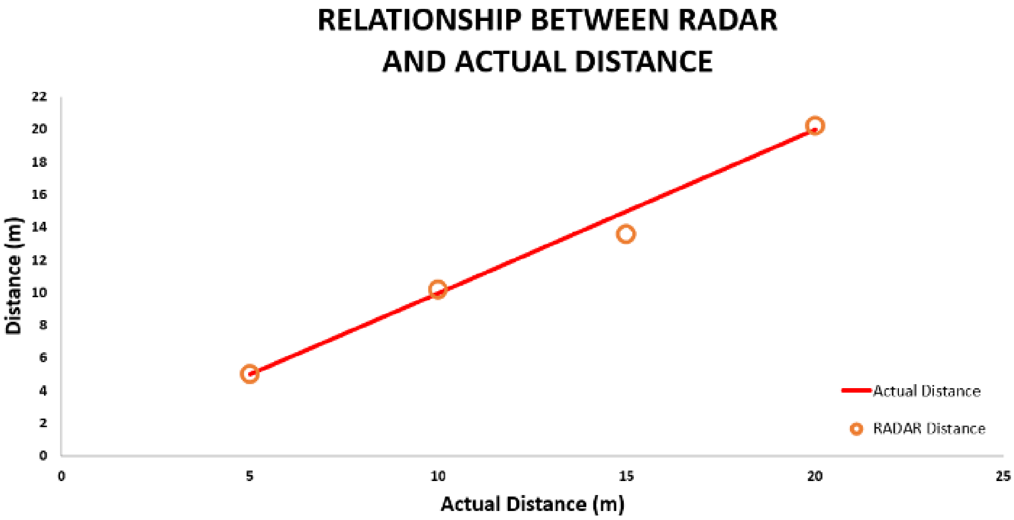

Fig. 17 *Result between RADAR distance and actual distance*

Fig. 18 illustrates the contrast between the distance estimates produced by the proposed MVDE approach and the actual distances. The graph demonstrates the level of accuracy attained by the MVDE approach provided in the research for all three sedan-sized cars analysed. Upon careful graph analysis, distinct accuracy patterns in distance estimates become apparent. The distance estimates obtained by the proposed MVDE approach demonstrate a notable level of precision, mainly when measured at distances of 5 meters and 15 meters. These estimations reveal a maximum measured error of ±1 meter. The previously described level of accuracy highlights the MVDE approach's effectiveness in providing accurate estimations of distances within these specified ranges.

Upon further examination of the graph, it can be shown that the suggested MVDE approach exhibits satisfactory accuracy at 20 meters. However, it is worth noting that there is a minor increase in the margin of error. In the given context, the upper limit of the measured error is ±2 meters, indicating that the predicted distances nearly correspond to the actual lengths. Notably, the MVDE method's distance estimate at 10 meters has a comparatively wider margin of error when compared to the actual lengths. The most significant observed error ranges from ±4 meters at this distance. The discovery above underscores the need to exercise caution while using the suggested MVDE approach within this specific range of distances. In conclusion, the graph successfully demonstrates the high level of accuracy in distance estimates achieved by the proposed MVDE method for varied distances. Additionally, it reveals subtle differences in accuracy across several distance intervals.

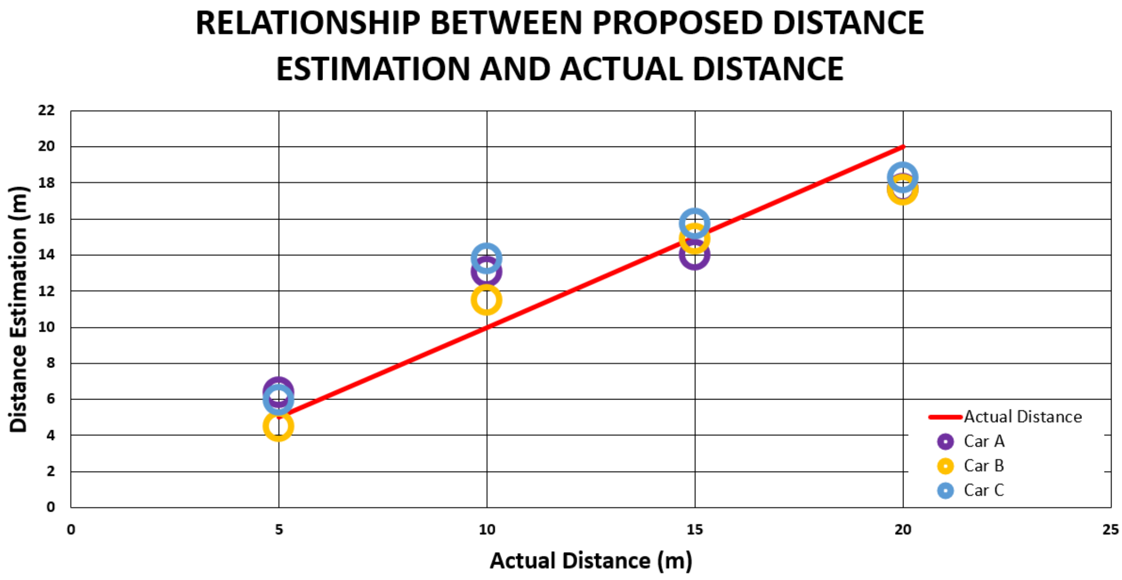

Fig. 18 *Result between proposed MVDE method and actual distance*

Moving forward to Fig. 19, the graphical representation highlights the correlation between the proposed MVDE method and the RADAR distance. Notably, the accuracy of distance estimation by the proposed MVDE method at 5 meters emerges as the most precise among the various distances considered. This is underscored by the maximum measured error at 5 meters being limited to ±1 meter. At a separation of 10 meters, there is a notable difference in accuracy between the distance estimation of the MVDE method and the RADAR distance, as seen by the highest observed error of ±4 meters. This implies that distance estimation accuracy somewhat decreases at this specific distance.

Similarly, the graph illustrates that the precision of the distance estimation based on the proposed MVDE method, compared to the RADAR distance at 15 meters, remains at a reasonable level. The observed maximum error is limited to a range of ±2 meters, strengthening the method's effectiveness in accurately calculating distances within this interval. Finally, in the case of distances spanning 20 meters, the distance estimations generated by the proposed MVDE method maintain a relatively stable accuracy but with a slightly larger margin of error. The maximum measured error at 20 meters is ±3 meters, indicating that the method's estimations remain reasonably close to the RADAR distance. In summary, the graph in Fig. 19 provides an insightful visualization of the varying degrees of accuracy between the proposed MVDE method's distance estimations and the RADAR distance across different intervals.

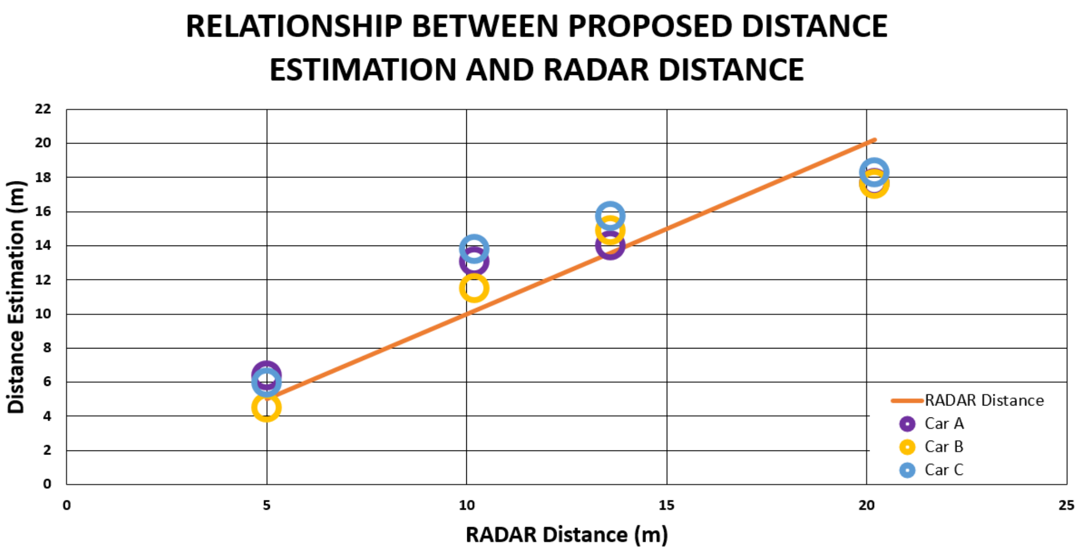

Fig. 19 *Result between proposed MVDE method and RADAR distance*

Fig. 20 comprehensively compares three essential elements: the MVDE method, the RADAR distance, and experimentally determined distances. This comparison's primary objective is to evaluate the performance of the MVDE method and its potential as a substitute for the traditional RADAR system in determining the distance between vehicles. The results shown in this figure exhibit significant promise and underscore the exceptional quality of the MVDE approach. It becomes apparent that the MVDE approach is not only a feasible option but also a very efficient one since it consistently generates distance estimates that nearly correspond to those obtained from experimental measurements.

The significance of this alignment lies in its emphasis on the possibility of the MVDE approach to function as a dependable alternative to RADAR for estimating distances. The MVDE approach demonstrates notable precision in predicting distances, vital in augmenting the effectiveness of advanced safety systems like AEB and AES integrated inside the vehicle's system. Moreover, it is essential to acknowledge that any mistakes or flaws in the distance assessment might directly influence the performance of AEB and AES. The potential consequences of inaccurate distance calculations include increased probabilities of delayed and incorrect reactions, amplifying the likelihood of vehicular accidents. Hence, the efficiency of the MVDE approach in delivering accurate distance estimates plays a crucial role in guaranteeing the safety and effectiveness of these devices.

In conclusion, Fig. 20 provides a compelling visual representation of the MVDE method's ability to accurately estimate distances between vehicles, positioning it as a promising replacement for conventional RADAR systems in this crucial automotive application.

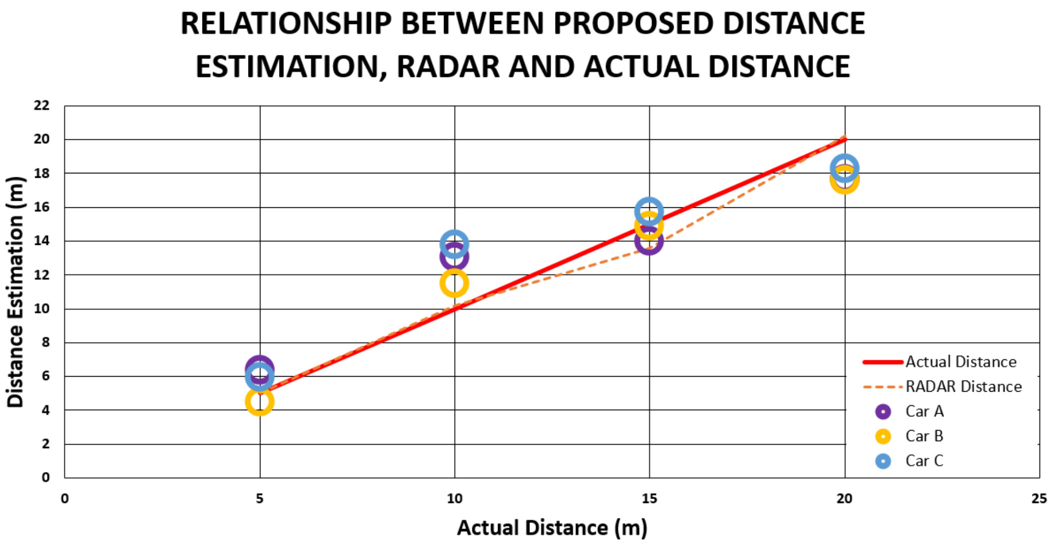

Fig. 20 *Result between MVDE method, RADAR distance, and actual distance*

Fig. 21 illustrates the discrepancies or inconsistencies between the RADAR distance measurements and the estimated distances when the actual distances are considered. For 5 and 15 meters, it is evident that the discrepancies are relatively consistent, lingering around one meter. The small error margin confirms that the RADAR measurements and estimated distances have a reasonable degree of precision and congruence, as indicated by the correlation between the two. However, a discernible divergence becomes apparent as we approach 10 and 20 meters. The differences between the RADAR distance measurements and the estimated distances become greater at these distances. For 10 meters, the error is approximately four meters, whereas for 20 meters, the error decreases but remains substantial at around two meters. These discrepancies between distance ranges emphasize the difficulties and complexities involved in distance estimation, particularly at greater distances. Discrepancies in measurement precision can be caused by sensor limitations and environmental conditions, which become more pronounced as distance increases.

In conclusion, Fig. 21 effectively illustrates the disparities or errors between RADAR distance measurements and estimated distances regarding actual distances. It demonstrates that while the estimation method maintains reasonable precision at certain distances, there are still obstacles to attaining precision at greater distances. These insights help comprehend the method's efficacy at varying distances. The results provide valuable insights into the accuracy of the MVDE method compared to RADAR distance measurements. The MVDE method exhibits an accuracy rate of 93%, whereas RADAR distance measurements achieve a notably higher accuracy of 98.4%. It is worth noting that while RADAR distance measurements outperform the MVDE method in terms of accuracy, there are essential cost considerations to consider. RADAR technology is associated with high costs, which can present barriers to widespread adoption, especially in regions with limited resources.

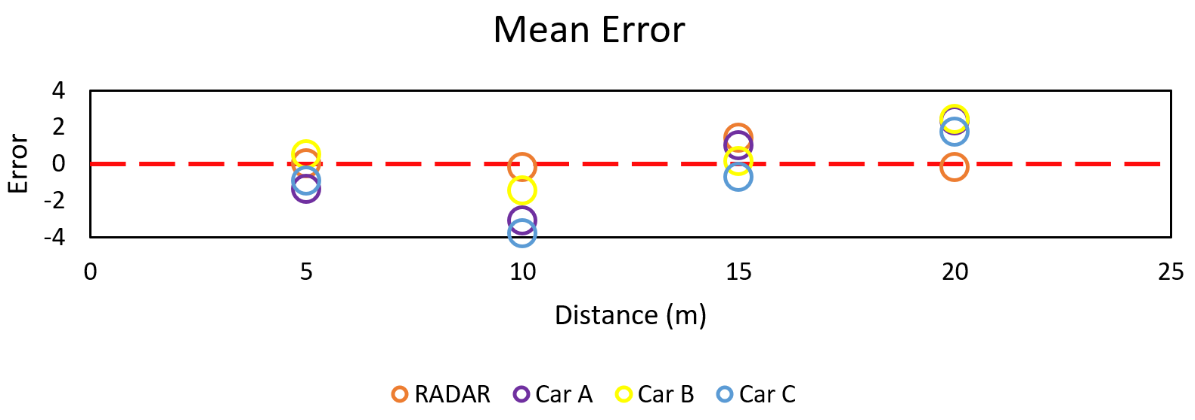

Fig. 21 *Mean error of the result between MVDE method, RADAR distance, and actual distance*

The primary motivation for utilizing monocular vision instead of RADAR is cost-effectiveness. By employing monocular vision and the MVDE method, the aim is to make AEB and AES systems more accessible to a broader range of countries and regions. This approach seeks to balance affordability and effectiveness, allowing these safety systems to benefit a wider global population. The study underscores the trade-off between accuracy and cost-effectiveness in AEB and AES systems. While RADAR offers superior accuracy, the MVDE method's affordability makes it a promising choice for enhancing road safety in areas where cost constraints are a significant consideration.

Section 2 discussed two primary methods for deep learning-based distance estimation: depth estimation and object distance estimation. We opted for the latter method in our study, while Adz-Dzikri et al. [18] employed the former approach. Both our MVDE method and the method used by Adz-Dzikri et al. relied on the SSD technique for object detection. A comprehensive comparison of the two methods was conducted, focusing on their maximum measured errors, as presented in Table 6. Notably, the method employed by Adz-Dzikri et al. exhibited a maximum measured error of ±14m. In contrast, our MVDE method, as implemented in this study, displayed a significantly improved performance with a maximum measured error of ±4m compared to actual distance measurements.

These findings emphasize that the MVDE method introduced in our study outperforms the method used by Adz-Dzikri et al. [18]. Consequently, our research demonstrates the superiority of the object distance estimation method over depth estimation for accurate distance estimation, particularly in the context of AEB and AES systems.

Table 6 *The measured error between the MVDE method and Adz-Dzikri et al. method*

Actual Distance (m)	Measured Error (m)			
	MVDE Method			MonoDepth2 by Adz-Dzikri et al.
	Car A	Car B	Car C	
5	1.37	0.49	0.93	2.723
10	3.08	1.47	3.78	6.166
15	1.01	0.12	0.72	9.17
20	2.35	2.41	1.73	13.631

6. Conclusion

In summary, this study introduced a MVDE method, leveraging deep learning techniques, to estimate distances between vehicles, particularly for AEB and AES systems. Our MVDE method demonstrated remarkable accuracy through extensive real-world experiments, with a maximum measured error of ±4m compared to actual distances. Although RADAR achieved higher accuracy at 98.4%, our method offers a cost-effective alternative with 93% accuracy, making AEB and AES systems more accessible. Additionally, our MVDE method outperformed depth estimation methods, showcasing its potential as a superior solution. This research contributes to improved vehicle safety and reduces costs, making these safety features more widely available. Future work can further enhance the MVDE method's versatility and expand its applications, ultimately shaping the future of automotive safety standards and intelligent mobility.

Acknowledgement

This research is supported by the Ministry of Higher Education Malaysia and Universiti Teknologi Malaysia under UTM Fundamental Research Grant Vot. No. 22H27.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The corresponding author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

References

[1] Zou, Q., Jiang, H., Dai, Q., Yue, Y., Chen, L., & Wang, Q. (2019). Robust lane detection from continuous driving scenes using deep neural networks. *IEEE transactions on vehicular technology, 69*(1), 41-54.
[2] Zakaria, N. J., Shapiai, M. I., Fauzi, H., Elhawary, H. M. A., Yahya, W. J., Rahman, M. A. A., Kassim, K. A. A., Bahiuddin, I., & Ariff, M. H. M. (2020). Gradient-based edge effects on lane marking detection using a deep learning-based approach. *Arabian Journal for Science and Engineering, 45*, 10989-11006.
[3] Zakaria, N. J., Shapiai, M. I., Abd Ghani, R., Yassin, M. N. M., Ibrahim, M. Z., & Wahid, N. (2023). Lane detection in autonomous vehicles: A systematic review. *IEEE Access, 11*, 3729-3765.
[4] Monsaingeon, N., Caroux, L., Langlois, S., & Lemercier, C. (2023). Multimodal interface and reliability displays: Effect on attention, mode awareness, and trust in partially automated vehicles. *Frontiers in Psychology, 14*, 1107847.
[5] Formosa, N., Quddus, M., Man, C. K., Singh, M. K., Morton, C., & Masera, C. B. (2024). Evaluating the impact of lane marking quality on the operation of Autonomous Vehicles. *Journal of Transportation Engineering, Part A: Systems, 150*(1), 04023126.
[6] Vater, C., Wolfe, B., & Rosenholtz, R. (2022). Peripheral vision in real-world tasks: A systematic review. *Psychonomic bulletin & review, 29*(5), 1531-1557.
[7] Halimi, S. N. A., Rasmana, S. T., Adiputra, D., Yahya, W. J., Rahman, M. A. A., Ariff, M. H. M., Husain, N. A., & Kassim, K. A. A. (2023). A review of safety test methods for new car assessment program in Southeast Asian countries. *Open Engineering, 13*(1), 20220501.
[8] Ahmed, M. M., Khan, M. N., Das, A., & Dadvar, S. E. (2022). Global lessons learned from naturalistic driving studies to advance traffic safety and operation research: A systematic review. *Accident Analysis & Prevention, 167*, 106568.
[9] Yang, L., Yang, Y., Wu, G., Zhao, X., Fang, S., Liao, X., Wang, R., & Zhang, M. (2022). A systematic review of autonomous emergency braking system: impact factor, technology, and performance evaluation. *Journal of advanced transportation, 2022*.
[10] Yang, W., Zhang, X., Lei, Q., & Cheng, X. (2019). Research on longitudinal active collision avoidance of autonomous emergency braking pedestrian system (AEB-P). *Sensors, 19*(21), 4671.
[11] Lowe, E., & Guvenç, L. (2021). A Review of Autonomous Road Vehicle Integrated Approaches to an Emergency Obstacle Avoidance Maneuver. *arXiv preprint arXiv:2105.09446*.
[12] Lin, Y. M., & Chen, B. C. (2021). Handling Enhancement of Autonomous Emergency Steering for Reduced Road Friction Using Steering and Differential Braking. *Applied Sciences, 11*(11). doi:10.3390/app11114891
[13] Bhardwaj, A., Lu, Y., Pan, S., Sarter, N., & Gillespie, B. (2020). *The effects of driver coupling and automation impedance on emergency steering interventions.* Paper presented at the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC).
[14] Dirgantara, F. M., Rohman, A. S., & Yulianti, L. (2019). *Object Distance Measurement System Using Monocular Camera on Vehicle.* Paper presented at the 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI).
[15] Khan, F., Salahuddin, S., & Javidnia, H. (2020). Deep learning-based monocular depth estimation methods—A state-of-the-art review. *Sensors, 20*(8), 2272.
[16] Xu, D., Wang, W., Tang, H., Liu, H., Sebe, N., & Ricci, E. (2018). *Structured attention guided convolutional neural fields for monocular depth estimation.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
[17] Masoumian, A., Marei, D. G., Abdulwahab, S., Cristiano, J., Puig, D., & Rashwan, H. A. (2021). *Absolute Distance Prediction Based on Deep Learning Object Detection and Monocular Depth Estimation Models.* Paper presented at the CCIA.

[18] Adz-Dzikri, A. A., Virgono, A., & Dirgantara, F. M. (2021). *Advance Driving Assistance Systems: Object Detection and Distance Estimation Using Deep Learning.* Paper presented at the 2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI).
[19] Back, M., Lee, J., Bae, K., Hwang, S. S., & Chun, I. Y. (2021). *Improved and efficient inter-vehicle distance estimation using road gradients of both ego and target vehicles.* Paper presented at the 2021 IEEE International Conference on Autonomous Systems (ICAS).
[20] Ma, J., Yu, C., Xia, Y., Ren, X., Tsviatkou, V. Y., & Boriskevich, A. A. (2022). *Framework for Estimating Distance and Detecting Object on Mono-camera.* Paper presented at the 2022 International Conference on Intelligent Systems and Computer Vision (ISCV).
[21] Parmar, Y., Natarajan, S., & Sobha, G. (2019). DeepRange: deep‐learning‐based object detection and ranging in autonomous driving. *IET Intelligent Transport Systems, 13*(8), 1256-1264.
[22] Karthika, K., Adarsh, S., & Ramachandran, K. I. (2020). *Distance Estimation of Preceding Vehicle Based on Mono Vision Camera and Artificial Neural Networks.* Paper presented at the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT).
[23] Vajgl, M., Hurtik, P., & Nejezchleba, T. (2022). Dist-YOLO: Fast Object Detection with Distance Estimation. *Applied Sciences, 12*(3), 1354.
[24] Arabi, S., Sharma, A., Reyes, M., Hamann, C., & Peek-Asa, C. (2022). Farm Vehicle Following Distance Estimation Using Deep Learning and Monocular Camera Images. *Sensors, 22*(7), 2736.
[25] Geiger, A., Lenz, P., & Urtasun, R. (2012). *Are we ready for autonomous driving? the kitti vision benchmark suite.* Paper presented at the 2012 IEEE conference on computer vision and pattern recognition.
[26] Bao, D., & Wang, P. (2016). *Vehicle distance detection based on monocular vision.* Paper presented at the 2016 International Conference on Progress in Informatics and Computing (PIC).
[27] Forsyth, D., & Ponce, J. (2012). Computer vision: A modern approach. Always learning. In: Pearson.
[28] Zhang, H., & Hong, X. (2019). Recent progresses on object detection: a brief review. *Multimedia Tools and Applications, 78*(19), 27809-27847.
[29] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). *Ssd: Single shot multibox detector.* Paper presented at the European conference on computer vision.
[30] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
[31] Aziz, L., Salam, M. S. B. H., Sheikh, U. U., & Ayub, S. (2020). Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review. *IEEE Access, 8*, 170461-170495.
[32] Li, Y., Wang, H., Dang, L. M., Nguyen, T. N., Han, D., Lee, A., Jang, I., & Moon, H. (2020). A deep learning-based hybrid framework for object detection and recognition in autonomous driving. *IEEE Access, 8*, 194228-194239.
[33] Panchal, F. S., & Panchal, M. (2014). Review on methods of selecting number of hidden nodes in artificial neural network. *International Journal of Computer Science and Mobile Computing, 3*(11), 455-464.
[34] Gupta, T. K., & Raza, K. (2020). Optimizing deep feedforward neural network architecture: A tabu search based approach. *Neural Processing Letters, 51*(3), 2855-2870.
[35] Zhang, Q. S., & Zhu, S. C. (2018). Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering, 19*(1), 27-39.
[36] Saleh, H., Saleh, S., Toure, N. T., & Hardt, W. (2021). *Robust Collision Warning System based on Multi Objects Distance Estimation.* Paper presented at the 2021 IEEE Concurrent Processes Architectures and Embedded Systems Virtual Conference (COPA).
[37] STURDeCAM20 - IP67 Full HD GMSL2 HDR camera for Surround view system. Retrieved from https://www.e-consystems.com/camera-modules/gmsl2-surround-view-camera.asp
[38] Liebske, R. (2018). ARS 404-21 Entry and ARS 408-21 Premium-Short Description. from Continental