Journal of Autonomous Intelligence (2024) Volume 7 Issue 3
doi: 10.32629/jai.v7i3.1059

ORIGINAL RESEARCH ARTICLE

Detection of lanes, obstacles and drivable areas for self-driving cars using multifusion perception metrics

A. Kishore Kumar, Venkatesh Palanisamy*

School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore 632014, India

*** Corresponding author:** Venkatesh Palanisamy, venkatesh.palanisamy@vit.ac.in

ABSTRACT

Autonomous vehicles have been a recent trend and active research area from the onset of machine learning and deep learning algorithms. Computer vision and deep learning techniques have simplified the operations of continuous monitoring and decision-making capabilities of autonomous vehicles. A navigation system is facilitated by a visual system, where sensors and collectors process input in form of images or videos, and the navigation system will be making certain decisions to adhere to the safety of drivers and passers-by. This research article contemplates the model of obstacle detection, lane detection, and how the vehicle is supposed to act in terms of autonomous driving situation. This situation should resemble human driving conditions and should ensure maximum safety to both the stakeholders. A unified neural network for detecting lanes, objects, obstacles and to advise the driving speed is defined in this architecture. As far as autonomous driving is considered, these target elements are considered to be the predominant areas of focus for autonomous driving vehicles. Since capturing the images or videos have to be performed in real-time scenarios and processing them for relevant decision making have to be completed at a swift pace, a concept of context tensors is introduced in the decoders for discriminating the tasks based on priority. Every task is associated with the other tasks and also the decision-making process and hence this architecture will continue to learn every day. From the obtained results, it is evident that multitask networks can be improved using the proposed method in terms of accuracy, decision-making capability and reduced computational time. This model investigates the performance using Berkeley deep drive datasets which are considered to be a challenging dataset.

Keywords: autonomous vehicle; deep learning; image processing; self-driving car, multi-task network

ARTICLE INFO

Received: 27 July 2023
Accepted: 1 November 2023
Available online: 8 January 2024

1. Introduction

Recent statistics released in global report on road safety by World Health Organisation reveal that nearly 1.35 million people died in road traffic accidents, according to 2018. Another recent study claims that nearly 50,000 accidents occur in America alone, and most of them occur due to presence of obstacles[1]. These press reports indicate that obstacles are the serious concern to be addressed irrespective of autonomous or manual driving scenarios. Having accounted for over hundred deaths and nearly 10,000 accidents, obstacles on the road must be removed by the respective officials. From various reports, it is also understood that 90% of the drivers cause these accidents due to errors and misjudgment. In order to address these issues, departments of the government and car manufacturers should work together by considering different factors and extending their support to predict and act accordingly. There has been intensive research and huge investments deployed by car manufacturers for developing at enormous vehicles and cognitive robots to be launched in this society[2]. European Nation has invested nearly 1 billion euros for accomplishing futuristic cars and accident-free roads. Google, and Waymo have defined the onset of autonomous vehicles and tested them in different states of America long back in 2009. Ford and BMW as also assisted Uber for designing and deployment of autonomous vehicles on roads. Ever since its onset from 2009, this technology has become a common ground for many states in America and consistently monitored with specialized regulations[3]. It is a projection by Victoria transport policy that autonomous vehicles would be common in the Year 2040. Adaptive cruise control is the miniature version of autonomous cars upon which intelligent Lane changing control, emergency braking systems, light detection and ranging, traffic signal detection, street sign detection, vehicle to vehicle communication, object avoidance system, collision avoidance system are deployed.

Autonomous Road vehicles are becoming more common in the recent past where many manufacturers have involved themselves in research and development for bringing in more sophisticated and fully automated vehicles. Google has been a significant example where they have invested almost 20 years of research and Tesla has been successful in launching such vehicles[4,5]. The core expectation of autonomous vehicles is to sense the presence of obstacles, altering the lanes for positioning the vehicle to avoid any accidents and ends maintaining a particular route. These two are the common problems that need to be addressed in any research and still, there is a minimal progression in this domain[6]. The proposed solutions, apart from solving these two predominant problems, have to be reliable, robust, efficient, and affordable. The common techniques used for detecting obstacles in a long-range are Radar, laser scanners, computer vision, and various deep learning algorithms. Many techniques have investigated the approach of sensor fusion to address the drawbacks of individual methodologies and mechanisms[7]. Concerning the detection of neighbouring vehicles, computer vision, and Laser scanners detect the presence of search vehicles and guide the autonomous vehicle accordingly. The continuous changes to highways, roads, streets add a burden to these autonomous vehicles as the system may consider small obstacles or debris, water droplets, lights, shadows, and other factors as obstacles. The vehicles' sensors may capture different objects and classify them as potential elements even when they are not. Sensors may be classified into active and passive, where LIDAR and RADAR are considered to be active and cameras are considered to be passive[8]. The purpose of active sensors is used to measure the distance and speed of the vehicle in accordance to be object obstacle with huge precision. Displacement of a vehicle from time to time and its correspondence to obstacles are identified with greater accuracy when active sensors are deployed. The problem with active sensors utilization is associated with the high-cost factor and lower resolution than a camera. This is where the camera will be introduced as a passive sensor, where accuracy is lesser but resolution and proper identification are better[9]. Timely detection of obstacles, transferring them into the system, and making a timely decision will be a challenge when passive sensors are part of the system.

One common observation is that, both in the cases of active and passive sensors, detection of small objects in a very small area in different forms, shapes and sizes is going to be difficult. There are high chances that the gradients of paper, sawdust, mud after the rain, etc., possess the same texture in terms of active and passive sensors. All these factors may contribute to false positives for a Lane detection system obstacle detection system. Research and development have taken autonomous vehicles to new heights and in recent developments, convolutional neural networks have been majorly implemented for obstacle detection and safe navigation[10]. During the training phase of a conventional neural network, a very close association that depicts the relationship between the road conditions and the driver's behaviour is monitored, corresponding to the inputs updated from the steering wheel and the inputs from active and passive sensors. When the model is deployed into the testing phase, based on the obstacles detected during run time, the convolutional neural network provides a prediction of a possible Steering Wheel angle for safe navigation. But the chances of training a model completely for different types of obstacles are nearly impossible and false positives will provide inaccurate information which may lead to unwanted decisions and hence may result in a collision or accident[11,12]. Failure of an automatic navigation system may be due to any one of the following reasons, where the sensors fail, the transmission of information failure or the software failure. One such example is where a Tesla failed to detect an overturned truck on the highway and a pedestrian trying to cross a road in its autopilot mode.

This approach investigates the model and changes required in the conventional system for ensuring accuracy and robustness in obstacle detection. A Markov Random Field will be altered for obstacle detection and lane detection, where the model considers the areas under a region suitable for driving. Road Segmentation and drivable areas were not part of the ecosystem in conventional approaches[13], and the proposed system introduced lane line, traffic objects, obstacles and additional perception strategies for intelligent and autonomous vehicles. A model of multifusion Network is proposed to consider a wide range of activities ranging from classification, detection and semantic segmentation will be fused into the architecture for enhanced performance. The proposed architecture intends to process the perception of lanes, other drivable areas, traffic objects such as cones, and obstacles using a deep learning model[14]. The image captured from active and passive sensors will yield and mark the obstacles using a Markov random field. To be more specific, the model includes intensity gradient parameter, curvature cue parameter, and differentiation of variance to discriminate the obstacles. Moreover, to increase the computational performance, subnets are included as a context tensor for processing the information in a parallel fashion. On the whole, the proposed method proves to improve efficiency and performance without any computational overheads. The computational resources are used in parallel subnets, thus eliminating the serial processing approach. After analysing the distance between the vehicle and the obstacle, speed of the vehicle, the deep learning model will navigate the autonomous vehicle into probable and drivable areas[15]. Deep learning and sensor fusion for obstacle detection in autonomous vehicles indeed hold significant promise, but a comprehensive exploration of their advantages and limitations within the context of our research is crucial.

1.1. Advantages

Enhanced Perception: Deep learning algorithms, particularly Convolutional Neural Networks (CNNs), excel at recognizing complex patterns in sensor data, improving the vehicle's ability to perceive and classify obstacles accurately.

Adaptability: Deep learning models can adapt to changing environments and learn from vast datasets, making them versatile for different road conditions and scenarios.

Real-time processing: Modern hardware acceleration allows for real-time processing of deep learning models, enabling rapid decision-making for obstacle avoidance.

Multimodal sensor fusion: Combining data from various sensors like LiDAR, radar, cameras, and ultrasonic sensors through sensor fusion improves redundancy and robustness in obstacle detection.

Generalization: Deep learning models can generalize from training data to handle novel obstacles or scenarios not explicitly encountered during training.

1.2. Limitations

Data dependency: Deep learning models require extensive labeled data for training, which can be time-consuming and expensive to acquire.

Computation intensive: Deep learning algorithms are computationally intensive and may require specialized hardware for efficient implementation.

Overfitting: Without proper regularization techniques, deep learning models may overfit the training data, leading to poor generalization.

Interpretability: Deep learning models are often considered "black-boxes," making it challenging to interpret their decision-making processes, which is crucial for safety-critical applications.

Sensitivity to adverse conditions: Extreme weather conditions like heavy rain, snow, or fog can challenge the performance of sensor fusion systems, affecting obstacle detection accuracy.

Sensor reliability: The effectiveness of sensor fusion heavily relies on the reliability and calibration of individual sensors, and sensor failures can lead to erroneous obstacle detection.

Section 2 reviews and presents related models carried out in previous research works. Section 3 presents the proposed architecture used for detecting obstacles and drivable areas, section 4 process simulation model, and results after the demonstration. Section 5 concludes the presented work with relevant discussions and future directions.

2. Related work

2.1. Obstacle detection techniques

Obstacle detection has been primarily focused in this research work, and it has been carried out by generating an occupation map in previous methodologies. The process of creating an occupation map involves the orthogonal projection of 3D objects which are protruding from ground levels[16]. The road surface is considered to be a planar surface and all surrounding environments will be the stated differently. The entire planar structure will be divided into number of cells to form a grid. Implemented algorithm will look out for presence of any object in either of these cells. Stereo vision sensor was responsible for capturing the occupants of each cell and has suitably recognised large objects such as pedestrians, cyclist and passers-by. The road boundary was detected using LIDAR, where the Precision and accuracy are affected by the probability function which is responsible for estimating the number of measurements available in one cell of the whole grid[17]. The common problem with stereo vision sensors is that failure to detect smaller and sudden obstacles rising in front of the vehicle. Detection accuracy can also be affected due to presence of excessive noise in the images.

The next algorithm is known as digital elevation map (DEM), which is considered to be one of the prominent algorithms for detecting obstacles and that are highly differentiated from the planar surface[18]. Density of every object, including these roads, lanes, and other traffic elements are measured and marked inside the DEM cells, based on which the algorithm functioned. The road surface was considered to be a planar surface and constituted a surface model for the entire algorithm. This model was extended into a random sample consensus applied on to stereo vision sensors, and finally the input was constructed on DEM cell. A density base classifier was implemented classify the type of obstacle present in the surface. The theory of random sample consensus was further extended into a polynomial fitting approach where curb detection was applied over DEM cells[19]. The common problem with all these models was the inability to perform the classification of smaller obstacles due to variation of disparity. Outcome in form of DEM cells, expected clear images without any noise level and demanded at least one cell which is smooth to indicate the road surface for drivable areas. Random sample consensus approach was successful in in estimating the number of vehicles, their relative position and displacement with the current autonomous vehicle and the road surface[20].

2.2. Fusion techniques

These deficient algorithms demanded the fusion of high-resolution cameras and three-dimensional LIDAR for sensing the surface and better curb detection. From the article surveyed, it is understood that the accuracy of obstacle detection depends on equality of fusion methods that includes ranges and visuals[21]. Conventional methods of fusion relied upon range information completely for curb detection, that exhibited the benefit of obstacles present at a long distance. The purpose of detecting small obstacles is still not resolved and hills fusion of three-dimensional LIDAR is recommended along with visual data and range data.

Considering the quality of obstacle detection and edge detection in traditional methodologies[22], fusion of multi sensors prove to be a better solution for clear images. Concerning the fusion methodologies, the depth and density of an image for every scene is reconstructed using sparse range points. The geometrical properties of high resolution image followed by propagation of depth information, edge detection, image recovery, depth recovery, curb point detection, mapping curb points, factoring and refactoring curb points, refining the curbs and parametrization have been simplified[23].

Another variant of obstacle detection algorithms is based on scene flow mechanism where a cloud of images is constructed virtually and temporally to analyse how the images relate with each other. Objects present in the cloud of images can be discriminated in to surface of the road, object on the road such as bicycles, passer-by, traffic signals and so on. Relationship between a particular object in different scenes, are carefully analysed for predicting the next movement[24]. This approach was greatly helpful in determining the trajectories of moving objects and hence assisted in better decision making. The flow of points connecting between different scenes may not be as usual in case of standard object. This approach also lacked recognition of smaller obstacles and in case of sparse information[25]. Despite using advanced stereo vision sensors and algorithms, these models were capable of composing three-dimensional motion-based approach mapping and were efficient in detecting bigger obstacles such as pedestrians, cyclists and other vehicles.

The next model introduced a vision based neural network with a dedicated classifier which was available online. The algorithm was capable of detecting and predicting obstacles present at a long distance, which is followed by deep stereo analysis for allowing the vehicle to safely navigate autonomously[26]. With a more focus on pedestrian or neighboring vehicle detection, an algorithm implemented category based parameters of different images collected at runtime along with geometric parameters. Since the information was vast, differences between vehicles with respect to shape, size and other perceptions, the vision sensor faced additional challenges to capture smaller obstacles on the path. Numerous research work has been carried out to document the different algorithms available for obstacle detection based on stereo based sensors. Stixel algorithm is one of the predominant algorithm which used geometric parameters constructed into a cluster-based access points[27]. The benefit of Stixels algorithm was that, the model was capable of differentiating the ground surface on the road and other obstacles by placing them on different verticals on a scenery. The algorithm was best known for its representation of surface and obstacles in different verticals. The problem with this method was the inability to detect a cluster point for specific obstacles and medium size obstacles would be detected unlike the previous methods but not from long distance. This is due to the fact that the size of the obstacle decreases with respect to distance[28]. Similar to other methods, this approach failed to detect the obstacles with great accuracy in terms of position.

The next model to be discussed implemented a plane estimation technique based on homographic subjected a lot of parameters as a homographical scores, pixel segmentation, line segmentation in a Markov random field architecture. Since multiple parameters were included[29], they always collide with each other minion the process and appearance based indexing process failed to correlate with each other and hence this model failed to detect intricate details of obstacles in form of gradient information. All the methods discussed so far as considered the road and driver by areas a flat surface or a free space and hence the obstacles are considered to be slightly higher than the planar surface. Deviation in form of geometrical parameters can be derived from image and calculated information of cloud resources, or by constructing a disparity histogram to determine the height of obstacles. In some techniques, the obstacles were considered to be clothoid or splines based on the three dimensional variations from the planar surface. These models required a careful analysis in disparity variations and domains after applying multistage filtering. The commonly used data sets were Udacity from USA, KTTI, Cityscape which supported proper segmentation of data and images but lack the information about optical parameters[30].

Convolutional neural networks (CNNs): CNNs are the backbone of our obstacle detection system, primarily used for image-based obstacle recognition. We used popular pre-trained CNN architectures like ResNet, VGG, or custom-designed CNNs tailored to the specific requirements of our dataset. These CNNs were responsible for extracting features from camera images.

Recurrent neural networks (RNNs): In addition to CNNs, we utilized RNNs, such as Long Short-Term Memory (LSTM) networks or Gated Recurrent Units (GRUs), for sequences of sensor data, such as LiDAR or radar measurements. RNNs helped capture temporal dependencies in the sensor data, allowing for better context-aware obstacle detection.

Fusion techniques: To combine information from multiple sensors like cameras, LiDAR, radar, and ultrasonic sensors, we employed sensor fusion techniques. This involved merging feature representations from various sensors at different stages of the neural network architecture. Fusion strategies included early fusion (combining data at input layers), late fusion (combining data at later layers or after feature extraction), and attention mechanisms to dynamically weigh sensor contributions.

YOLO (you only look once): YOLO is an object detection algorithm that we integrated into our pipeline for real-time bounding box detection of obstacles in camera images. YOLO is known for its speed and accuracy, making it suitable for fast-paced autonomous driving scenarios.

Semantic segmentation: For precise pixel-level obstacle segmentation in images, we implemented semantic segmentation networks like U-Net, FCN (Fully Convolutional Network), or DeepLab. These networks provided fine-grained information about obstacle boundaries, which is crucial for path planning and obstacle avoidance.

Transfer learning: Leveraging pre-trained models on large datasets like ImageNet, we fine-tuned these models on our specific obstacle detection dataset. Transfer learning helped reduce training time and improve detection performance, especially when data for our specific task was limited.

Data augmentation: To mitigate the data scarcity challenge, we applied data augmentation techniques like random rotations, translations, flips, and brightness adjustments. Data augmentation increased the diversity of our training dataset, leading to more robust models.

Loss functions: We designed custom loss functions tailored to our detection task, which incorporated both classification and regression components. These loss functions helped optimize the model's performance by penalizing false positives and false negatives differently.

The current study required modifications in the existing datasets, since they lack the potential information for detecting small obstacles in short ranges. However,

1) The dataset may not adequately capture rare or extreme events (e.g., uncommon traffic scenarios, severe weather conditions, or rare road obstacles).
2) The dataset may not account for the evolution of obstacles over time, such as changes in object behaviour or road conditions.
3) The dataset may over represent certain types of obstacles (e.g., vehicles) while underrepresenting others (e.g., animals or unusual objects).

The optimal dataset would require steering angle data from autonomous vehicles for the deep learning model to safely navigate after processing information from real time high resolution sensors. ZED stereo devices are mounted on vehicles in the Lost and Found Dataset, which would be a significant element for this research work. The obstacles of small size and shapes are annotated and mapped by MRF architecture[31], along with drivable area detection technique implemented in the deep learning model for safe navigation. The model processes the real time information captured from the sensors and advice the steering angle to maintain or modify for a hindrance free driving experience.

3. Proposed methodology: Multifusion perception technique

The proposed architecture contemplates a Multi-Fusion network, which is divided into three major modules, with respective functionalities. Various stochastic techniques including curvature potential, the gradient of grid cells and measuring the depth of each cell are potential elements of the Markov random field, to determine the obstacles present on the roads. The predominant three techniques analyse every image[32] with respect to pixel-based information to extract meaningful patterns and organize them for further orientation of the training model. In the proposed architecture, regions of interest are generated based on gradient information and depth variance of every pixel of captured images. The benefits of implementing AND Gates are analysed along with the functionalities of Markov random field technique. The second stage of this proposal concentrates on understanding the segmentation process is required for determining traffic lanes, drivable areas, and other filtering techniques required for outlier detection. The final model would integrate the prediction of changes required in the steering wheel angle for safe navigation without hitting the objectives[33]. The sensitivity of a sudden presence of an obstacle is measured by an Obstacle Closeness Factor (OCF) to portray a level of impact and accident risk. The Obstacle Closeness Factor (OCF) is a critical parameter in assessing the sensitivity of an autonomous vehicle's obstacle detection system to sudden obstacles and plays a significant role in enhancing safety during navigation. OCF is a metric that quantifies how close an obstacle is to the autonomous vehicle concerning its current speed, trajectory, and time to collision. It essentially measures the proximity of an obstacle in relation to the vehicle's ability to react and avoid a collision.

The architecture of the proposed method is depicted in **Figure 1**, where the different stages are split into different modules for their dedicated specificities. The input of the architecture is provided by optical encoders where the images are attained from two stereo vision sensors and the input image is processed for depth estimation. Semi global block matching algorithm[34] is implemented for measuring the depth of images obtained from stereo vision sensors. From these images, disparity variance, picture gradients, and depth curvature are calculated and presented for the next set of operations[35]. Altogether, these three parameters of combined into a unary cue to simplify the process of Markov random field and hence derive a pair wise potential element.

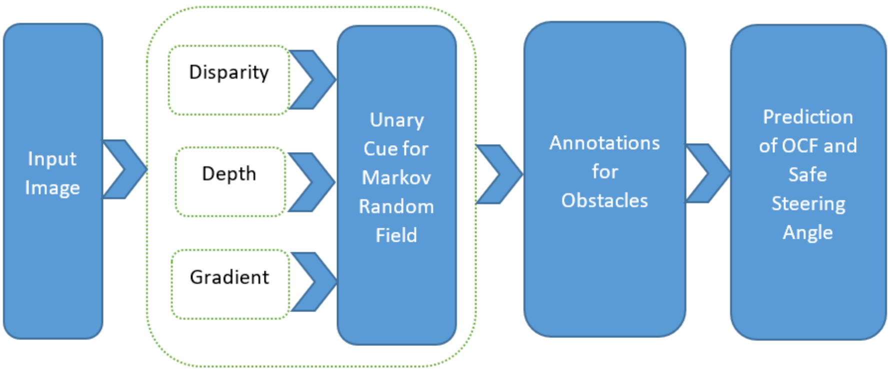


Figure 1. Architecture of the proposed model.

3.1. Image preprocessing using Markov random field technique

Different methods have implemented different forms of Markov random field due to its enhanced benefits of inference and learning methodologies. The various problems associated with long and short vision of the sensors have been removed when Markov random fields are implemented. Image restoration, construction of grid cells, edge detection, annotation, segmentation and various other processes are related to image and scene derivations. During this process, Markov random field has proven to be a suitable technique for deriving inferences. This approach concentrates on smaller and medium size obstacles, segmentation off images have to be derived with an additional cost function for Markov random things to operate with better performance.

Every image will be represented with various elements X_i spread over different random process of X_n. The position of individual elements is represented by *pos* which belongs to all the positions covered by all the elements P. Every pixel position will be indicated unary element and its association with other elements derived by the following cost function.

$$Cost(X) = \sum_{pos}^{P} Cost_u + \sum_{pos}^{P} Cost_{pair}\,(Xi, Xs) \tag{1}$$

where $Cost_u$ indicates the unary cost of individual elements and $Cost_{pair}\,(Xi, Xs)$ indicates the total cost of associated elements in the given space. $X_{i\ to\ n}$ represent the random variable associated with different nodes, which are identified by the texture, shape, size and ground truth of 0 or 1 is used to indicate whether this is an obstacle or not. Cost function this determined independently with respect to gradients, depth variance and curvature respectively. The gradient potential is represented by the following Equation (2). Whenever colour images are involved, partial derivatives are calculated in terms of horizontal and vertical axis. Image gradients are considered in this research rather than edge detectors tournament the need of thresholding whenever weaker gradient is detected. The proposed method builds stronger cues, implementing image gradients.

$$Cost_u^g\,(x, y) = \sqrt[2]{G_i\,(x, y)^2 + G_j\,(x, y)^2} \tag{2}$$

where $G_i\,(x, y)^2$ *and* $G_j(x, y)^2$ are considered to be horizontal and vertical partial derivatives respectively. In order to detect small obstacles, curvature is an important element to be considered and calculated along with gradient and depth. Curvature is defined as 1 for curves and 0 for straight lines, where the immediate curves indicate obstacles and 0 will indicate lanes or roads. The next parameter to be included is known as depth variant potential, to indicate the presence of any obstacle or changes in depth of pixel information. This is calculated by moving to the horizontal cell of the entire grid. Variance is measured by the following Equation (3) which is a representative form of multiple square Windows placed horizontally. Differences between each pixel are estimated by W_{hor} where the sudden changes are presented as depth.

$$Depth_{variance} = variance\left(\left[Depth\left(i - \frac{W_{hor}}{2}, j\right) : Depth\left(i + \frac{W_{hor}}{2}, j\right)\right]\right) \tag{3}$$

3.2. Multi-fusion perception

The previous module detected the presence of drivable areas, traffic objects which act as an obstacle and the present module we discuss how the Road conditions are perceived and the recommendations are given for safe navigation. The proposed architecture includes a context tensor for merging the inputs obtained from the previous module and attain a perception. The encoder is responsible for extracting meaningful information and patterns for detecting the presence of an obstacle. There are various convolutional neural networks presented in the literature for this purpose. The technique used in the proposed methodology is to implement a VGG16 and a feature pyramid map with deep learning modalities. The size of the input image would assume the size of the width, height, and channel. The recommended dimensions of the encoder would be W/16, H/16, 512. The connections between the layers are revoked and the purpose of this encoder is to differentiate the features between deep and shallow and hence that them into proper perceptions. Multi fusion network indicates that different layers would be merged together to consider both Deep and Shallow features together. Element Mean operation, for all X elements mentioned in the previous module, is proposed and hence the final feature map would be derived as the following dimensions W/4, H/4, 256.

The purpose of the decoder is to sense the drivable areas and represent them into zones between green or red lines after careful semantic segmentation. The original size of the input image will be recovered so that none of the important areas are eliminated from the consideration. The size of the drivable zones will be represented as W, H, 2. The important element of the proposed architecture is the context tensor where feature maps are merged together to derive the lanes and drivable areas. Drivable areas indicate that there is enough space for the vehicle to move without hitting any traffic objects, and it is also understood that even the drivable areas may process some obstacles such as traffic cones and pedestrians. In previous methodologies, curb detection has some important elements and usually gets confused with the lines drawn on the road for indicating different lanes. Using the context tensor, this approach has proven to eliminate the confusion and hence detect the lane lines with greater accuracy. Curvature cues, depth variance, gradient, and pairwise potential constituted the obstacle detection module, which was later merged into the context tensor for drivable area detection.

4. Results and discussions

There are very few data sets that have focused on obstacles of medium and small size and hence for the first module of obstacle detection, we have created our dedicated data set by retrieving nearly 5000 stereo images captured by ZED equipment, which are deployed on various vehicles in different environments. These sensors were able to capture 30 frames per second with a resolution of 1280 × 720. Another significant video database is the Berkeley deep drive dataset which contains videos of different environments conditions and vehicles and available open-source. Another significance in this data set is that the videos are highly annotated, with proper segmentation of drivable areas, objects on road, and lane marking. This data set has been assisting in various research purposes of autonomous driving and hence used in this research by considering 70 percent for training the model, 20% for testing, and the remaining 10% for validation of the proposed model.

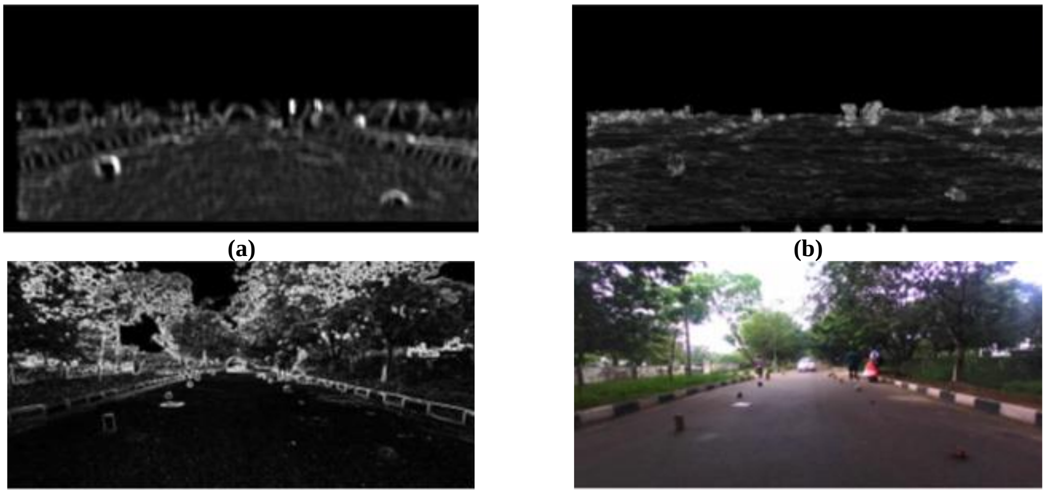
Table 1. Obstacle detection performance.

Parameter	DLT net	Multi net	ERF net	Proposed multifusion perception
Recall	89.4	81.6	77.2	91.2
AP	68.2	61.0	55.8	72.6
Speed	9.2	8.2	9.2	9.3

Table 2. Comparison of the proposed technique over the existing state of art techniques for lane detection.

Performance parameters	Our approach	YOLO (Obstacle Detection)	Faster R-CNN (Obstacle Detection)	Lane net (Lane Detection)
Real-time Performance	Yes	Yes	Not always	Yes
Detection Accuracy	High	Moderate	High	High
Sensor Fusion	Yes	No	No	No
Contextual Analysis	Yes	No	No	No
Efficiency	High	High	Moderate	Moderate
Adaptability	High	Limited	Limited	Limited
Lane Prediction	Yes	No	No	Yes
Semantic Context	Yes	No	No	Yes
Robustness	High	Limited	Limited	High

Parameters used for evaluating the performance of the model are mean intersection over Union and mean average precision. Intersection over Union indicates the performance in driving the area suitable for driving accurately and mean average Precision is used for detecting smaller and medium-sized obstacles. Multi Fusion Perception network has proven to achieve better results when compared to ERFNet which worked on cityscape data sets. Configuration of our simulation environment is an Intel(R) Xeon(R) E5-2630 v4, and our GPU is an NVIDIA GTX TITAN XP. According to **Table 1**, the performance of the Multifusion perception network has shown better results than ERFNet and DLT-Net without context tensor. Curvature cues, Depth variance, gradients measurement in the first module for object detection have increased the performance of the context sensor and hence delivered better results during semantic segmentation. The context tensor is a multi-dimensional data structure that encodes contextual information from various sources, such as environmental data, vehicle state, road conditions, and traffic patterns. This tensor is typically generated or updated at each time step as the vehicle navigates through its environment. By analyzing the context tensor, the system can predict the future trajectories of detected objects. This prediction is valuable for planning the vehicle's path and making decisions to avoid potential collisions. The runtime of MultiFusion perception network and ERFNet is almost similar and our model has shown betterments in the accurate detection of obstacles as listed in **Table 2**. The algorithm continues to learn under different scenarios and the final model has performed much better and faster due to better classification and regression strategies. **Figure 2** illustrates the performance of the proposed model, where **Figure 2a** illustrates the estimation of curvature cue, **Figure 2b** showcases how the model estimates the depth variance, **Figure 2c** estimates the image gradients and the decoded image is illustrated in **Figure 2d**. According to **Figure 3**, the model has successfully identified the drivable areas and **Figure 4** depicts the detection of other vehicles even during the low light conditions. **Figures 5** and **6** contemplate the collective performance of the proposed model, thereby addressing the driveable areas, lanes, and detected obstacles.



(a) (b) (c) (d)

Figure 2. (a) Estimation of curvature cue; **(b)** estimation of depth variance; **(c)** estimation of image gradients; **(d)** restored Image by the decoder.

Figure 3. Recognition of drivable Areas.

Figure 4. Detection of other vehicles.

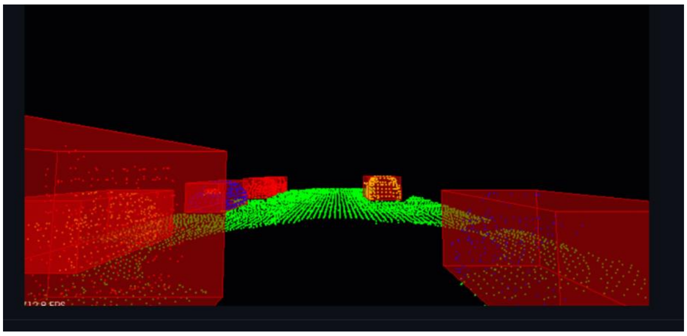


Figure 5. Green zones—Drivable areas, red zone—Other vehicles and obstacles.

Figure 6. Green zones—Drivable areas, red zone—Other vehicles and obstacles.

The challenges faced by the proposed approach are listed below.

Uncommon obstacles: Our system encountered difficulties when faced with uncommon obstacles, such as debris, animals, or objects not frequently encountered during training. These challenges highlight the importance of continually expanding the dataset to include a broader range of obstacles.

Fast-moving objects: Detecting and tracking fast-moving objects, such as motorcycles or high-speed vehicles, proved to be challenging. Improving the temporal resolution and prediction capabilities of our models is an ongoing focus.

Sensor failures: In cases of sensor failures or occlusions, the system's performance was adversely affected. Developing redundancy and sensor fault detection mechanisms is crucial for maintaining safety and reliability.

Sudden manoeuvres: The system faced difficulties in scenarios where other vehicles executed sudden manoeuvres, such as abrupt lane changes or hard braking. Improving the system's ability to anticipate and respond to such manoeuvres is an area of active research.

It is crucial to delve into the practical challenges and considerations associated with implementing our proposed obstacle detection system in real autonomous vehicles. These insights are valuable for the industry and can guide future research and development efforts. Here, we highlight some key practical challenges and considerations:

1) **Hardware Requirements:** Computational Power: Implementing our system in real autonomous vehicles requires substantial computational power. High-performance GPUs or specialized hardware accelerators may be necessary to process sensor data and run deep learning models in real-time. The effectiveness of our system relies on a diverse sensor suite, including cameras, LiDAR, radar, and more. Ensuring the reliability and calibration of these sensors is a complex task.
2) **Data Collection and Annotation:** Data Diversity: Building a comprehensive and representative dataset is challenging. It requires collecting data across a wide range of geographic locations, weather conditions, and traffic scenarios. Annotating data with precise labels is labour-intensive and requires domain expertise. Ensuring high-quality annotations for diverse obstacles is critical for model training.
3) **Regulatory compliance:** Compliance with Regulations: Autonomous vehicles must adhere to strict regulatory standards. Our system needs to meet safety and performance requirements set by regulatory bodies, which may vary by region.
    3.1) **Ethical dilemmas:** Autonomous vehicles may encounter ethical dilemmas in collision-avoidance situations. Decisions related to human safety vs. pedestrian safety, for instance, require careful consideration.
    3.2) **Liability and legal framework:** Developing a clear legal framework for liability in case of accidents involving autonomous vehicles is an ongoing challenge.
4) **Continuous learning and updating:** Continuous Improvement: Our system needs mechanisms for continuous learning and updating to adapt to evolving traffic conditions, new obstacle types, and emerging technology.

5. Conclusion and future work

The proposed method has shown significant presence in terms of obstacle detection in a complex environment and predicting the drivable areas the integrating Markov random fields. The process of obstacle detection can be challenging at times when there are sudden changes on roads caused by pedestrians, passers-by, or traffic elements. The process of autonomous vehicles begins with accurate detection of distance and altering the trajectory of the vehicle for safe navigation. The proposed deep learning model suggested the cost estimation of such obstacles and suggesting the probable areas without any accidents. On the whole, the proposed model has delivered nearly 30% of better analysis and prediction of obstacles irrespective of autonomous or human driving. Vision-based Navigation systems mentioned in the literature are bound to certain limitations which have been removed in this research work. High rate values are provided to sudden obstacles which are measured in close calls or short distances. The size of an obstacle is also a huge concern for accident-free navigation. In this regard, the proposed multi-fusion perception network combined different components of obstacle detection, lane detection, and safe navigation with a context tensor and suitable attested using Berkeley deep drive data set. Markov random field has significantly improved the detection accuracy of obstacles and other vehicles when compared to other state of art technologies.

Future work

Exploration of novel deep learning architectures tailored specifically for obstacle detection and lane tracking, with a focus on reducing computational requirements and enhancing accuracy is the future direction of the proposed research work. We shall investigate methods for handling occlusions and rare object types more effectively. We should be considering ethical decision-making frameworks for autonomous vehicles, addressing scenarios where safety considerations may conflict with legal or moral obligations. This approach has exhibited the modern architecture for intelligent transportation in smart cities. Focus on address privacy concerns related to sensor data collection and processing, implementing robust data security and anonymization measures should be given more emphasis in future.

Author contributions

Conceptualization, study conception and design, AKK; methodology, AKK; software, AKK; data collection and validation, AKK and VP; formal analysis, AKK; investigation, AKK; resources, AKK; data curation, AKK; writing—original draft preparation, AKK; writing—review and editing, AKK; visualization, AKK and VP; supervision, AKK and VP; project administration, AKK; funding acquisition, VP. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Kumar AK, Venkatesh P. A survey on handling events, localization and perception of autonomous vehicles in real time scenarios. Design Engineering. 2021, 2021(6): 7323-7337.
2. Wang Q, Zheng J, Xu H, et al. Roadside Magnetic Sensor System for Vehicle Detection in Urban Environments. IEEE Transactions on Intelligent Transportation Systems. 2018, 19(5): 1365-1374. doi: 10.1109/tits.2017.2723908
3. Qian Y, Yang M, Wang C, et al. Pedestrian Feature Generation in Fish-Eye Images via Adversary. 2018 IEEE International Conference on Robotics and Automation (ICRA); May 2018. doi: 10.1109/icra.2018.8460565
4. Tran N. Global Status Report on Road Safety. World Health Organization: Geneva, Switzerland; 2018. pp. 5–11.
5. Jeppsson H, Östling M, Lubbe N. Real life safety benefits of increasing brake deceleration in car-to-pedestrian accidents: Simulation of Vacuum Emergency Braking. Accident Analysis & Prevention. 2018, 111: 311-320. doi: 10.1016/j.aap.2017.12.001
6. Tet BC. The Prevalence of Motor Vehicle Crashes Involving Road Debris, United States, 2011–2014. Age 2016, 20, 10-1.
7. Xu X, Fan CK. Autonomous vehicles, risk perceptions and insurance demand: An individual survey in China. Transportation Research Part A: Policy and Practice. 2019, 124: 549-556. doi: 10.1016/j.tra.2018.04.009
8. Guo Y, Xu H, Zhang Y, et al. Integrated Variable Speed Limits and Lane-Changing Control for Freeway Lane-Drop Bottlenecks. IEEE Access. 2020, 8: 54710-54721. doi: 10.1109/access.2020.2981658
9. Puglia L, Brick C. Deep learning stereo vision at the edge. Available online: http://arxiv.org/abs/2001.04552v1 (accessed on 29 November 2023).
10. Azeta J, Bolu C, Hinvi D, et al. Obstacle detection using ultrasonic sensor for a mobile robot. IOP Conference Series: Materials Science and Engineering. 2019, 707(1): 012012. doi: 10.1088/1757-899x/707/1/012012
11. Suzen AA, Duman B, Sen B. Benchmark Analysis of Jetson TX2, Jetson Nano and Raspberry PI using Deep-CNN. In: Proceedings of the 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA).
12. Ou X, Yan P, Zhang Y, et al. Moving Object Detection Method via ResNet-18 With Encoder–Decoder Structure in Complex Scenes. IEEE Access. 2019, 7: 108152-108160. doi: 10.1109/access.2019.2931922
13. Orozco-Rosas U, Picos K, Montiel O, et al. Environment Recognition for Path Generation in Autonomous Mobile Robots. Studies in Computational Intelligence. 2019, 273-288. doi: 10.1007/978-3-030-34135-0_19
14. Pan X, Shi J, Luo P, et al. Spatial as Deep: Spatial CNN for Traffic Scene Understanding. Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1). doi: 10.1609/aaai.v32i1.12301
15. Fu C, Hu P, Dong C, et al. Camera-Based Semantic Enhanced Vehicle Segmentation for Planar LIDAR. 2018 21st International Conference on Intelligent Transportation Systems (ITSC); November 2018. doi: 10.1109/itsc.2018.8569413
16. Caltagirone L, Scheidegger S, Svensson L, et al. Fast LIDAR-based road detection using fully convolutional neural networks. 2017 IEEE Intelligent Vehicles Symposium (IV); June 2017. doi: 10.1109/ivs.2017.7995848
17. Mehmood A, Liaquat M, Bhatti AI, et al. Trajectory Planning and Control for Lane-Change of Autonomous Vehicle. 2019 5th International Conference on Control, Automation and Robotics (ICCAR); April 2019. doi: 10.1109/iccar.2019.8813737
18. Ko W, Chang DE. Cooperative adaptive cruise control using turn signal for smooth and safe cut-in. In: Proceedings of the 2018 18th International Conference on Control, Automation and Systems (ICCAS); 17–20 October 2018; Yongpyong Resort, Seoul, Korea.

19. He K, Gkioxari G, Dollár P, et al. “Mask R-CNN,” in Proc. ICCV, Oct. 2017, pp. 2961–2969.
20. Orozco-Rosas U, Picos K, Montiel O. Acceleration of Path Planning Computation Based on Evolutionary Artificial Potential Field for Non-static Environments. Studies in Computational Intelligence. 2020, 271-297. doi: 10.1007/978-3-030-35445-9_22
21. Rawat P. Environment Perception for Autonomous Driving: A 1/10 Scale I Implementation of Low Level Sensor Fusion Using Occupancy Grid mapping [Master’s thesis].
22. Inthanon P, Mungsing S. Detection of Drowsiness from Facial Images in Real-Time Video Media using Nvidia Jetson Nano. In: Proceedings of the 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON).
23. Romera E, Alvarez JM, Bergasa LM, et al. ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation. IEEE Transactions on Intelligent Transportation Systems. 2018, 19(1): 263-272. doi: 10.1109/tits.2017.2750080
24. Teichmann M, Weber M, Zollner M, et al. MultiNet: Real-time Joint Semantic Reasoning for Autonomous Driving. 2018 IEEE Intelligent Vehicles Symposium (IV); June 2018. doi: 10.1109/ivs.2018.8500504
25. Katare D, El-Sharkawy M. Embedded system enabled vehicle collision detection: An ANN classifier. In: Proceedings of the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC); 7–9 January 2019; Las Vegas, NV, USA.
26. He D, Zou Z, Chen Y, et al. Obstacle detection of rail transit based on deep learning. Measurement. 2021, 176: 109241. doi: 10.1016/j.measurement.2021.109241
27. Che E, Jung J, Olsen M. Object recognition, segmentation, and classification of mobile laser scanning point clouds: A state of the art review. Sensors. 2019, 19(4): 810. doi: 10.3390/s19040810
28. Islam K, Wijewickrema S, Raj R, et al. Street sign recognition using histogram of oriented gradients and artificial neural networks. Journal of Imaging. 2019, 5(4): 44. doi: 10.3390/jimaging5040044
29. Chen C, Liu B, Wan S, et al. An Edge Traffic Flow Detection Scheme Based on Deep Learning in an Intelligent Transportation System. IEEE Transactions on Intelligent Transportation Systems. 2021, 22(3): 1840-1852. doi: 10.1109/tits.2020.3025687
30. Zhao J, Liu J, Yang L, et al. Future 5G-oriented system for urban rail transit: Opportunities and challenges. China Communications. 2021, 18(2): 1-12. doi: 10.23919/jcc.2021.02.001
31. Wang S, Yang S, Zhao C. SurveilEdge: Real-time Video Query based on Collaborative Cloud-Edge Deep Learning. IEEE INFOCOM 2020 IEEE Conference on Computer Communications; July 2020. doi: 10.1109/infocom41043.2020.9155284
32. Yang Z, Zhang Y, Yu J, et al. End-to-end multi-modal multi-task vehicle control for self-driving cars with visual perceptions. In: Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR); 20–24 August 2018; Beijing, China.
33. Han X, Lu J, Zhao C, et al. Semisupervised and Weakly Supervised Road Detection Based on Generative Adversarial Networks. IEEE Signal Processing Letters. 2018, 25(4): 551-555. doi: 10.1109/lsp.2018.2809685
34. Wan S, Ding S, Chen C. Edge computing enabled video segmentation for real-time traffic monitoring in internet of vehicles. Pattern Recognition. 2022, 121: 108146. doi: 10.1016/j.patcog.2021.108146
35. Cong Z, Li X. Track obstacle detection algorithm based on YOLOv3. In: Proceedings of the 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI).