Deep Learning Techniques for Road Lane Detection for Traffic Control

T.Srilalith[1]
Assistant Professor, Talla Padmavathi College of Engineering., Warangal, India.
Email id: **sreethuluva@gmail.com**
Dr. N. Venkatesh[2]
Assistant Professor, SR University,Warangal, India.
Email id: **naramulavenkatesh@gmail.com**

Abstract

The development of intelligent traffic systems and autonomous driving has depended heavily on road lane identification in recent years. To improve road lane recognition accuracy and resilience, this research introduces a unique strategy that combines classic machine learning approaches with hybrid deep learning. The suggested approach combines the effectiveness of machine learning techniques for lane border classification and tracking with the advantages of convolutional neural networks (CNNs) for feature extraction and pattern identification. Using a hybrid model, the system retains the interpretability and flexibility of machine learning while gaining access to deep learning's high-dimensional data processing capabilities. This research presents a comparative analysis of many deep neural network designs that can infer surface normal information on the traditional KITTI road dataset under a range of demanding conditions. By testing it on three cutting-edge deep learning models—"Resnet-50," "Xception," and "MobileNet-V2"—we want to streamline the process of how current approaches interpret ground-related data and offer a solution. This will allow us to better understand and use the capabilities of these models. This comparative study's primary contribution has been to assess these networks' performance for edge deployment. Therefore, the little DNN model of MobileNet-V2 has been taken into consideration, which has about 80% less adjustable parameters than the other models. The acquired outcomes demonstrate that the suggested networks may get a segmentation accuracy of over about 96%, even in a variety of difficult situations. The method is tested on a large dataset with different surroundings and road conditions. Comparing experimental results to state-of-the-art techniques, one can observe notable gains in computing efficiency and detection accuracy. The integration of this sturdy lane identification technology into traffic management systems can facilitate the development of autonomous driving solutions that are more dependable and safer.

Keywords: Autonomous driving, Driver assistance system, Semantic segmentation, Machine learning, Deep learning

Introduction

The development of intelligent traffic management systems, where precise road lane recognition is essential, has been accelerated by the quick progress of autonomous driving technology. Road lane recognition is critical in autonomous driving safety, navigation, and vehicle localization. Complex road conditions, changing illumination, and occlusions are common challenges for traditional lane recognition approaches, which rely on manually created features and conventional image processing techniques. More resilient and flexible methods must be developed to overcome these constraints. Autonomous driving is seen as a difficult, open, and quickly developing study topic in the automotive industry. The Society of Automobile Engineers' definition of level five automation has been the subject of several academic and industrial research projects (SAE International, 2016). The data required for a modern car's Advanced Driver Assistance System (ADAS) to effectively handle tasks like adaptive cruise control, parking assistance, pedestrian detection, path planning, or collision avoidance is provided by sensors like cameras, RADAR, or LIDAR, which refers to remote sensing technology using laser light to measure distances. Most of these ADAS capabilities have been greatly impacted by current developments and the effectiveness of deep learning methods for image processing applications. This occurred mostly as a result of the Convolutional Neural Network's (CNN) ability to extract semantic information from pictures and make inferences about them in tasks such as 3D point cloud segmentation and semantic segmentation. The area of autonomous vehicles has advanced significantly over the past couple of decades, and DARPA has contributed significantly to these advancements.

Advanced driver assistance systems (ADAS) are systems that self-driving cars have been designed to employ a variety of onboard sensors, including as cameras, LiDARs, and GPS, to collectively detect the changing surrounding scene and make the appropriate judgments for safe navigation. Recent advancements in deep learning and multi-sensor fusion techniques have facilitated the creation of autonomous driving systems that are safe, effective, and ready for consumers. Multi-modal sensor fusion techniques and artificial intelligence are commonly employed in tandem to create perception systems that are able to sense their surroundings, anticipate traffic patterns, plan routes or assist with lane changes, and carry out these decisions in real-world scenarios. These intelligent perception systems should be precise, reliable, and responsive. All of these will contribute to the development of autonomous intelligent vehicle systems, which will lessen traffic, decrease accidents on the roads, and improve the efficiency and economy of transportation. The goal of the current effort is to construct a deep neural network architecture that can identify the areas of a driving scenario that are suitable for driving. The fundamental method for the semantic segmentation of the road surfaces in the suggested RoadSegNet is Google's DeepLavV3+. Typically, the RoadSegNet employs weights from three separate pretrained networks: one small DNN from MobileNet-V2, two high-accuracy models from ResNet50 and XceptionNet, and one small model from ResNet50. In this work, the RoadSegNet was trained using the Vision Benchmark Suite Data Set. The DeepMind-V3+ encoder-decoder architecture is used for segmentation, and the study compares the three cutting-edge DNNs—ResNet50, XceptionNet, and MobileNet-V2—in it. The design of these pretrained networks is a significant consideration, in addition to their use for weight initialization. The design and quantity of training parameters shared by these DNNs are distinctive: ResNet50 has 23 million trainable parameters, XceptionNet has 22.8 million, and MobileNet-V2 has just 4.2 million. This comparative study's primary contribution has been to assess these networks' performance for edge deployment. Thus, the MobileNet-V2 small DNN model has been taken into consideration; it has around 80% less configurable parameters than the other models, which makes it ideal for edge deployment. Three criteria have been used to assess the performance of these trained models: mean BF score, weighted IOU, and global accuracy. The worldwide accuracy provided by the trained models ranges from 96% to 97%. Additionally, it has been noted that MobileNet-V2's performance, in spite of its small deep neural network design, is on par with XceptionNet's and, in certain situations, even outperforms ResNet50.

Road lane detection is a critical first step toward safe and dependable autonomous driving as passenger safety is the intelligent system of any autonomous vehicle's top priority. As it controls the vehicle's driving area and the driver's actions (such as lane changes or overtakes), accurate lane recognition is a crucial need for all the previously described ADAS system capabilities. Despite some standardization in lane designs, accurate real-time lane recognition is a challenging problem since actual lanes may differ in form and color. Conventional computer vision-based lane recognition methods identify the road lanes by employing image processing methods such as the Hough Transform, Sobel filters, and Canny Edge detection (Feniche & Mazri, 2019; Marzougui et al., 2020). However, when the road picture scenarios are complicated in terms of weather, illumination, fading road lines, or occlusions produced by objects and pedestrians, these antiquated approaches are not strong enough to recognize the lanes (H. Hou et al., 2021).

Deep learning has made significant advances in computer vision in recent years by making it possible to extract high-level characteristics from unprocessed input data. A subtype of deep learning called convolutional neural networks (CNNs) has shown remarkable performance in object detection, picture segmentation, and lane detection.. To the best of our knowledge, an end-to-end deep learning strategy was first put out in 2015 (Huval et al., 2015). Since then, a number of alternative model architectures have been investigated in an effort to overcome the shortcomings of the conventional methods. Following 2015, a plethora of CNN-based models were proposed, including those that combined Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) (Zou et al., 2019), Spatial Convolutionals (SCNNs) (Pan et al., 2018), convolutional architectures combined with Polynomial Regression Neural models (L. Chen et al., 2021; Tabelini, Berriel, Pasaxão, et al., 2021), and Transformer-based architectures that utilize various attention-based mechanisms (Vaswani et al., 2017) to refine the extracted semantic feature maps from the road image scenes (R. Liu et al., 2021). All of these techniques do, however, have certain drawbacks and restrictions. For example, the CNN's local kernel filters prevent it from extracting global information from images (Strudel et al., 2021), and training a Transformer network requires a significant amount of computer power.

Furthermore, the prediction ability of a lane identification neural network for various road environment circumstances, such as extreme temperatures or blocked lanes, is just as significant as real-time inference. However, traditional machine learning methods have advantages in terms of interpretability and computational efficiency, even though they are not as effective in feature extraction. The drawbacks of using deep learning and conventional machine learning separately may be solved by combining their respective advantages. Apart from the architectural limitations specific to each neural model, the absence of data variance on the available datasets is a significant problem that complicates road lane recognition. A sufficient balance of different conditions and diverse road scenes is lacking in the majority of these widely used datasets, such as TUSimple (Tomatosliu et al., 2020), CULane (Pan et al., 2018), KITTI (Geiger et al., 2012), and others, according to comparative studies (Hafiz et al., 2021; Kumar et al., 2022). This result emphasizes how crucial it is to thoroughly assess a suggested deep learning model for lane detection across several datasets, as strong performance in one dataset does not always imply sufficient and trustworthy performance across the board. This paper presents a hybrid strategy to improve the resilience and accuracy of road lane recognition by integrating machine learning and deep learning approaches. To extract high-dimensional characteristics from input photos and capture fine-grained aspects of the road environment, the system uses a CNN. To balance computing economy and detection accuracy, these characteristics are processed by machine learning algorithms for lane border categorization and tracking. By exploring the potential integration of two distinct models—a CNN network serving as the backbone and a Vision Transformer (ViT) (Dosovitskiy et al., 2021)—to detect road lanes, this research endeavour aims to advance the area. By feeding the extracted feature maps into a ViT model, we hope to improve the quality of the retrieved feature maps and maximize the CNN backbone network's feature extraction capabilities. We shall endeavour to leverage the benefits of both architectures in this manner, striving for enhanced and dependable overall performance. To the best of our knowledge, how well a ViT can be used to enhance CNN-extracted features in the feature maps of the backbone and improve lane recognition performance. The suggested hybrid model is tested using a wide range of datasets that cover different kinds of roads, climates, and lighting situations. Experiments show that our method works better in real-time and detection accuracy compared to current state-of-the-art techniques. The aim is to aid in developing safer and more dependable autonomous driving technology by incorporating this strong lane identification system into traffic control systems. The increased capacity for lane recognition may greatly optimize vehicle navigation, lower the number of traffic accidents, and promote smooth traffic movement in metropolitan areas.

The following four tasks—perception and localization, high-level path planning, behaviour negotiation, and intelligent mobility control—are often the main uses of the data from these sensors. For the sake of safety, these four high-level jobs also require supervision. The general architecture of an autonomous vehicle's perception, planning, and control process is depicted in Figure 1.

Figure 1. Architecture of a perception, planning, and control workflow in autonomous vehicles

Trajectory or path planning is the next step once sensing and localization are successfully completed, which allows the vehicle to be driven through traffic. Path planning is the most crucial and difficult activity, and it will affect the decision-making process. The car will use the sensed data to attempt to comprehend the specific driving situation, such as a right turn or junction, the conditions and actions of other cars up ahead, different road signs, avoiding collisions, etc. The car will use its perception of the environment to learn and plot every conceivable path. It will then use machine learning models or state models to draw conclusions that will help it navigate the road. The vehicle's motion control is the penultimate stage of the procedure. Taking into account the dynamics of the vehicle, the motion control system affects both the vehicle's lateral and longitudinal movement. It involves controlling the car's braking, steering, and cruise control to ensure that it stays on the intended route while driving safely.

The most crucial tasks for sensing the dynamic traffic environment are perception and localization, which make use of a variety of vehicle sensors. The many approaches for road detecting are shown in Fig 2.

Some of the sensors used are discussed as follows:

- Mono cameras have limitations such as being extremely light-sensitive and performing poorly in low-light conditions like fog and rain. Nevertheless, they can be used for obstacle detection and classification; they provide an affordable solution and are good for two-dimensional mapping and lane detection. Furthermore, it is quite challenging to comprehend distance estimate with these kinds of cameras.

- The same features as mono cameras are also available with stereo-vision cameras, such as depth estimation and three-dimensional mapping; however, these cameras are more expensive to compute; furthermore, velocity and distance estimation cannot be estimated; and like mono cameras, they are sensitive to light and perform poorly in difficult lighting conditions.

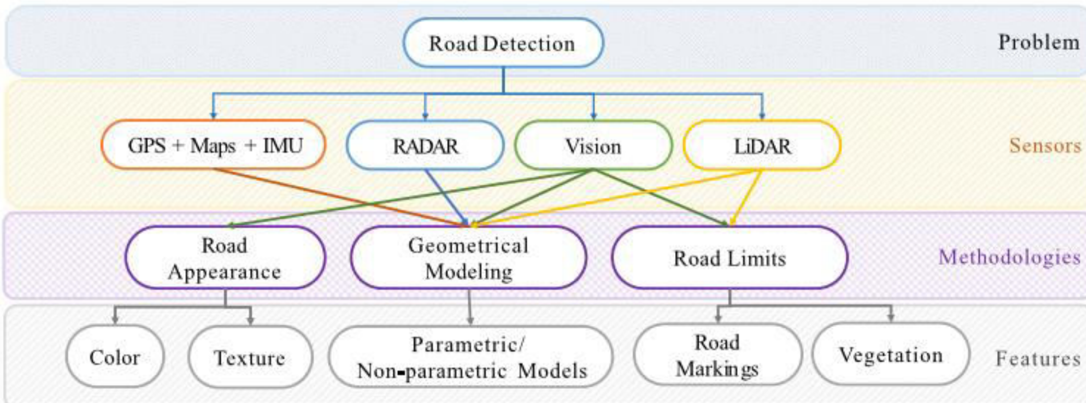


Figure 2. Various approaches to road detection

- LiDAR is also used for obstacle detection, accurate 3D mapping of the driving environment and scenario using multi-layer LiDAR, direct distance estimation, effectiveness in light weather conditions, etc. However, object classification is a difficult task because of reflective surfaces and generally bad weather.
- RADAR can be used for obstacle detection; it also provides velocity information; long- and short-range options are available; it detects well in poor weather conditions but performs poorly in terms of classification, static object detection, angular rotation, and interference due to multiple reflective surfaces.
- Other sensors, like IMUs, GPS, GIS, are also used for estimating the various inertial measurements and real-time positioning of the vehicle on the road.

Since there isn't a single, effective method for providing good sensing and perception functionality, a variety of these technologies are combined to provide correct perception. The perceived information from several sensors is combined to provide an accurate picture of the driving environment. Methodologies including odometry, particle filters, Kalman filters, and simultaneous localization and mapping (SLAM) approaches are used to assess the state of the vehicle in a driving situation in order to locate it independently. The entire process of path planning, motion control of the vehicle, sensing, perception, and localization is shown graphically in Figure 3. Table 1 lists some methods for detecting roads.

Trajectory or path planning is the next step once sensing and localization are successfully completed, which allows the vehicle to be driven through traffic. Path planning is the most crucial and difficult activity, and it will affect the decision-making process. The car will use the sensed data to attempt to comprehend the specific driving situation, such as a right turn or junction, the conditions and actions of other cars up ahead, different road signs, avoiding collisions, etc. The car will use its perception of the environment to learn and plot every conceivable path. It will then use machine learning models or state models to draw conclusions that will help it navigate the road. The vehicle's motion control is the penultimate stage of the procedure. Taking into account the dynamics of the vehicle, the motion control system affects both the vehicle's lateral and longitudinal movement. It involves controlling the car's braking, steering, and cruise control to ensure that it stays on the intended route while driving safely.

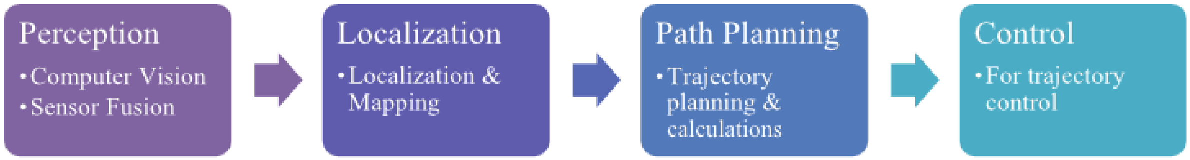


Figure 3. Stages of the autonomous driving system

Table 1. Different Road detection Methodologies

Methodology	Sensors	Features
Road appearance	Vision	Colors, textures
Road limits	Vision, LiDAR	Markings, lanes
Geometrical modeling	Maps, GPS, LiDAR, RADAR, vision	Parametric and non-parametric models

Literature Review

Road lane recognition has advanced significantly with the introduction of classic machine learning techniques and, more recently, deep learning approaches. The present study examines the development of lane identification techniques, emphasizing the advantages and disadvantages of conventional and deep learning-based approaches. Additionally, it establishes the framework for the hybrid strategy that is being suggested. The CNN, a deep learning model architecture based on the mathematical linear process of convolution, has had a significant impact on image data analysis.

Hand-crafted features and traditional image processing techniques are the mainstays of traditional lane detection systems. These techniques include clustering algorithms like K-means and RANSAC, edge detection, and the Hough Transform. In order to determine lane borders, early works—like those by Canny (1986) and Hough (1962)—focused on distinguishing edges and lines in pictures. Even though these methods are computationally effective, they frequently have trouble with intricate road conditions, uneven illumination, and occlusions. Other techniques, such as those created by Aly (2008), performed better in controlled settings and classified lane boundaries by using Random Forests and Support Vector Machines (SVM). However, because these methods relied on specified characteristics, which could not generalize well across diverse datasets, they still encountered difficulties in practical settings.

Lane detection has gone much farther since deep learning, and Convolutional Neural Networks (CNNs) in particular, were introduced. From raw picture data, CNNs automatically extract hierarchical features, enabling robust performance in challenging circumstances. Research has demonstrated great performance in image segmentation tasks, including lane detection. Examples of these works include He et al. (2016) using the ResNet architecture and Ronneberger et al. (2015) with the U-Net design. Modern deep learning models, such as Pan et al. (2018)'s SCNN, use spatial CNNs to identify long-range relationships in road imagery and achieve excellent lane recognition accuracy. However, huge labeled datasets and substantial computer resources are often needed for training deep learning models. Image data are often stored as a fixed uniform grid of parameters that represent each pixel and its corresponding color in order to do image analysis using deep learning. The method by which the connections between the neurons of subsequent layers occur distinguishes a CNN network from an Artificial Neural Network (ANN) with fully-connected layers in a revolutionary way. Through the use of convolution, a CNN network creates local connections rather than simply joining the output neurons of each layer to the subsequent layer. In this manner, there are much less created network parameters (Albawi et al., 2017). Three different types of layers make up a CNN network: the pooling layers, the convolutional layer, and the non-linear layers, which are activation functions like sigmoid, tanh, and Rectified Linear Unit (ReLu). As feature extractors, the convolutional layers use various regional matrix kernels to the input picture in order to locate and extract important characteristics. The network may extract various information from the picture by using distinct convolutional kernels for each convolutional layer, which function as filters.

The non-linear layer in most CNN designs comes after the convolutional layer and uses an activation function of some sort to modify the output of the previous layer. The pooling layer is yet another crucial operational element of a CNN network. Its primary contribution is to reduce the complexity of the CNN architecture by maintaining high-value features when down sampling the feature map.

Detecting the free (drivable) road has been an area of focus over the past few decades. It is one of the primary tasks involved in detecting, perceiving, and localizing the present driving environment. This visual perception is used to identify areas in the driving environment devoid of collisions, which helps advanced driving assistance systems make decisions independently. One of the key computer vision methods utilized in autonomous driving is road scene segmentation. For collision-free navigation, extracting or segmenting the drivable region from the collected road picture is crucial. Typical driving scenarios may include buildings, cars, roadways, people, etc. Estimating the length of the road, the different lanes and their junctions, splits, and termination locations in the various driving situations are all included in road detection. A linked road surface free of people and other barriers is referred to as a drivable zone. Imposing geometrical limitations on the several elements that make up the driving picture is the aim of road segmentation. In essence, road segmentation makes it possible to create an occupancy map of the observed driving environment, which is then used by autonomous driving systems to facilitate safe navigation. Therefore, it becomes crucial to precisely and effectively divide the area of drivable road from the driving environment. Road segmentation is often accomplished through the use of several computer vision algorithms that make use of techniques like edge detection and histograms. Driving assistance systems may employ comparable information to safely traverse the driving environment. Color, texture, limits, and lane markings are the essential indicators that help people perceive information about the road. The development of improved driver assistance systems has made extensive use of vision-based perception. Various machine learning algorithms are being combined with this technology to provide a proof of concept for autonomous vehicles that meet the SAE stage 2 to stage 3 autonomy requirements. However, this is extremely difficult to perform since road conditions and design differ globally and are not always the same, therefore these computer vision algorithms won't provide results that are consistently the same everywhere. The use of convolutional neural networks (CNNs) in autonomous driving has increased in recent years due to the development of full CNNs for semantic segmentation. Recent developments in the creation of massive or deep CNNs, such as SegNet, will help the driving assistance system handle a variety of driving scenarios. Deep CNNs have been employed by several researchers to partition the driving scene semantically. A DCNN has been reported for detecting obstacles and segmenting roads. The study suggests creating a disparity map for obstacle identification in a driving environment using a stereo-based method. A weighted mixture of the different features has been employed for road recognition in the two networks, LaneNet and ENet, that have been suggested to identify road features. An exact and accurate representation of the driveable road is obtained by combining the output from two CNNs: one for road surface recognition and another for lane detection. A centerline extraction and road detecting deep recurrent convolutional neural network (U-Net). The work entails creating a unique RNN unit that is integrated into the U-Net architecture for road extraction. Next, a multi-task learning scheme is developed to handle road detection and centerline extraction at the same time. Road detection has been accomplished with ResNet-101. Road and road boundary network (RBNet), a deep neural network (NN) that reduces the chance of a pixel being incorrectly classified as a road or road border, is designed for unified road and road boundary identification concurrently. In order to quickly and accurately segment roads, a CNN with gated recurrent units has been suggested. This solution addresses the issue of complicated computation that arises when fusing pixels for road segmentation using the more traditional extremely deep encoder-decoder structure. The division of unmarked roads has been suggested to be done using a CNN with colored lines. The work segments the road from the backdrop using a graphical model based on a conditional random field, which is created using a score-based approach. Moreover, deep learning-based multi-modal systems for autonomous vehicles have been developed recently with various sensor fusion technologies. The network takes a multi-layer feature as input, solves the sequential regression problem, and generates an output of similar width as the input. The network consists of three sections: a CNN-based local feature encoder, an LSTM-based feature processor, and a CNN-based output decoder.

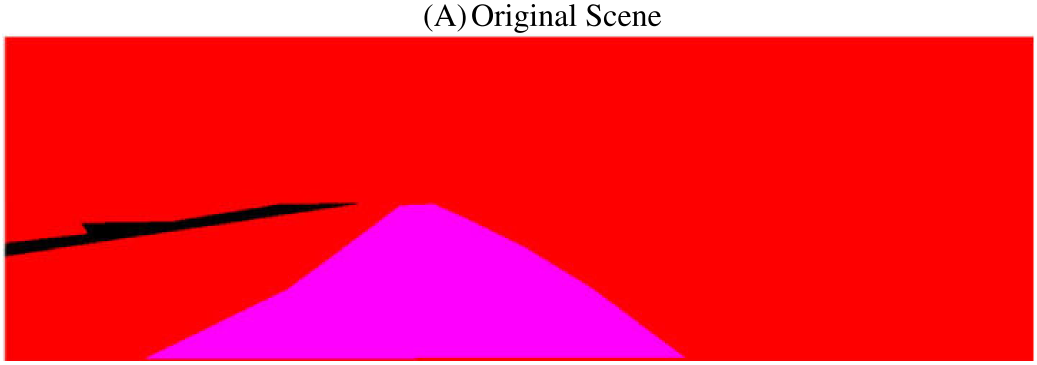
A 3D object detection system has been developed by fusing the data sensed from the RGB camera and LiDAR point cloud. The deep multi-modal detection and classification methodologies sense and fuse data from multiple sensing mechanisms, like mono and stereo vision, LiDAR, RADAR, GPS, and IMU, to generate complex features. The work predicts 3D bounding boxes using the fused information. The network is made up of two subnetworks: one for Multiview feature fusion and the other for 3D object identification. Similar research has been published, in which 3D object identification is achieved by fusing camera data with LiDAR point clouds. The utilization of RGB, far-, middle-, and near-infrared camera photos in combination with other multispectral photographs has also been the subject of some study, which has focused on perceiving multilateral information about the driving scene and depth perception.

Proposed System

Through the integration of deep learning and conventional machine learning approaches, the proposed system seeks to improve road lane recognition. Suitable for intelligent traffic systems, this hybrid technique combines the best features of both paradigms to deliver reliable, precise, and real-time lane recognition. Using the CNN-extracted features, traditional machine learning methods are used to monitor and identify lane borders. For every recognized lane marking, a feature vector is created using the CNN-generated feature maps. These vectors have characteristics like curvature, direction, and location. The feature vectors are divided into groups based on lane boundaries and non-lane boundaries using machine learning techniques like Support Vector Machines (SVM) and Random Forests. A tracking algorithm, such a particle filter or Kalman filter, is used to guarantee reliable lane detection over time. This procedure aids in preserving lane continuity even in the event of transient lane marker losses or occlusions. The lane detection dataset is used to refine a CNN model that has already been trained, such as ResNet or VGGNet. These models have an impressive track record of successfully extracting features. When creating feature maps from input photos, the CNN extracts pertinent information such as road borders, lane markings, and other pertinent facts. Lane markings are emphasized in a segmented image that is produced using the output feature maps. According to Pavlidis, segmentation is a pixel classification technique that divides an input image into subgroups by classifying each pixel. To distinguish bright objects from dark backgrounds or vice versa, we categorize the pixels into dark and light classes while segmenting a picture by thresholding its gray level. According to reports in the literature, deep learning models that are enhanced with stacked layers (depth) allow users to obtain extremely accurate and high-quality outcomes. These models may use the most unstructured data.In the context of autonomous driving, semantic segmentation holds great promise for the advancement of visual perception systems. Road and lane detection systems are only two examples of the driving assistance systems that may be developed using the pictures from the many cameras that are present. The figure illustrates the segmentation procedure for roads.

Figure 4 shows an example of road segmentation process. Figure 4A displays the image of the road taken by a camera installed on the vehicle, and Figure 4B displays the segmented image with three classes: (a) The surroundings are depicted in red, (b) the driveable right road is depicted in magenta, and (c) the non-drivable left road is depicted in black. Deep neural networks have shown to be advantageous for semantic segmentation in a wide range of applications, including autonomous vehicles and medical imaging. Thus, the current work has investigated the application of deep learning models for the segmentation of the driveable road.

(A) Original Scene



(B) Segmented Image

Figure 4. Image of the driving environment captured by a camera in the KITTI dataset. (a) Original scene. (b)Segmented image

Figure 5 illustrates the trade-off between object sensitivity and distance sensitivity in conventional vehicle systems. When an item is close, its sensitivity is strong, making classification easier; when the object is farther away, the sensitivity decreases, which might result in subpar classification. The existing methods would demand too much computer resources to handle excellent distance and object sensitivity. We may identify objects and their distances by determining the nature and location of the ground region in a picture. Additionally, understanding the drivable zone in a driving scenario or environment is crucial for autonomous cars. Using the KITTI road dataset, the suggested system in this study seeks to recognize and segment the road area. This will be useful for applications like navigation systems and autonomous driving. For the sake of this definition, "ground" refers to a surface that is comparatively smooth, driveable, and immediately identifiable from its surroundings. It might be made up of typical flaws or abnormalities or different lighting situations.

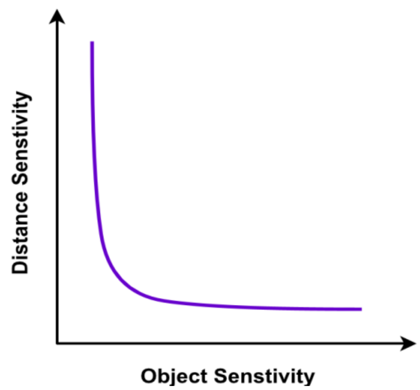
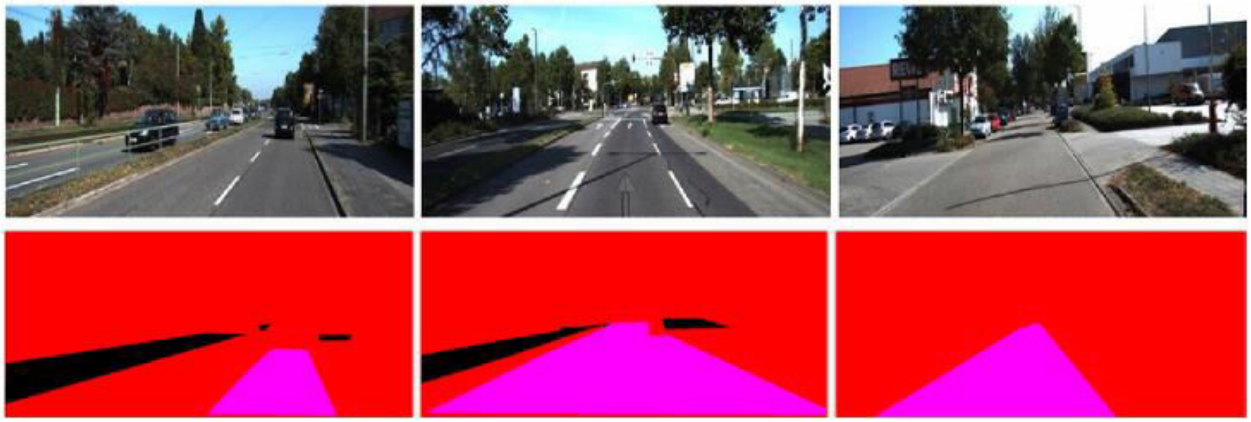


Figure 5. Plot between the object sensitivity and the distance sensitivity

Methods

For object and road/lane detection, there is a dataset called the KITTI Vision Benchmark Suite. There are 289 training and 290 testing photos in the road/lane dataset. The dimension of each image is 372 × 1242 pixels. Five distinct days were used to take all of the photos. As can be seen, the information is further separated into three groups of road sceneries as can be seen in Fig 6.

- *Single-lane, designated roads that are urban marked (UM)*
  - Comprising 96 testing and 95 training photos.
- *A single-lane, unmarked urban unmarked (UU) road*
  - Comprising 100 testing and 98 training photos.
- *A multi-lane, urban multiple-marked (UMM) road with markings*
  - Comprising 94 testing and 96 training photos.



(a) UM (b) UMM (c) UU

Figure 6. The three different road scene categories with three RGB color codes. a UM. b UMM. c UU

Two sets of label pictures, color-coded with areas of interest, are included in the input photos. It identifies every route in one group and only the lane the automobile is traveling in the other set. The collection that corresponds to every road surface has been employed in the current investigation. The road is colored magenta, non-road sections are colored red, and left road surfaces are colored black in these labels, which are RGB pictures. Before being fed, the dataset has undergone pre-processing and augmentation in accordance with the network's input layers. The data set has been scaled to 224 × 224 pixels for ResNet50 and MobileNet-V2, and to 299 × 299 pixels for Xception.

The architecture and weights of the three pretrained networks—two high accuracy models of ResNet50 and XceptionNet, and one small DNN of MobileNet-V2—were initialized for the present work, which is based on Google's DeepLabV3+ semantic segmentation model. This is usually done for edge deployment. All of these networks and models are discussed as follows:

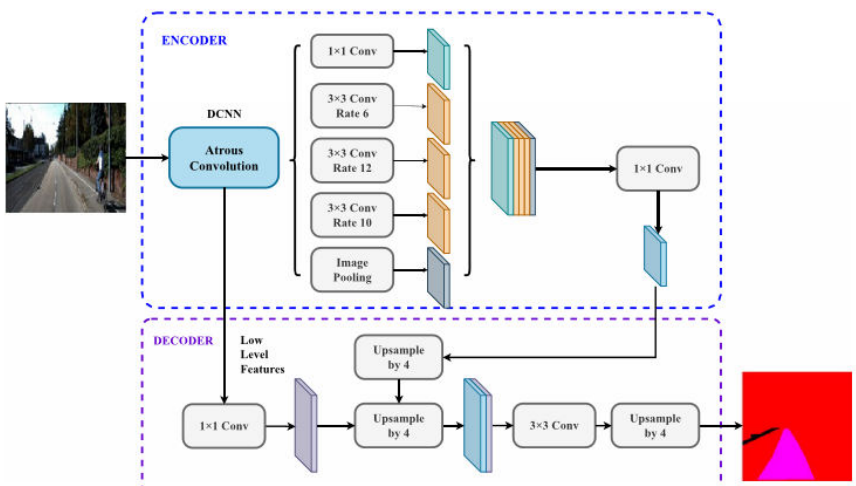


Figure 7. Encoder-decoder of DeepLabV3+ architecture

The state-of-the-art DeepLabV3+ is the foundation of the RoadSegNet. Google created DeepLab, an open-source semantic segmentation model. It functions by incorporating a basic decoder module that aids in object segmentation along boundaries and enhances the segmentation outcomes. Depth-wise separable convolution is used for both the Decoder Module and Atrous Spatial Pyramid Pooling to get faster results. The transfer learning approach was used to initialize the weights. Three cutting-edge DNNs have been employed. The paper examines the usage of one small DNN, MobileNet-V2, and two high-accuracy models, ResNet50 and XceptionNet. In addition to the above modifications, DeepLabV3+'s primary feature extractor is an aligned Xception network.

a) Striding and depth-wise separable convolution take the role of the max pool layers.

b) After every 3 x 3 depth-wise convolution, more batch normalization and ReLU activation are applied.

c) The model's depth is enhanced without affecting the topology of the entrance flow network.

The encoder operates using an output stride, which is the proportion between the size of the initial picture and the finished encoded features. The encoded features are initially unsampled with a factor of 4 and concatenated with equivalent low-level features from the encoder module with the same spatial dimensions, as opposed to employing bilinear up-sampling with a factor of 16. Prior to concatenating on the low-level features, 1 × 1 convolution is used in order to decrease the number of channels. Following concatenation, a few 3 × 3 convolutions are performed, and a factor of 4 is used to unsample the features. As a result, the output and picture have the same size. Semantics of the DeepLabv3+-based RoadSegNet architecture that is being proposed are shown in Fig 8 as below.

Results and Discussions

Evaluation metrics

The metrics (a) global accuracy, (b) mean accuracy, (c) mean IoU, (d) weighted IoU, and (e) mean BF score have been utilized to assess the effectiveness of the acquired segmentation findings. The following words are used to describe these assessment metrics:

- *False negative (FN)*: pixels that belong to lesions but are incorrectly classified as background;
- *False positive (FP)*: pixels that belong to the background but are incorrectly classified as lesions; and
- *True positive (TP)*: pixels that belong to lesions and are correctly classified as lesions.
- *True negative (TN)*: Pixels that are appropriately classified as background and belong such.

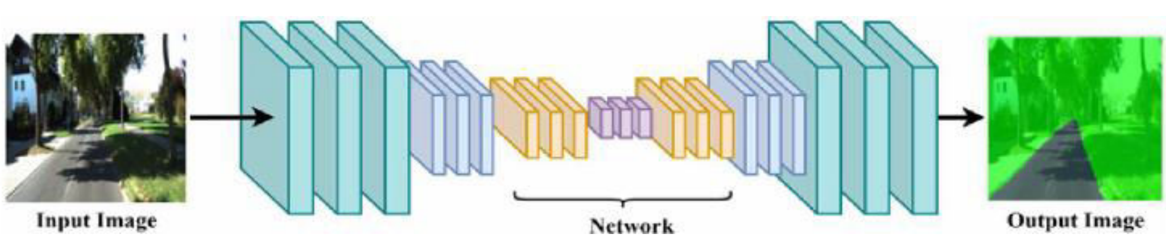


Figure 8. The Road Lane Detection Segment (RoadSegNet) architecture

Accuracy

It can be computed globally for all classes or individually for each class. Equation 1 provides the accuracy, which is the percentage of accurately classified pixels in each class.

$$\text{Accuracy} = \frac{\left(\frac{\text{TP}}{\text{TP+FN}}\right) + \left(\frac{\text{TN}}{\text{TN+FP}}\right)}{2} \quad (1)$$

Global accuracy

The global accuracy, which can be found in Eq. 2, is the ratio of properly classified pixels to all pixels.

$$\boldsymbol{Global\ Accuracy = \frac{TP+TN}{TP+TN+FP+FN}} \quad (2)$$

Mean accuracy

The mean accuracy is the ratio of the total of the accuracy of each class to the number of classes.

Intersection over Union (IoU)

It estimates the inaccurate classifcation of the pixels and is provided in Eq. 3.

$$\boldsymbol{IoU = \frac{Lesions + Background}{2}} \quad (3)$$

where

$$\boldsymbol{Lesions = \frac{TP}{TP+FP+FN}} \text{and} \quad \boldsymbol{Background = \frac{TN}{TN+FP+FN}}$$

Weighted IoU

If there is an unequal correlation between the class sizes in the pictures, the weighted IoU is employed to reduce the impact of incorrect classification in smaller classes. It is given in the equation as follows:

WeightedIoU = (LesionWeight ∗ Lesion) + (BackgroundWeight ∗ background)

where

$$\text{Lesion Weight} = \frac{\text{No.of Pixels belonging to Lesion}}{\text{Total No.of Pixels}}$$

$$\text{Background Weight} = \frac{\text{No.of Pixels belonging to Background}}{\text{Total No.of Pixels}}$$

BF score

The alignment between the projected boundaries and the gold standard boundary is computed. According to Eq. 4, it is provided by the harmonic mean of recall and accuracy as follows:

$$\text{BF Score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

where

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad \text{and} \quad \text{Precesion} = \frac{\text{TP}}{\text{TP+FP}}$$

Training performance

The KITTI Road/Lane Detection Evaluation Dataset 2013 has been taken into consideration in this suggested effort. Preprocessing of the dataset to satisfy the needs of ResNet50, Xception, and MobileNet-V2 deep neural networks has been done in order to make it compatible with the suggested architecture of RoadSegNet. The KITTI Road dataset's LabelIDs and ColorMaps have changed as a result of the different class labels being redefined as the environment, the left road, and the right road. All three networks have been trained using the algorithm-specific learning option of stochastic gradient descent with momentum (sgdm). For each network, the maximum number of epochs has been set at 100 and the starting learning rate has been set at 0.001. The remaining settings are left unchanged, and the mini-batch sizes are adjusted based on the GPU specifications. All of the models were trained using the MATLAB 2020b environment on a Windows 10 computer equipped with an Nvidia 2060 super 8 GB GPU, a Ryzen 9 12 Core CPU, and 16 GB of RAM. Figures represent the training loss, training accuracy, and base learning rate plots for each of the three networks—ResNet50, XceptionNet, and MobileNet-V2—considered for the RoadSegNet. Plots show that the training loss function reduces once all of these networks reach good training accuracy of around ~ 96 to 97%.

Segmentation results

The RoadSegNet is given different driving scenario photos from the KITTI Road Eval Dataset once it has been trained. The MobileNet-V2, XceptionNet, and ResNet50 segmented pictures. The first six images in each table show the best segmentation results, while the last three images show the segmentation results for harsh driving scenarios in heavily shadowed regions. The tables also show the plot for intersections over the union (IOU) between the segmented image and the ground truth image. The IOU plots in each table show that the RoadSegNet can detect the drivable road in each driving scenario. In Tables 3 and 4, every evaluation parameter has been tabulated. Table 3 presents a comparison of the different performance indicators for each built network, including global and mean accuracy, mean and weighted IOU, and mean BF score for the full training and testing datasets. Table 4 provides details on each class's accuracy, IOU, and mean BF score—that is, the precision with which a given class has been identified throughout all training, testing, and validation datasets for every network that has been created. The radar plots for all performance parameters are displayed for each network in Figures 12, 13, and 14. It is evident from the data obtained in Tables 2, 3, and 4 that the created networks provide extremely excellent accuracy; the mean BF score varies between about 0.75 and 0.83, the weighted IOU also spans between approximately 92 and 97%, and the global accuracy ranges between approximately 96 and 97%. The findings obtained further show that, in spite of its small size and deep neural network design, the MobileNet-V2 performs virtually as well as the XceptionNet and, in certain situations, even better than the ResNet50.

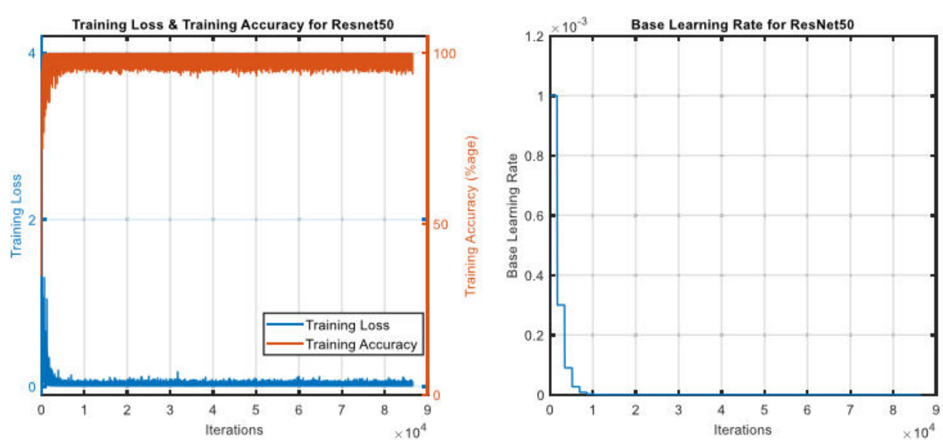


Figure 9. Plot for training loss, accuracy, and base learning rate for ResNet50

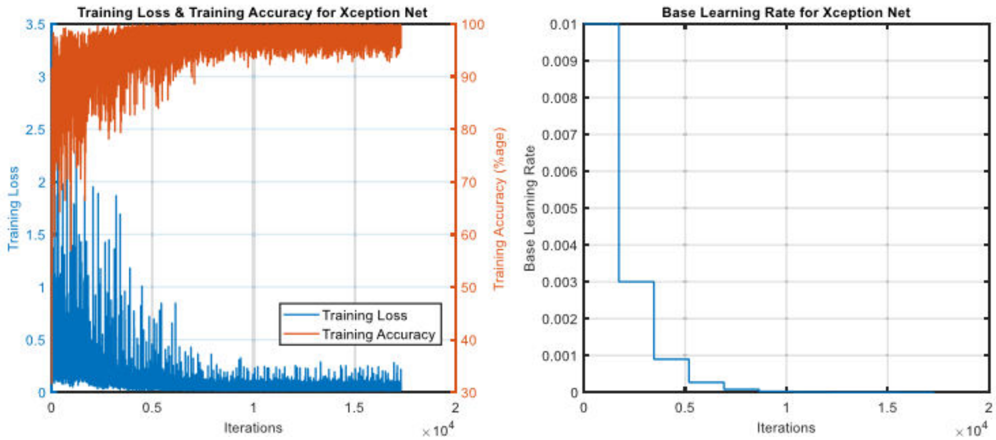


Figure 10. Plot for training loss, accuracy, and base learning rate for XceptionNet

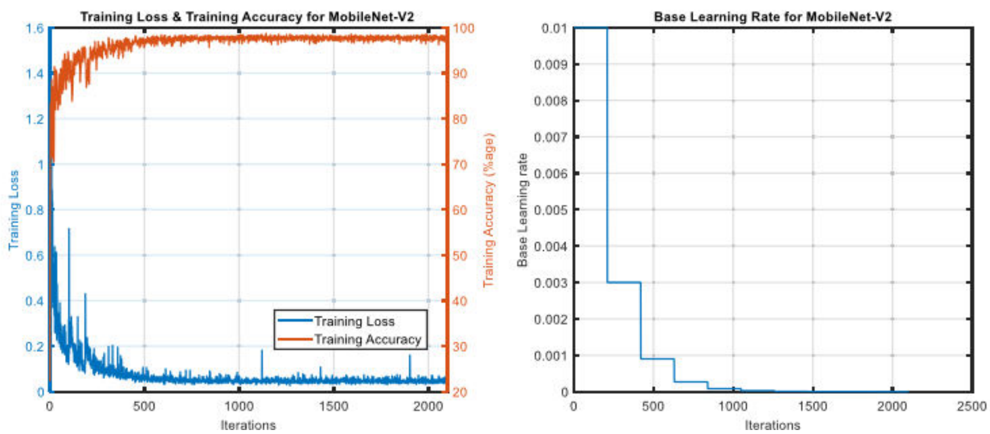
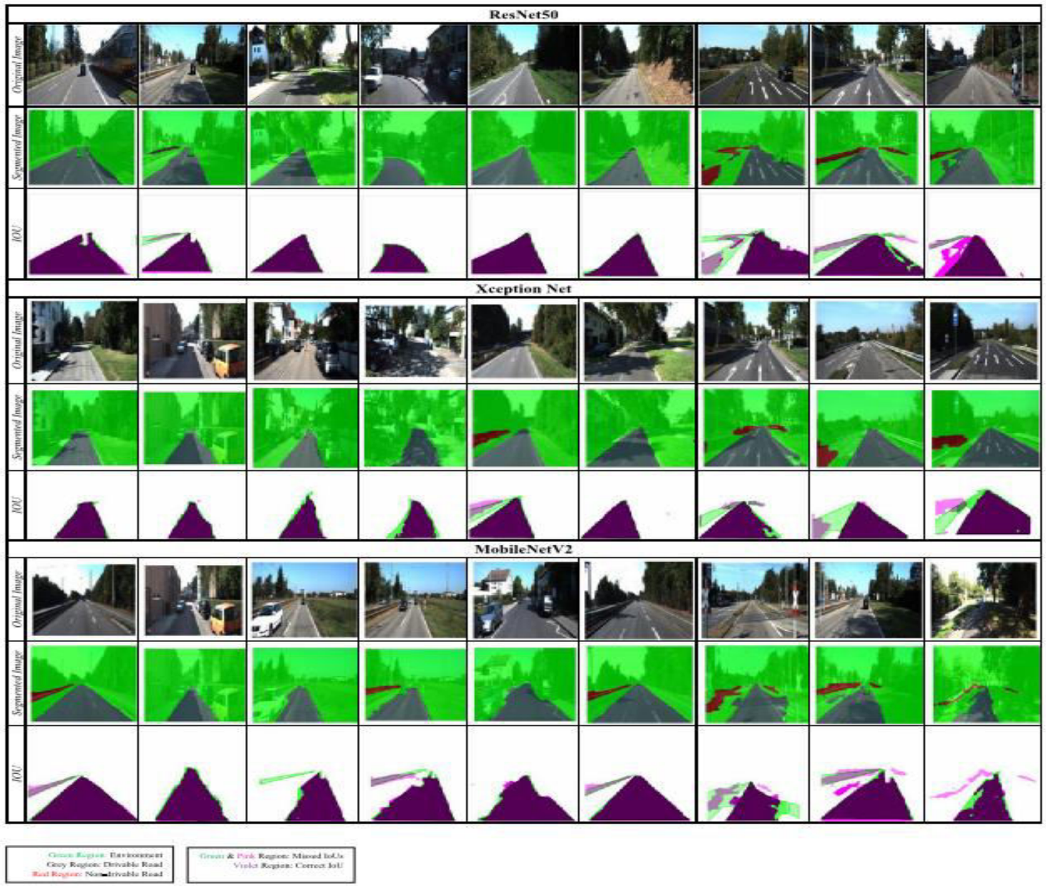


Figure 11. Plot for training loss, accuracy, and base learning rate for MobileNet-V2

Table 2. The segmented road and environment regions and their respective IOUs obtained using ResNet50, XceptionNet, and MobileNet-V2



Discussion

The four steps of any autonomous driving system are perception, localization, path planning, and control. Perception tasks are the main focus of the current study. The goal of the study described in this paper is to develop a ground detection system based on deep learning.

The findings shown in the "Segmentation results" section demonstrate the system's resilience by identifying a sizable portion of the road, even in areas with little lighting. Due to fewer photos being tagged for the left road sections, they are not as well spotted. Utilizing a dataset with more of these photos can help with this. The optimized framework functions optimally with bright images. The study investigates the use of three distinct pretrained networks, namely MobileNet-V2 (a small DNN) and ResNet50 and XceptionNet (high accuracy models), usually for edge deployment. It is evident that MobileNet-V2's accuracy is comparable to that of ResNet50 and XceptionNet, two high-accuracy models. The suggested model may be used to effective road detecting tasks with the addition of features including lane detection, depth estimation, and intersection detection. Autonomous cars still face a difficulty even though the model works well in daytime situations; its performance in night-time scenarios has not been evaluated. The research presents a comparative analysis of the state-of-the-art DNNs, namely ResNet50, XceptionNet, and MobileNet-V2. Table 2 qualitatively demonstrates that the trained models' IOU performs exceptionally well in both highly complicated shaded situations and well-lit roadways. Tables 3 and 4 provide quantitative evidence for this finding.

Table 3. Compared dataset performance metrics for each network

S. no.	Metric	Network		
		ResNet50	XceptionNet	MobileNet-V2
Training dataset				
1	Global accuracy	97.52	97.34	97.80
2	Mean accuracy	89.83	93.80	98.70
3	Mean IoU	81.49	77.02	80.32
4	Weighted IoU	95.34	95.45	96.26
5	Mean BF score	0.8139	0.8008	0.8386
Testing dataset				
1	Global accuracy	95.79	96.39	96.35
2	Mean accuracy	73.99	82.20	87.03
3	Mean IoU	69.02	72.99	75.28
4	Weighted IoU	92.15	93.63	93.59
5	Mean BF score	0.7572	0.7669	0.7885
Validation dataset				
1	Global accuracy	96.62	97.14	97.13
2	Mean accuracy	86.92	94.25	95.67
3	Mean IoU	76.24	74.32	74.33
4	Weighted IoU	93.67	94.92	94.99
5	Mean BF score	0.7939	0.7979	0.8153

The multiple state-of-the-art DNNs of ResNet50, XceptionNet, and MobileNet-V2 have been compared in the research. Table 2 qualitatively demonstrates that the trained models' IOU performs exceptionally well in both highly complicated shaded situations and well-lit roadways. Tables 3 and 4 provide quantitative evidence for this fact. In Table 3 After analyzing the segmentation's global accuracy metrics, it was found that the models provided an accuracy of more than 97% for the training dataset and more than 96% for the test database. In addition, all three DNN models have been assessed for the additional metrics of mean accuracy, mean IOU, weighted IOU, and mean BF scores; these have been determined for both the training and testing datasets. Similarly, in Table 4, The ability of the three DNNs to precisely segment and categorize the several classes in the dataset—left road, right road, and environment—has been compared based on their class-wise accuracy. The effectiveness of the three DNNs has been assessed using the metrics of accuracy, IOU, and mean BF score. The assessment has been conducted on the training, testing, and validation datasets, and Table 4 shows that positive outcomes have been achieved. The "right road," or driveable portion of the dataset, shows that MobileNet-V3 has achieved an accuracy of nearly 99%, while the other two networks provide accuracy of roughly 91% and 97%. Comparably, MobileNetV3 achieves an accuracy of 97% for the environment, while other networks also provide an accuracy of greater than 97%. Similar to this, both the testing and validation datasets' performances have been assessed. When the current work is compared to other studies that have been published in the literature, Table 5 shows that the current study offers one of the best accuracies while using the least amount of runtime.

Table 4. Compared classification performance metrics for each network

S. no.	Metrics	Right road			Left road			Environment		
		ResNet50	XceptionNet	MobileNet-V3	ResNet50	XceptionNet	MobileNet-V3	ResNet50	XceptionNet	MobileNet-V3
Training dataset										
1	**Accuracy**	94.96	97.77	99.15	76.17	86.24	99.46	98.36	97.40	97.50
2	**IoU**	89.53	92.41	94.29	57.87	41.84	49.33	97.08	96.78	97.34
3	**Mean BF score**	0.8008	0.8093	0.8679	0.7185	0.6631	0.7256	0.8675	0.8443	0.8502
Testing dataset										
1	**Accuracy**	92.7	96.52	95.94	31.32	52.79	68.10	97.91	97.28	97.07
2	**IoU**	85.07	90.23	89.29	26.81	33.06	40.82	95.20	95.67	95.73
3	**Mean BF score**	0.7640	0.7794	0.8080	0.5115	0.5689	0.6331	0.8479	0.8279	0.8253
Validation dataset										
1	**Accuracy**	91.42	94.61	95.73	71.38	90.40	93.82	97.97	97.74	97.46
2	**IoU**	84.99	89.39	90.07	47.78	37.04	36.39	95.94	96.54	96.51
3	**Mean BF score**	0.7585	0.7917	0.8251	0.6823	0.6151	0.6774	0.8559	0.8524	0.8449

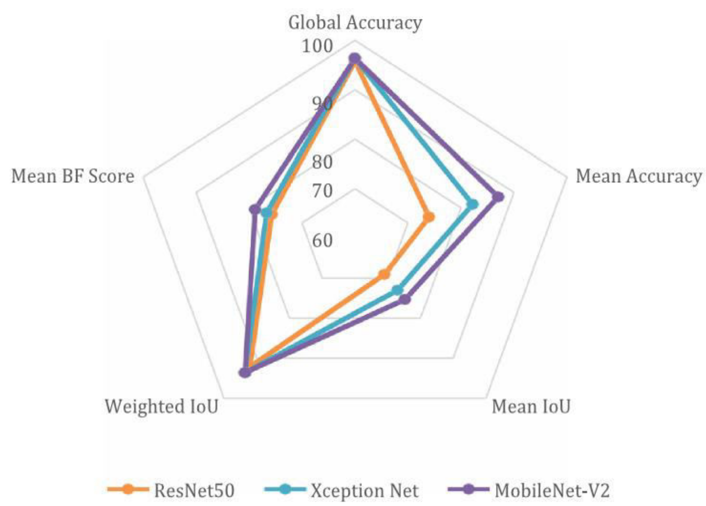


Figure 12. The radar plot for each network for all the performance metrics for the testing dataset

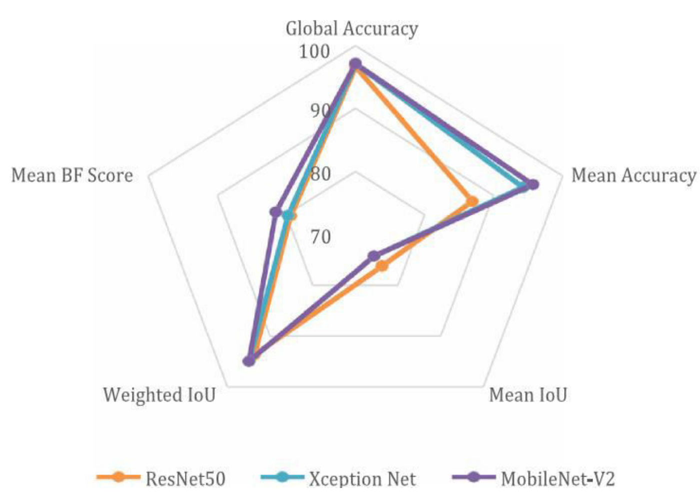


Figure 13. The radar plot for each network for all the performance metrics for the validation dataset

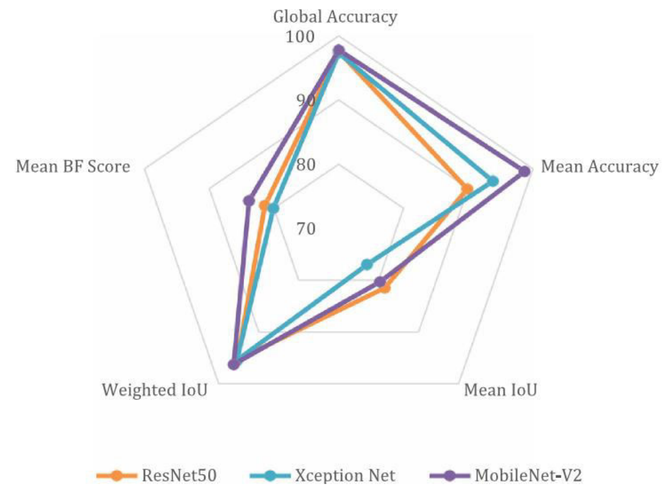


Figure 14. The radar plot for each network for all the performance metrics for the training dataset

Table 5 Overall accuracy comparison

Methodology		Accuracy	Runtime	Environment
DeepLabV3+ (current work)	ResNet50	97.52%	0.14 s	GPU @ 1.1 GHz (MATLAB)
	XceptionNet	97.34%	0.11 s	GPU @ 1.1 GHz (MATLAB)
	MobileNet-V2	**97.80%**	**0.07 s**	GPU @ 1.1 GHz (MATLAB)
PLARD [3]		97.27%	0.16 s	GPU @ 2.5 GHz (Python)
SNE-RoadSeg+ [4]		96.95%	0.25 s	GPU @ 2.5 GHz (Python)
USNet [5]		96.46%	0.02 s	GPU @ 1.5 GHz (Python)
DFM-RTFNet [6]		96.46%	0.08 s	GPU @ 2.5 GHz (Python)
SNE-RoadSeg [7]		96.42%	0.18 s	GPU @ 2.5 GHz (Python)
RBANet [8]		95.78%	0.16 s	GPU @ 1.5 GHz (Python + C/C++)
NIM-RTFNet [9]		95.71%	0.05 s	GPU @ 2.5 GHz (Python)
CLCFNet [10]		95.65%	0.02 s	GPU @ 1.5 GHz (Python)
LidCamNet [11]		95.62%	0.15 s	GPU @ 2.5 GHz (Python)

Conclusion

In this current study, it shows that the resilience and accuracy of lane detection systems may be greatly enhanced by fusing the feature extraction powers of Convolutional Neural Networks (CNNs) with the effectiveness and versatility of conventional machine learning techniques. An important step forward in creating intelligent traffic systems is using hybrid deep learning and machine learning techniques for road lane recognition. An automated road detecting system based on deep learning has been suggested in this work. The suggested framework is based on Google's DeepLab-V3+ architecture, a cutting-edge semantic segmentation network. Three image classification networks—ResNet-50, MobileNet-V2, and Xception—initialize the network's weights. The KITTI road dataset is used as a standard for evaluating the outcomes. The model achieves noteworthy results on the assessment criteria and is evaluated for inadequate circumstances and general ground difficulties. Furthermore, the suggested architecture performs well on MobileNet-V2, a tiny yet potent network that may be utilized for edge deployment and low-power devices. For road lane recognition, the suggested hybrid deep learning and machine learning technique is a viable way to improve the precision, resilience, and real-time functionality of intelligent traffic systems. Through a successful integration of the advantages of both paradigms, the system tackles the shortcomings of current approaches and paves the way for autonomous driving systems that are more dependable and safer. The topic of intelligent transportation and traffic management might be greatly advanced by more study and development in this area.

References

[1] Feng D, Haase-Schuetz C, Rosenbaum L, Hertlein H, Glaeser C, Timm F, Wiesbeck W, Dietmayer K (2020) Deep multimodal object detection and semantic segmentation for autonomous driving: datasets, methods, and challenges. In: IEEE Transactions on Intelligent Transportation Systems

[2] Cui Y, Chen R, Chu W, Chen L, Tian D, Li Y, Cao D (2021) Deep learning for image and point cloud fusion in autonomous driving: a review. In: IEEE Transactions on Intelligent Transportation Systems

[3] Chen Z, Zhang J, Tao D (2019) Progressive lidar adaptation for road detection. IEEE/CAA J Automat Sinica 6(3):693–702

[4] Wang H, Fan R, Cai P, Liu M (2021) SNE-RoadSeg+: rethinking depth-normal translation and deep supervision for freespace detection. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp 1140–1145

[5] Chang, Yicong, Feng Xue, Fei Sheng, Wenteng Liang, and Anlong Ming. “Fast road segmentation via uncertaintyaware symmetric network.” arXiv preprint arXiv:2203.04537 (2022).

[6] Wang H, Fan R, Sun Y, Liu M (2021) Dynamic fusion module evolves drivable area and road anomaly detection: a benchmark and algorithms. In: IEEE transactions on cybernetics

[7] Fan R, Wang H, Cai P, Liu M (2020) SNE-RoadSeg: incorporating surface normal information into semantic segmentation for accurate freespace detection. In: European Conference on Computer Vision. Springer, Cham, pp 340–356

[8] Sun J-Y, Kim S-W, Lee S-W, Kim Y-W, Ko S-J (2019) Reverse and boundary attention network for road segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, p 0

[9] Wang H, Fan R, Sun Y, Liu M (2020) Applying surface normal information in drivable area and road anomaly detection for ground mobile robots. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp 2706–2711

[10] Gu S, Yang J, Kong H (2021) A cascaded lidar-camera fusion network for road detection. In: In 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp 13308–13314

[11] Caltagirone L, Bellone M, Svensson L, Wahde M (2019) LIDAR–camera fusion for road detection using fully convolutional neural networks. Robot Autonom Syst 111:125–131

[12] Committee SAE (2014) Taxonomy and defnitions for terms related to on-road motor vehicle automated driving systems

[13] Dickmanns E, Graefe V (1988) Dynamic monocular machine vision. Machine Vision Appl 1:223–240

[14] EUREKA Network. Programme for a European trafc system with highest efficiency and unprecedented safety(PROMETHEUS), Brussels, Belgium. http://www.eurekanetwork.org/project/-/id/45
[15] Dickmanns ED (2007) Dynamic vision for perception and control of motion. Springer Science & Business Media
[16] Thorpe C, Herbert M, Kanade T, Shafer S (1991) Toward autonomous driving: the cmu navlab. i. perception. IEEE Expert 6(4):31–42
[17] Behringer R, Sundareswaran S, Gregory B, Elsley R, Addison B, Guthmiller W, Daily R, Bevly D (2004) The DARPA grand challenge- development of an autonomous vehicle. In: IEEE Intelligent Vehicles Symposium, 2004. IEEE, pp 226–231
[18] Thrun S, Montemerlo M, Dahlkamp H, Stavens D, Aron A, Diebel J, Fong P et al (2006) Stanley: the robot that won the DARPA Grand Challenge. J Field Robot 23(9):661–692
[19] Hillel AB, Lerner R, Levi D, Raz G (2014) Recent progress in road and lane detection: a survey. Machine Vision Appl 25(3):727–745
[20] Yoo H, Yang U, Sohn K (2013) Gradient-enhancing conversion for illumination-robust lane detection. IEEE Transact Intell Transport Syst 14(3):1083–1094
[21] Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440
[22] Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39(12):2481–2495
[23] Levi D, Garnett N, Fetaya E, Herzlyia I (2015) StixelNet: a deep convolutional network for obstacle detection and road segmentation. BMVC 1(2):4
[24] Almeida T, Lourenço B, Santos V (2020) Road detection based on simultaneous deep learning approaches. Robot Autonom Syst 133:103605
[25] Yang X, Li X, Ye Y, Lau RYK, Zhang X, Huang X (2019) Road detection and centerline extraction via deep recurrent convolutional neural network U-Net. IEEE Transact Geosci Remote Sens 57(9):7209–7220
[26] Munoz-Bulnes J, Fernandez C, Parra I, Fernández-Llorca D, Sotelo MA (2017) Deep fully convolutional networks with random data augmentation for enhanced generalization in road detection. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp 366–371
[27] Chen Z, Chen Z (2017) RBNet: a deep neural network for unifed road and road boundary detection. In: International Conference on Neural Information Processing. Springer, Cham, pp 677–687
[28] Lyu, Yecheng, and Xinming Huang. “Road segmentation using CNN with GRU.” arXiv preprint arXiv:1804.05164 (2018).
[29] Yadav S, Patra S, Arora C, Banerjee S (2017) Deep CNN with color lines model for unmarked road segmentation. In: 2017 IEEE International Conference on Image Processing (ICIP). IEEE, pp 585–589
[30] Lyu Y, Bai L, Huang X (2019) Road segmentation using cnn and distributed lstm. In: 2019 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, pp 1–5
[31] Chowdhuri S, Pankaj T, Zipser K (2019) MultiNet: multi-modal multi-task learning for autonomous driving. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp 1496–1504

[32] Ni J, Chen Y, Chen Y, Zhu J, Ali D, Cao W (2020) A survey on theories and applications for self-driving cars based on deep learning methods. Appl Sci 10(8):2749

[33] Chen X, Ma H, Wan J, Li B, Xia T (2017) Multi-view 3d object detection network for autonomous driving. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp 1907–1915

[34] Asvadi A, Garrote L, Premebida C, Peixoto P, Nunes UJ (2018) Multimodal vehicle detection: fusing 3D-LIDAR and color camera data. Pattern Recognit Lett 115:20–29

[35] Oh S-I, Kang H-B (2017) Object detection and classifcation by decision-level fusion for intelligent vehicle systems. Sensors 17(1):207

[36] Wang Z, Zhan W, Tomizuka M (2018) Fusing bird's eye view lidar point cloud and front view camera image for 3d object detection. In: 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp 1–6

[37] Kim T, Ghosh J (2016) Robust detection of non-motorized road users using deep learning on optical and LiDAR data. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp 271–276

[38] Sindagi VA, Zhou Y, Tuzel O (2019) MVX-Net: multimodal voxelnet for 3d object detection. In: 2019 International Conference on Robotics and Automation (ICRA). IEEE, pp 7276–7282

[39] Takumi K, Watanabe K, Ha Q, Tejero-De-Pablos A, Ushiku Y, Harada T (2017) Multispectral object detection for autonomous vehicles. In: Proceedings of the on Thematic Workshops of ACM Multimedia, vol 2017, pp 35–43

[40] Ha Q, Watanabe K, Karasawa T, Ushiku Y, Harada T (2017) "MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes," 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 5108–5115. https://doi.org/10.1109/IROS.2017.8206396

[41] Horowitz SL (1974) Picture segmentation by a directed split-and-merge procedure. In: IJCPR, pp 424–433

[42] Geiger A, Lenz P, Stiller C, Urtasun R (2013) Vision meets robotics: the KITTI dataset. Int J Robot Res 32(11):1231–1237

[43] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. IEEE Trans Pattern Anal Mach Intell 40(4):834–848.