Pal *et al.*
Journal of Engineering and Applied Science (2022) 69:110
https://doi.org/10.1186/s44147-022-00162-9

Journal of Engineering and Applied Science

RESEARCH Open Access

RoadSegNet: a deep learning framework for autonomous urban road detection

Kushagra Pal, Piyush Yadav and Nitish Katal*

*Correspondence: nitishkatal@gmail.com

School of Electronics, Indian Institute of Information Technology, Una, India

Abstract

Ground detection is an essential part of the perception system in self-driving cars. The ground can be imagined as a fairly smooth, drivable area that is even textured and easily distinguished from the surrounding area. It can have some common imperfections, like shadows and differing light intensities. In this paper, a comparative study of several deep neural network architectures has been reported that can deduce surface normal information on the classic KITTI road dataset in various challenging scenarios. Our goal is to simplify the task of how the recent methods perceive the ground-related information and propose a solution by testing it on three state-of-the-art deep learning models, which are "Resnet-50," "Xception," and "MobileNet-V2" to understand and exploit the capabilities of these models. The main significance of this comparative study has been to evaluate the performance of these networks for edge deployment. So, the tiny DNN model of MobileNet-V2 has been considered, which has approximately 80% fewer tunable parameters as compared to the others. The obtained results show that the proposed networks are able to achieve a segmentation accuracy of more than ~ 96% and that too in various challenging scenarios.

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

Introduction

Over the course of the last few decades, significant progress has been made in the field of autonomous vehicles, and DARPA has played a significant role in these developments [1]. The self-driving cars have been developed to use various onboard sensors like cameras, LiDARs, and GPS to collectively sense the dynamic environmental landscape and make the necessary decisions for safe navigation, and such systems are called *advanced driver assistance systems* (ADAS). Now, recent developments in the field of deep learning and multi-sensor fusion techniques have fostered the development of consumer-ready, safe, and efficient autonomous driving systems [2]. Technologies like multi-modal sensor fusion techniques and artificial intelligence are usually used collectively for the development of perception systems to sense the driving environment, predict the course of the traffic, plan the trajectory or lane assistance, and execute these decisions in the real world. It is desired that these intelligent perception systems be accurate, robust, and real-time. All of these will aid in the development of autonomous intelligent vehicle

systems and thus reduce road accidents, decongest the roads, and make commuting much more efficient and economical too.

The present work explores the development of a deep neural network architecture for detecting the drivable road regions in a driving scene. The proposed RoadSegNet uses Google's DeepLavV3+ at its core for the semantic segmentation of the road surfaces. The RoadSegNet typically uses weights from three different pretrained networks, namely the two high-accuracy models of ResNet50 and XceptionNet and one tiny DNN of MobileNet-V2. To train the RoadSegNet, the Vision Benchmark Suite Data Set has been used in the present study.

The study presents a comparative study between the three state-of-the-art DNNs, of ResNet50, XceptionNet, and MobileNet-V2, and uses the DeepMind-V3+ encoder-decoder architecture for the segmentation. Apart from using these pretrained networks for weight initialization, another important aspect is their architecture. All these DNNs have a characteristic architecture and number of training parameters: the ResNet50 has 23 million trainable parameters, the XceptionNet has 22.8 million, and the MobileNet-V2 has only 4.2 million trainable parameters.

The main significance of this comparative study has been to evaluate the performance of these networks for edge deployment. So, the tiny DNN model of MobileNet-V2 has been considered, which has approximately 80% fewer tunable parameters as compared to the others, which makes it perfect for edge deployment. The execution time has also been compared in Table 5, and it can be observed that MobileNet-V2 offers a justifiable time for the segmentation and classification of the roads.

The performance of these trained models has been evaluated using the metrics of global accuracy, weighted IOU, and mean BF score. The trained models offer a global accuracy of between 96 and 97%. It has also been observed that the performance offered by MobileNet-V2, despite being a tiny deep neural network architecture, is comparable with that of XceptionNet and, in some cases, offers better performance than ResNet50.

Related work

The self-driving cars are autonomous decision-making systems, and this self-driving autonomy is divided into five SAE levels. The lower SAE levels offer basic driver assistance features like automatic braking, lane departure warnings, and adaptive cruise control, while the higher SAE levels are aimed at offering driverless navigation in all road conditions [12]. In the 1980s, Ernest Dickmanns developed the first autonomous car [13]. This was followed by various research efforts, like the development of Prometheus [14], VaMP [15], and CMU NAVLAB [16]. These advancements laid the groundwork for self-driving cars. In the early 2000s, DARPA's grand challenges [17] were one of the major turning points in the development of self-driving cars, where machine learning was used for the first time for navigation [18].

As these self-driving cars are autonomous decision-making systems and are being designed to assure road safety and efficient navigation, it is desired that the autonomous vehicle be able to not only perceive the current state of the driving environment, but also be able to foresee future behavior too. So, to estimate the current state and predict the future states of the driving environment, the self-driving vehicles use an amalgamation of onboard sensors like mono and stereo cameras, depth estimation sensors, LiDARs,

Fig. 1 Broad architecture of a perception, planning, and control workflow in autonomous vehicles

EMUs, GPS units, and ultrasonic sensors, and based on the sensed data, the autonomous vehicle will make navigational decisions. Broadly, the data from these sensors is primarily used for the following four tasks: (a) perception and localization, (b) high-level path planning, (c) behavior negotiation, and (d) intelligent motion control. These four high-level tasks also need to be monitored for safety. Figure 1 shows the representation of the broad architecture of a perception, planning, and control workflow in autonomous vehicles.

Perception and localization are two of the most important tasks to sense the dynamic traffic environment, and they leverage the use of various vehicle sensors. The various methodologies used in road detection are shown in Fig. 2. Some of the sensors used are discussed as follows:

- *Mono cameras* can be used for obstacle detection and classification; they offer a cost-effective solution and are good for two-dimensional mapping and lane detection, but they suffer from drawbacks like the fact that they are very sensitive to light and in poor lighting scenarios, like fog and rain; they offer a very poor performance. Also, it is very difficult to perceive the estimation of distance using such cameras.
- *Stereo-vision cameras* provide the same functionality as mono cameras, but they also allow for three-dimensional mapping and depth estimation. However, these cameras are computationally expensive; additionally, velocity and distance estimation cannot be estimated, and, like mono cameras, these are light-sensitive and do not provide good results in challenging lighting.
- *LiDAR* is also used for obstacle detection, robust 3D mapping of the driving scenario and environment using multi-layer LiDAR, direct estimation of the distances, efficacy in light weather conditions, etc., but the object classification is

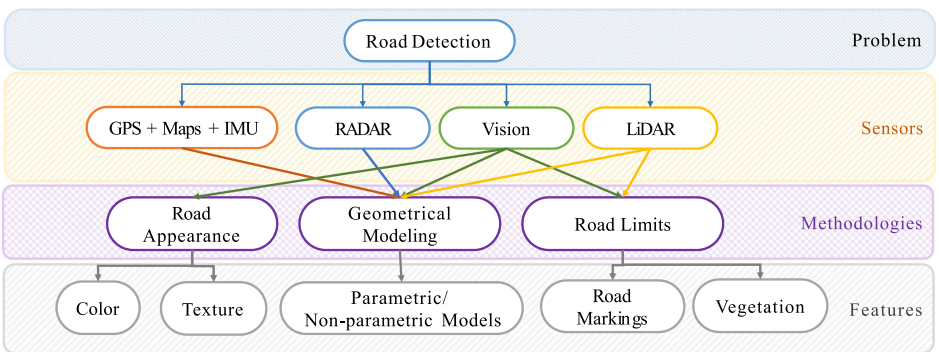


Fig. 2 Various approaches to road detection

a challenge; some inaccuracies can occur due to reflective surfaces and typically severe weather conditions.

- *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.

So, there is no one unique solution that offers good sensing and perception functionality, so multiple such technologies are used in conjunction with each other to offer accurate perception. The sensed data from various sensors are fused together to accurately perceive the driving environment. To localize the vehicle independently, methodologies like odometry, Kalman filters, particle filters, and simultaneous localization and mapping (SLAM) techniques are employed to estimate the state of the vehicle in a driving scenario. Figure 3 graphically illustrates the whole process of sensing, perception, localization, path planning, and vehicle motion control. Various road detection methodologies are given in Table 1.

After the successful completion of perception and localization, the next task is the trajectory or path planning to navigate the vehicle through the traffic. Path planning will influence the decision-making process and is the most important and challenging task. From the sensed data, the vehicle will try to understand the particular driving scenario, whether it is an intersection or a right turn, the states and behavior of the vehicles ahead, the various road signs, collision avoidance, etc. From this perceived information, the vehicle will learn and plan out all the possible trajectories, and using the machine learning models or state models, an inference will be made for navigating the vehicle through the road.

The last step in the process is the motion control of the vehicle. The vehicle motion control system influences the longitudinal and lateral movement of the vehicle, considering its dynamics. It engulfs the control of the steering, braking, and cruise control of the vehicle to assure that it sticks to the desired path on the road safely.

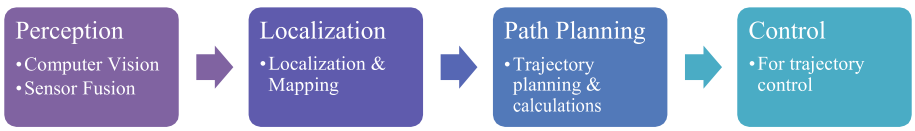


Fig. 3 Stages of the autonomous driving system

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

One of the main tasks while sensing, perceiving, and localizing the current driving environment is to detect the free (drivable) road, which has been of interest for the last few decades. This visual perception is done in order to detect collision-free space in the driving environment that will aid the advanced driving assistance systems in autonomous decision-making. Road scene segmentation is one of the important computer vision techniques used in autonomous driving. A typical driving scenario may consist of buildings, vehicles, roads, pedestrians, etc., so it is essential to obtain or segment the drivable area from the captured road scene for collision-free navigation [19]. Road detection includes the estimation of the extent of the road, the various lanes and their intersections, splits, and termination points in the diverse driving scenarios. A drivable region is a connected road surface that is not occupied by any obstacles like other vehicles, and people. The objective of road segmentation is to impose geometrical constraints on the various objects that are present in the driving scene [19]. Road segmentation basically allows the generation of an occupancy map of the perceived driving environment and uses this information in the automated driving workflow to navigate safely. Thus, it becomes essential to accurately and efficiently segment the drivable road region from the driving environment.

Traditionally, road segmentation is carried out using various computer vision algorithms that employ methodologies such as edge detection and histograms [20]. The key markers that aid humans in perceiving information about the road are color, texture, boundaries, and lane markings, and similar information can be used by driving assistance systems to safely navigate the driving environment. Vision-based perception has been prominent in the development of advanced driving assistance systems and is being coupled with various machine learning algorithms to develop the proof of concept for the SAE stage 2 to stage 3 level of autonomy in self-driving vehicles. But it is very difficult to do so, as road design and conditions vary throughout the globe and are not universally the same, so these computer vision algorithms will not offer universally uniform results.

Over the last few years, the development of full convolutional neural networks (CNNs) for semantic segmentation [21] boosted their adoption in autonomous driving, and the recent advancements in the development of massive or deep convolutional neural networks, like SegNet [22], will aid the driving assistance system in handling several diverse driving scenarios. Several researchers have used deep CNNs for the semantic segmentation of the driving scene. In [23], a DCNN has been reported for obstacle detection and road segmentation. The work proposes the use of a stereo-based approach to build a disparity map for obstacle detection in a driving environment. In [24], two networks, ENet and LaneNet, have been proposed to detect road features, and a weighted combination of the various features has been used for road detection. One CNN works on the detection of the road surface, and the other one is used to detect the lanes, and the output from both is merged to get an accurate and precise representation of the drivable road. A deep recurrent convolutional neural network (U-Net) for road detection and centerline extraction is discussed in [25]. The work involves the development of a novel RCNN unit incorporated into the U-Net framework for road extraction, followed by the multi-task learning scheme that handles both the tasks of road detection and centerline

extraction simultaneously. In [26], ResNet-101 has been used for the detection of the road. In [27], a deep NN, road and road boundary network (RBNet), is developed for unified road and road boundary detection simultaneously and eliminates the possibility of a pixel being misclassified as a road or road boundary. In [28], a CNN with gated recurrent units has been proposed for the fast and accurate segmentation of the road and solves the problem of complex computation that is prominent in the conventionally used very deep encoder-decoder structure to fuse pixels for road segmentation. In [29], a DCNN with color lines has been proposed for the segmentation of unmarked roads. The work uses a score-based mechanism to create a conditional random field-based graphical model to segment the road from the background. In [30], CNN, along with distributed LSTM, has been used to segment the road. 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 comprises three sections: the first one is a CNN-based local feature encoder, followed by a LSTM-based feature processor, and finally the CNN-based output decoder.

Also, recently, with the development of various sensor fusion technologies, deep learning-based multi-modal systems are being developed for autonomous vehicles [1, 31, 32]. 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. In [33], a 3D object detection system has been developed by fusing the data sensed from the RGB camera and LiDAR point cloud. By using the fused information, the work predicts 3D bounding boxes, and the network consists of two subnetworks, one meant for 3D object detection and another for multi-view feature fusion. Similar work has been reported in [34–38] where the data from the cameras has been fused with LiDAR point clouds for 3D object detection. Some research has also been focused on using multi-spectral camera images [39, 40], where the RGB images along with the far-, middle-, and near-infrared images have been used to perceive the multilateral information about the driving scene and for the perception of the depth.

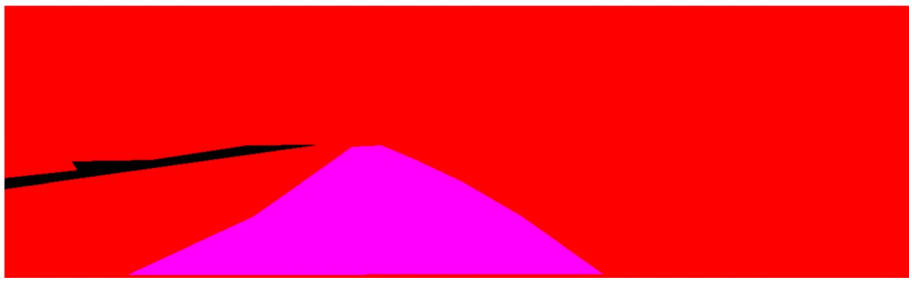
Background

Problem definition

Pavlidis [41] formally defined segmentation as *a process of pixel classification in which the input picture is segmented into subsets by assigning the individual pixels to classes*. For example, while segmenting a picture by thresholding its gray level, we are actually classifying the pixels into *dark* and *light* classes in an attempt to differentiate light objects from dark backgrounds or vice versa. In the literature, it has been reported that deep learning models are enriched with stacked layers (depth), and using these models, one can get high-quality results and that too with great accuracy. These models can utilize the maximum amount of unstructured data.

Semantic segmentation has a promising potential in autonomous driving for developing visual perception systems. The images captured from the various cameras present can be used to develop various driving assistance systems, like road and lane detection systems. Figure 4 shows an example of the road segmentation process. Figure 4a shows the image of the driving environment captured by a camera mounted on the car, and Fig. 4b shows the segmented image containing three classes: (a) the environment, shown

(a) Original Scene



(b) Segmented Image

Fig. 4 Image of the driving environment captured by a camera in the KITTI dataset. **a** Original scene. **b** Segmented image

in red color; (b) the drivable right road, shown in magenta color; and (c) the non-drivable left road, shown in black. Deep neural networks have proven to be beneficial for semantic segmentation in various diverse applications like medical imaging and autonomous driving. So, the usage of deep learning models for the segmentation of the drivable road has been explored in this current study.

Need for ground detection

In traditional automotive systems, there has been a tradeoff between distance sensitivity and object sensitivity as shown in Fig. 5. When the object is close, its sensitivity is high, allowing for better classification; as the distance increases, the object becomes farther away, potentially leading to poor results. To address good distance and object sensitivities, the current approaches would require too many computational resources. By knowing what and where the ground region in an image is, we can detect both objects and their distances. Also, for autonomous vehicles, it is essential to know about the drivable region in a driving scenario or environment. The proposed system in the present work aims at detecting and segmenting the road area using the KITTI road dataset [42], which will prove valuable in tasks like autonomous driving and navigation systems. For this purpose, "ground" has been defined as *a relatively smooth, drivable, and easily distinguishable from the surrounding surface. It may consist of common irregularities or imperfections or differing light conditions*.

The paper has been organized into the following sections: The "Results and discussion" section sheds light on the state-of-the-art research in the field of the development of advanced driver assistance systems for self-driving cars and the various techniques that are being used for perception and localization tasks. The "Evaluation metrics" section deals with establishing the background for various deep learning models used in the

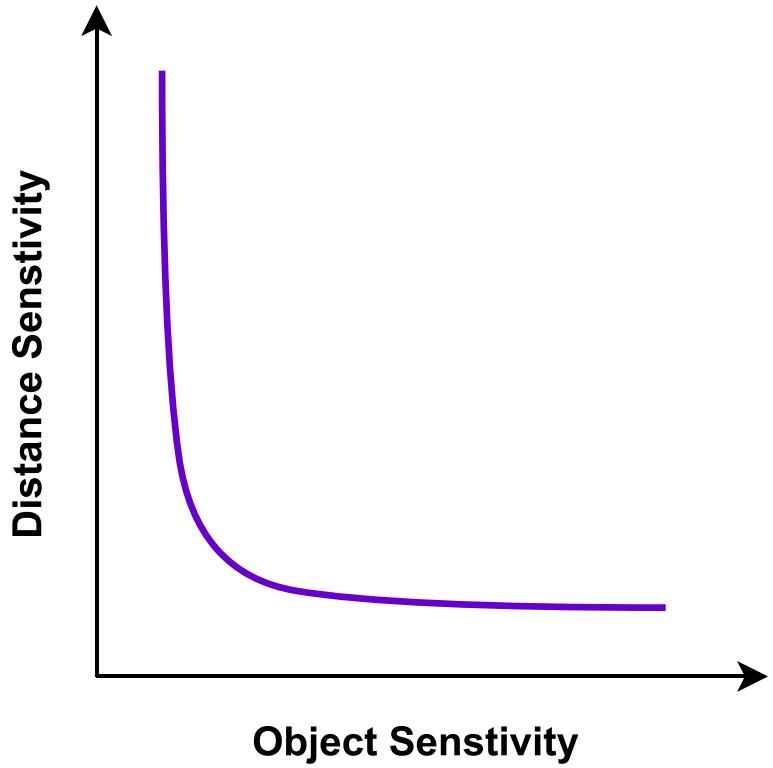


Fig. 5 Plot between the object sensitivity and the distance sensitivity

present study and how these can be used for road segmentation purposes. The "Training performance" section deals with the various DNN architectures, methods, and datasets used in the present work. The "Segmentation results" section deals with the segmentation results and the evaluation of the various performance indices, followed by discussions and the scope for future work in the "Discussion" section.

Methods

KITTI Vision Benchmark Suite Data Set

The KITTI Vision Benchmark Suite [42] is a dataset designed for object and road/lane detection. The road/lane dataset consists of 289 training and 290 testing images. Each image is 372 × 1242 pixels in size. All the images were acquired on five different days. The dataset is further divided into three categories of road scenes, as can be seen in Fig. 6:

- *Urban marked (UM), single-lane road with markings*
  - Consisting of 95 training and 96 testing images
- *Urban unmarked (UU), single-lane road without markings*
  - Consisting of 98 training and 100 testing images
- *Urban multiple marked (UMM), multi-lane road with markings*
  - Consisting of 96 training and 94 testing images

The input images have two sets of label images; these images are color-coded with areas of interest. In one set, it identifies all the roads, and in the other set, it identifies

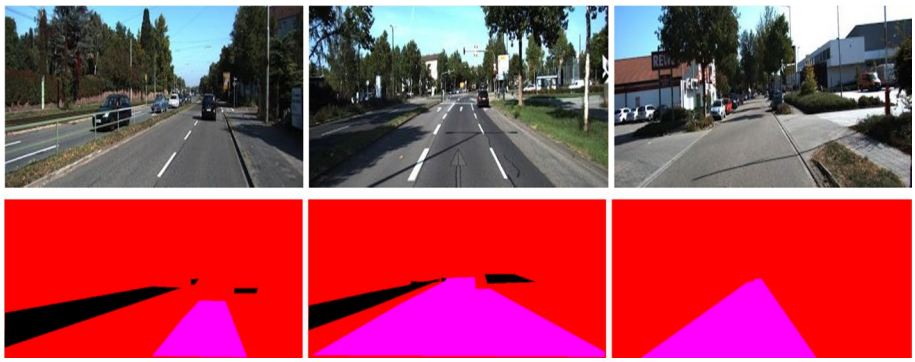


Fig. 6 The three different road scene categories with three RGB color codes. **a** UM. **b** UMM. **c** UU

just the lane where the car is moving. In the current study, the set corresponding to all the road surfaces has been used. These labels are RGB images that color code the road as magenta, non-road areas as red, and left road surfaces as black. The dataset has been pre-processed and augmented according to the input layers of the network before being fed. For ResNet50 and MobileNet-V2, the data set has been resized to 224 × 224 pixels, and for Xception, it has been resized to 299 × 299 pixels.

Methods

The present work is based upon Google's DeepLabV3+ semantic segmentation model, as shown in Fig. 7, and the architecture and weights have been initialized from three different pretrained networks, viz., two high accuracy models of ResNet50 and XceptionNet and one tiny DNN of MobileNet-V2, typically for the edge deployment. All of these networks and models are discussed as follows:

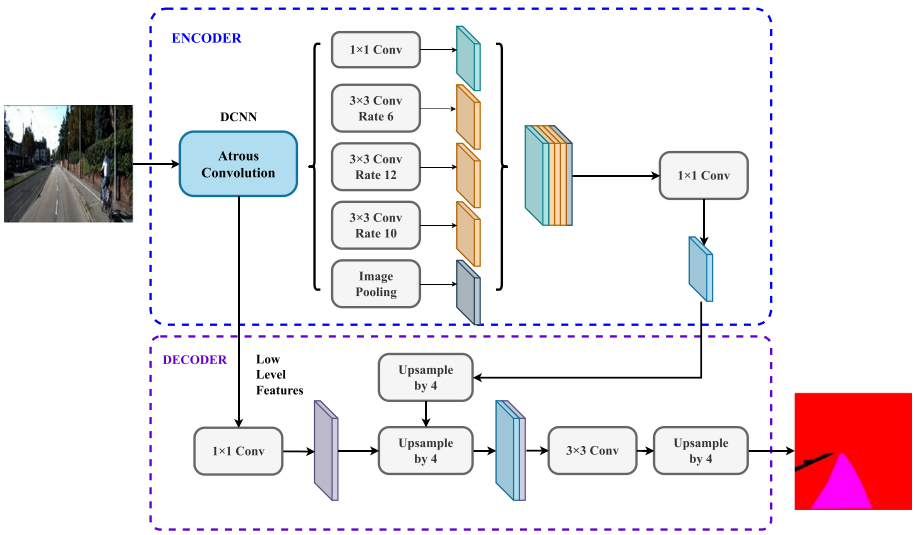


Fig. 7 Encoder-decoder of DeepLabV3+ architecture

RoadSegNet architecture

The RoadSegNet is built around the cutting-edge DeepLabV3+. DeepLab [43] is an open-source semantic segmentation model designed by Google and works by adding a simple decoder module that helps in segmenting objects along boundaries and also refines the segmentation results. More rapid results are achieved by using depth-wise separable convolution for both Atrous Spatial Pyramid Pooling and the Decoder Module [43]. The weights were initialized using the transfer learning method. The three state-of-the-art DNNs have been utilized. The work considers the use of two high-accuracy models, ResNet50 and XceptionNet, and one tiny DNN, MobileNet-V2. DeepLabV3+ uses an aligned Xception network as its key feature extractor, along with the following modifications:

a) The max pool layers are replaced by depth-wise separable convolution and striding.
b) Additional batch normalization and ReLU activation are added after each 3×3 depth-wise convolution.
c) The depth of the model is increased without changing the entry flow network structure.

The encoder works on an output stride, i.e., the ratio of the original image size to the size of the final encoded features. Instead of using bilinear up-sampling with a factor of 16, the encoded features are first unsampled with a factor of 4 and concatenated with corresponding low-level features from the encoder module having the same spatial dimensions. To reduce the number of channels, 1×1 convolution is applied before concatenating on the low-level features. After concatenation, a few 3×3 convolutions are applied, and the features are unsampled by a factor of 4. This gives the output the same size as the image. The semantics of the proposed RoadSegNet architecture based on DeepLabv3+ are shown in Fig. 8 as below.

Results and discussion

Evaluation metrics

To evaluate the efficacy of the obtained segmentation results, the metrics (a) global accuracy, (b) mean accuracy, (c) mean IoU, (d) weighted IoU, and (e) mean BF score have been used. For describing these evaluation metrics, the following terms are used:

- *False positive (FP):* pixels belonging to the background but misclassified as lesions
- *False negative (FN):* pixels belonging to lesions but misclassified as background
- *True positive (TP):* pixels belonging to lesions and correctly classified as lesions

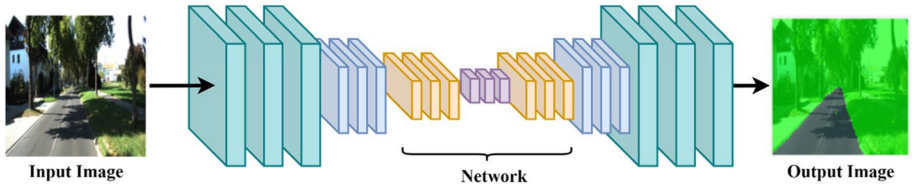


Fig. 8 RoadSegNet architecture

- *True negative (TN):* pixels that belong to the background and are correctly classified as such

Accuracy

It can be calculated for each class separately as well as globally for all classes. The accuracy gives the proportion of correctly classified pixels in each class and is given in Eq. 1

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

Global accuracy

The global accuracy is the ratio of pixels correctly classified to the total number of pixels and is given in Eq. 2

$$\boldsymbol{Global\ Accuracy} = \frac{\boldsymbol{TP} + \boldsymbol{TN}}{\boldsymbol{TP} + \boldsymbol{TN} + \boldsymbol{FP} + \boldsymbol{FN}} \tag{2}$$

Mean accuracy

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

Intersection over Union (IoU)

It calculates the incorrect classification of the pixels and is given in Eq. 3.

$$\boldsymbol{IoU} = \frac{\boldsymbol{Lesions} + \boldsymbol{Background}}{\mathbf{2}} \tag{3}$$

where

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

Weighted IoU

The weighted IoU is used when there is a disproportionate relationship between the class sizes in the images, minimizing the penalty of the wrong classification in smaller classes. It is given in the equation as follows:

$$\text{Weighted IoU} = \left(\text{Lesion Weight} * \text{Lesion}\right) + \left(\text{Background Weight} * \text{background}\right)$$

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

It calculates the alignment between predicted borders to the gold standard one. It is given by the harmonic mean of recall and precision as shown in Eq. 4 as:

$$\text{BF Score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \tag{4}$$

where

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

Training performance

In the proposed work, the KITTI Road/Lane Detection Evaluation Dataset 2013 [42] has been considered. To accommodate the dataset with the proposed architecture of RoadSegNet, the dataset has been preprocessed to meet the requirements of each of the individual deep neural networks of ResNet50, Xception, and MobileNet-V2. The various class labels have been redefined as *the environment*, *the left road*, and *the right road*, and accordingly, the LabelIDs and ColorMaps for the KITTI Road dataset [42] have been modified. For training the various networks, the algorithm-specific learning option of stochastic gradient descent with momentum (*sgdm*) has been used for all three networks. The initial learning rate has been considered as 0.001, and the maximum number of epochs has been taken as 100 for all the networks. The mini-batch sizes are set according to the GPU specifications, and the rest of the parameters are kept the same. All the models have been trained in MATLAB 2020b environment running on a Windows 10 PC, with Ryzen 9, 12 Core CPU with 16 GB of RAM, and a Nvidia 2060 super 8 GB GPU. Figures 9, 10, and 11 show the plot for the training loss, training accuracy, and base learning rate for all the networks considered for the RoadSegNet, namely, ResNet50, XceptionNet, and MobileNet-V2, respectively. From the plots, it can be observed that the training loss function minimizes as all of these networks achieve good training accuracy of approximately ~ 96 to 97%.

Segmentation results

After training the RoadSegNet, the network is fed with various driving scene images from the KITTI Road Eval Dataset. The segmented images for ResNet50, XceptionNet, and MobileNet-V2 are shown in Table 2. The first six images in each table show the best obtained segmentation results, and the last three images (S. nos. 7 to 9) show the segmentation results for very harsh driving scenarios in heavily shadowed regions where the segmentation becomes quite challenging. 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 with high precision, even in very shadowed areas. All the evaluation parameters have been tabulated in Tables 3 and 4. Table 3

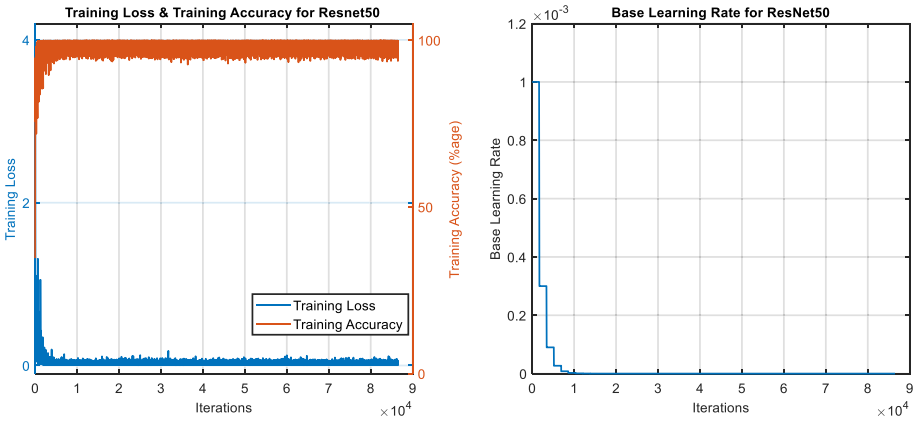


Fig. 9 Plot for training loss, accuracy, and base learning rate for ResNet50

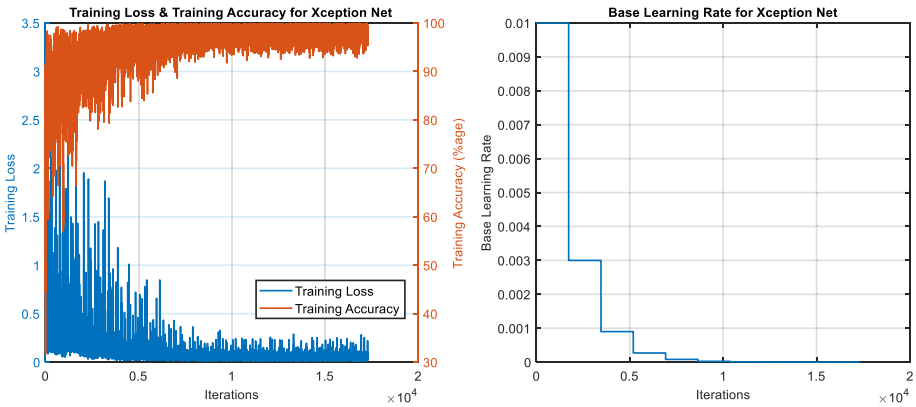


Fig. 10 Plot for training loss, accuracy, and base learning rate for XceptionNet

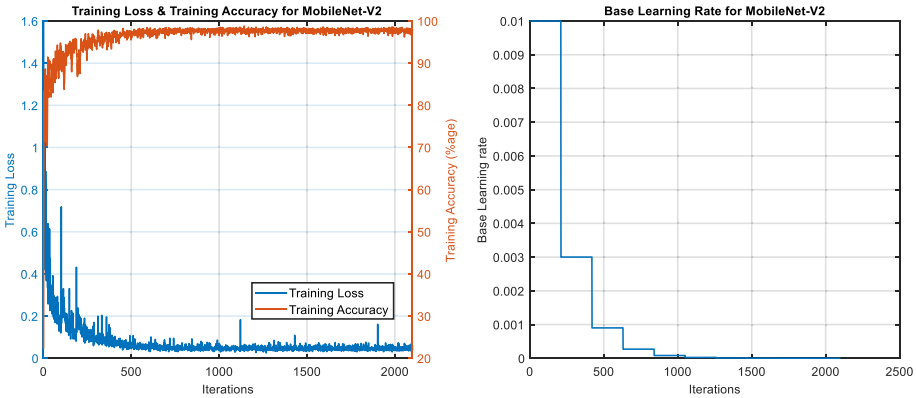
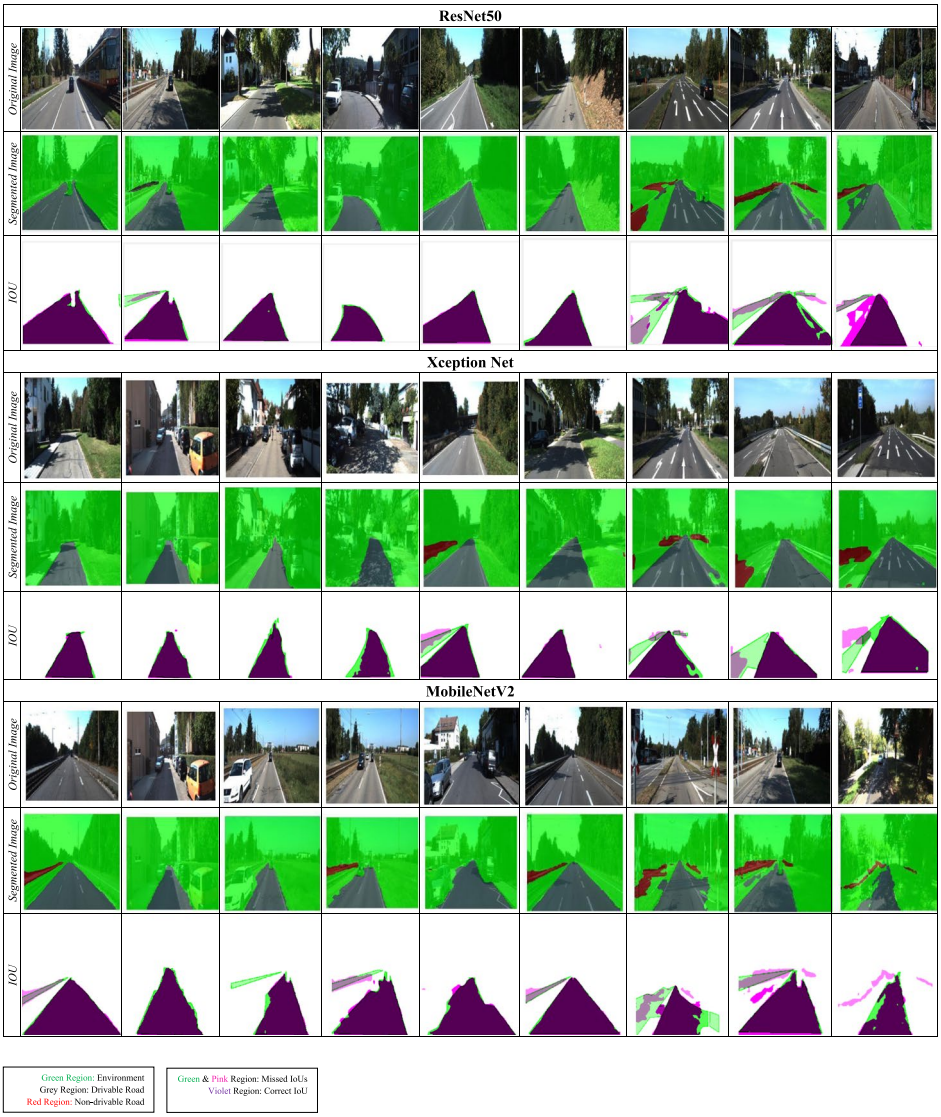


Fig. 11 Plot for training loss, accuracy, and base learning rate for MobileNet-V2

gives the comparison of the various performance metrics like global and mean accuracy, mean and weighted IOU and mean BF score for the entire training, and testing and validation datasets for each designed network. Table 4 gives the information regarding the accuracy, IOU, and mean BF score for each class, i.e., with what precession a particular class has been detected for the entire training, testing, and validation datasets for each designed network. Figures 12, 13, and 14 show the radar plot for

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



each network for all the performance metrics. From the obtained results in Tables 2, 3, and 4, it can be observed that the developed networks offer very good accuracy, the global accuracy ranges between ~ 96 and 97%, the weighted IOU also spans between ~ 92 and 97%, and the mean BF score too varies between ~ 0.75 and 0.83. It can also be observed from the obtained results that the MobileNet-V2, despite being a tiny deep neural network architecture, offers almost comparable performance with the XceptionNet and, in some cases, offers better performance than the ResNet50.

Discussion

Any autonomous driving system consists of four stages, viz., perception, localization, path planning, and control. The present work is focused on perception tasks. The scope of the work presented in this paper is to build a deep learning-based ground detection

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

system. The results as obtained in the "Segmentation results" section validate the robustness of the system by detecting a significant part of the road, even in the improperly illuminated regions. The left road regions are not detected well due to a smaller number of images being labeled for the region, as can be seen in Tables 2, 3, and 4 (S. no. 7–9). This can be improved by using a dataset with more of these images. The developed framework performs best on bright images, as can be seen in Tables 2, 3, and 4. The work explores the application of three different pretrained networks of ResNet50 and XceptionNet (high accuracy models) and MobileNet-V2 (tiny DNN) typically for edge deployment. It can be observed that the accuracy of MobileNet-V2 is on par with the accuracy of the high-accuracy models of ResNet50 and XceptionNet. With added capabilities like lane detection, depth estimation, and intersection detection, the proposed model can be used for efficient road detection tasks. Although the model performs well in daylight conditions, the capability of the model in nighttime scenarios has not been tested, which still poses a challenge for autonomous vehicles.

In the paper, a comparison has been made between the various state-of-the-art DNNs of ResNet50, XceptionNet, and MobileNet-V2. Table 2 shows qualitatively that the IOU for the trained models provides excellent performance for brightly lit roads as well as in very complex shady conditions. This observation has been established quantitatively in Tables 3 and 4.

In Table 3, the metrics of global accuracy for the segmentation have been analyzed, and it is observed that the models offer an accuracy above 97% for the training dataset and above 96% for the test database. Also, the other metrics of mean accuracy, mean IOU, weighted IOU, and mean BF scores have been evaluated for all three DNN models, and these have been established for both the training and the testing datasets.

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

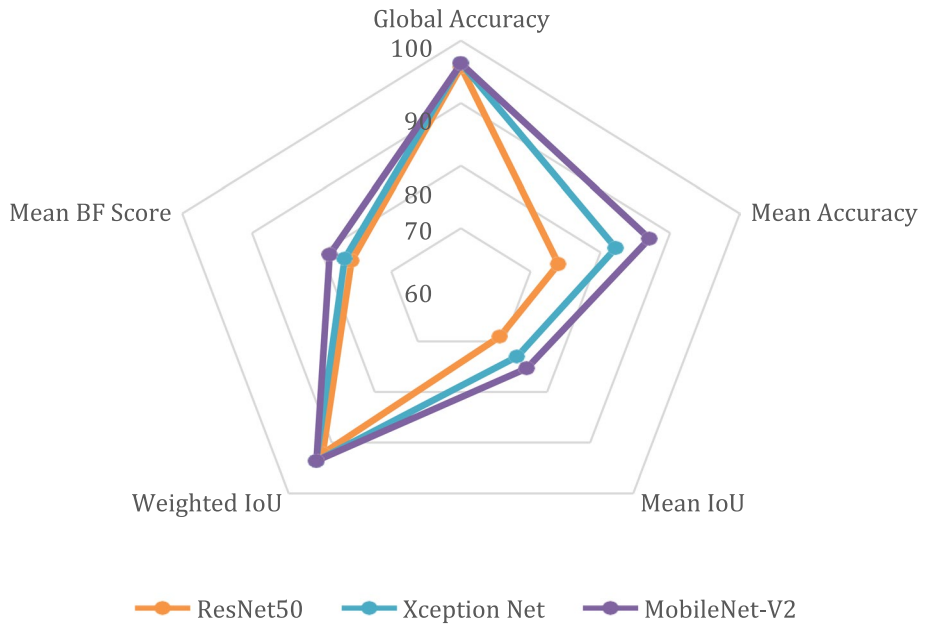


Fig. 12 The radar plot for each network for all the performance metrics for the testing dataset

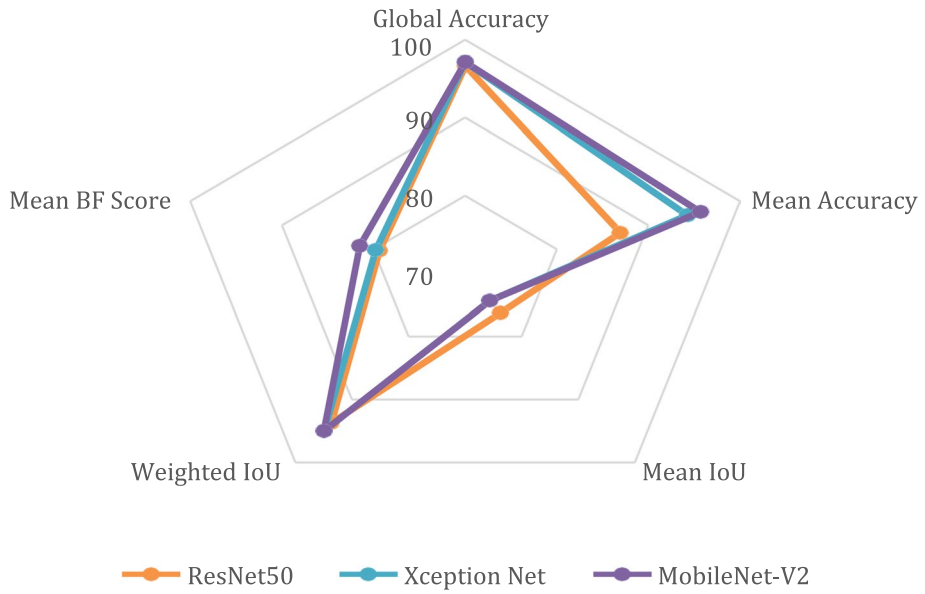


Fig. 13 The radar plot for each network for all the performance metrics for the validation dataset

Similarly, in Table 4, the comparison of the class-wise accuracy of the 3 DNNs has been made such that they are able to accurately segment and classify the various classes in the dataset, viz., left road, right road, and environment. The metrics of accuracy, IOU, and mean BF score have been used to evaluate the efficacy of the three DNNs, and the evaluation has been done on the training, testing, and validation datasets, and it can be observed from Table 4 that good results have been obtained. The drivable section in the

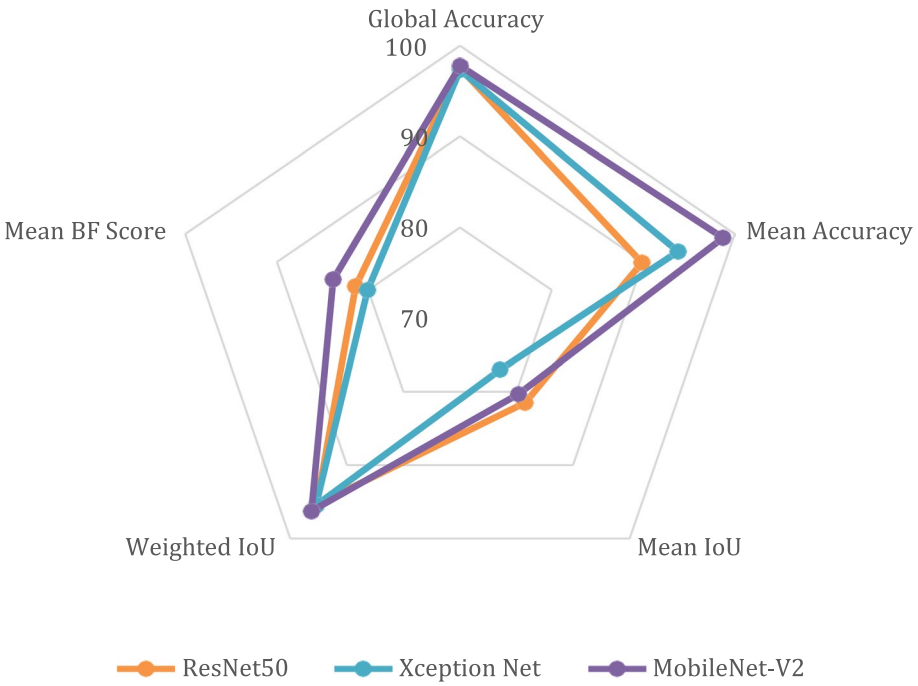


Fig. 14 The radar plot for each network for all the performance metrics for the training dataset

dataset is the "right road," and it can be observed that an accuracy of ~ 99% has been obtained for MobileNet-V3, and other two networks also offer an accuracy of about 91% and 97%. Similarly, for the environment, an accuracy of 97% is obtained for MobileNet-V3, and other networks too offer an accuracy of above 97%. Similarly, the performance has been evaluated for the testing as well as the validation dataset. Table 5 presents a comparison of the current work with the work already reported in the literature, and it can be observed that the current work offers one of the highest accuracies and that too in a minimum amount of runtime.

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)

Conclusions

In this study, a deep learning-based autonomous road detection system has been proposed. The proposed framework is built on the DeepLab-V3+ architecture, which is a state-of-the-art semantic segmentation network developed by Google. The weights of the network are initialized by three image classification networks, namely, ResNet-50, MobileNet-V2, and Xception. The results are evaluated on the benchmarked KITTI road dataset. The model is tested for adverse light conditions and general ground complexities, while also achieving significant results on the evaluation metrics. The proposed model also achieves good results on a small and yet powerful network, MobileNet-V2, that can be used in systems that require low power and can be used for edge deployment.

Abbreviations

ADAS	Advanced driver assistance systems
CMU	Carnegie Mellon University
CNN	Convolutional neural network
DARPA	Defence Advanced Research Projects Agency
DCNN	Deep convolutional neural network
DNN	Deep neural network
IMU	Inertial measurement unit
FN	False negative
FP	False positive
GIS	Geographic information system
GPS	Global Positioning System
IOU	Intersection Over Union
LiDAR	Light detection and ranging
LSTM	Long short-term memory
NAVLAB	Navigation Laboratory, CMU
NN	Neural networks
RCNN	Recurrent convolutional neural networks
ReLU	Rectified linear unit
RGB	Red, green, and blue
SLAM	Simultaneous localization and mapping
TN	True negative
TP	True positive
UM	Urban marked
UMM	Urban multiple marked
UU	Urban unmarked

Acknowledgements
Not applicable.

Authors' contributions
KP and PY are responsible for the initial conceptualization, coding, pre-filtering of the dataset, training of the models, and drafted of the initial manuscript. NK worked on the training of the models, revisions of the manuscript, and analysis of the results. All authors have read and approved the manuscript.

Funding
No funding was obtained for this study.

Availability of data and materials
The dataset is openly available at KITTI Repository [link: http://www.cvlibs.net/datasets/kitti/eval_road.php]

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 24 June 2022 Accepted: 18 November 2022
Published online: 12 December 2022

References

1. Feng D, Haase-Schuetz C, Rosenbaum L, Hertlein H, Glaeser C, Timm F, Wiesbeck W, Dietmayer K (2020) Deep multi-modal 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 uncertainty-aware 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 definitions 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 traffic 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 unified 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 classification 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

Publisher's Note