An Improved Traffic Sign Recognition and Road Lane Detection for Self-Driving Cars using YOLO-V8

Divya Elankumaran

School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

Abstract

Autonomous vehicles, often known as self-driving cars, are a paradigm-shifting breakthrough that has gained a lot of attention in recent years. They have the potential to transform mobility by offering safer and more efficient routes. This paper presents a unique way for improving the efficiency and efficacy of self-driving cars through enhanced traffic sign recognition and road lane detection using the deep learning framework. The YOLOv8 model is employed because of its superior object detection capability, which allows the model to process in a single run. The selected solution demonstrates good performance across a wide range of environmental challenges by thoroughly training the YOLO model on a massive dataset including various lighting factors, weather scenarios, and other aspects. To train the model for sign recognition, a huge number of varying traffic sign images is utilised. The model's effectiveness at recognising and categorising multiple traffic signs in a range of circumstances, including low-light and severe weather, has been enhanced. Considering YOLOv8's versatility, the model is trained on a wide range of road lane video datasets, ensuring exact vehicle localization between lanes. Deep convolutional neural network (DCNN) based on residual network 50 (resnet) architecture for sign and lane identification, as well as you only look once (YOLOv8), an advanced CNN technique for real-time object detection, were used to accomplish the proposed model. Furthermore, the suggested approach produces feedback for observed traffic signals via voice. The results show substantial rises in both traffic sign recognition and roadlane detection. Even in challenging conditions, the proposed method achieves over 99% accuracy for traffic signs and over 97% accuracy for lane recognition.)

Keywords: Self-driving cars, traffic sign recognition, road-lane detection, YOLOv8, Deep convolutional neural network, Resnet

1.Introduction

Autonomous or autonomous vehicles, often known as self-driving cars, are an innovative means of transport with the potential to revolutionise the face of transportation. Unexpected occurrences arise as a consequence of poor road maintenance, inclement weather, misinterpretation of traffic signs due to foggy conditions, faded or occluded boards, and other reasons that leave drivers unable to notice the appropriate sign boards. According to study, 1.2 million people die on the pavements each year, with an additional 20 to 50 million suffering the repercussions of nonfatal injuries. Certain safety measures have been implemented in order to reduce collisions caused by traffic and road lanes. Artificial intelligence and computer vision advancements over the previous decade have led to the creation of self-driving car.

Self-driving automobiles have far-reaching implications for society, transportation, and the economy. One of the primary promises of self-driving cars is a considerable reduction in traffic accidents and fatalities. They allow those who are unable to drive due to age, disability, or other circumstances to travel [2]. They may optimise driving actions, perhaps leading in reduced fuel use and emissions. The introduction of self-driving vehicles signalled the beginning of a new era in the automobile industry, resulting in safer and more efficient transportation. The ability of self-driving automobiles to perceive and interpret the complicated nature of their driving environment,which plays an essential role to their success [3]. Identification of traffic signs and recognition of roadways are key features of this perceptual system for ensuring reliability and security. In this scenario, employing skilled deep learning techniques, namely the YOLOv8 algorithm, provides a significant step towards enhanced accuracy and real-time processing capability. This research contributes to the advancement of self-driving automotive technology by providing a comprehensive and efficient perception system. Furthermore, text is transformed to voice for the detected traffic sign, together with the warning sound. The goal of this study is to dig into the complexities of applying YOLOv8 for traffic sign identification and road lane detection, solving existing issues in accuracy, efficiency, and flexibility.

Deep learning is essential in self-driving automobiles because it enables them to comprehend and study their environment. It is used in real time to recognise, categorise, and track things. It also assists in the construction of high-definition maps and precise localization, letting the vehicle to understand its precise position and orientation in relation to its surroundings. Yolo-v8, which stands for "You Only Look Once version 8," is a cutting-edge real-time object detection technique.This approach comprises fine-tuning the model with a specially chosen dataset that contains a range of traffic sign pictures from real-world environments. This unique learning mechanism enables the system to recognise and respond to a broad variety of traffic sign types and forms. Furthermore, the critical issue of road lane recognition is addressed, which is a need for autonomous vehicles to maintain safe and precise trajectories. One of the primary features of yolov8 is the stage detection technique, which is meant to identify objects in real time and with high accuracy. ResNet 50 is a deep neural network (DNN) that is widely recognised for its capacity to train exceptionally deep networks capable of producing cutting-edge results on a variety of applications such as image classification, object recognition, and natural language processing. The core notion of ResNet is the addition of residual blocks. Instead of the direct mapping from input to output, ResNet identifies the residual mapping. The utilisation of skipped connections, which are connections that bypass one or more network levels, is ResNet's fundamental innovation. Skipping connections allows the network to learn about residual functions, which are functions that are added to a layer's input for generating output. Deep Convolutional Neural Networks (DCNN or ConvNet) are a type of artificial neural network designed primarily for image recognition and computer vision applications. DCNN operates by converging a series of filters on an input image.

2.Related Works

Using a convolutional neural network, Hee Seok Lee et al. [1] introduced a unique method for concurrently detecting traffic signals and estimating their precise bounds. The suggested method frames the boundary estimation problem as a 2-D posture and shape class prediction task that a single CNN can address effectively. To do this, the object bounding box detection problem is enlarged and transformed into an object pose estimation problem, which is efficiently described using CNN based on recent

advances in object detection networks. Jigang Tang a b et al. [2] conducted research on a review of deep learning-based lane-detecting algorithms. Lane detection is an environmental perception application that detects lane areas or lane lines using a camera or lidar. The first attempt is made here to offer a comprehensive review of vision-based lane detection systems. The history of lane detection is initially explored, followed by typical lane identification methodologies and accompanying deep learning methods. Second, categorise current lane identification algorithms as two-step or one-step solutions. The lane-detecting technique is presented from two angles in the following overview. The first is network designs, which include classification and object recognition methodologies, end-to-end segmentation of images methods, and specific optimisation algorithms, and the second is the loss function, which is connected to network frameworks. Jia Li et al. [3] presented in-field real-time traffic sign identification using efficient CNNs. The architecture of faster R-convolutional neural networks (CNN) and the MobileNet structure were used to design and construct a detector. Colour and shape information were used to enhance the localization of small traffic signs that are difficult to regress accurately. Finally, as the traffic sign classifier, an efficient CNN with asymmetric kernels is used. Mohammed Gad et.al[4] investigated real-time instance segmentation using segnet and image processing. An encoder-decoder deep learning architecture is used in this approach to construct binary segmentation of lanes, which is then processed to separate lanes, and a sliding window extracts each lane to provide the lane instance segmentation image. This method was tested using a simple data set and yielded competitive results. Mohammed Ikhlayel et al. [5] proposed a method to identify traffic signs for the navigation of an autonomous automobile prototype employing CNN. In this work, the autonomous car may navigate based on the traffic sign that it detects. An autonomous car will recognise and classify traffic signs depending on their function using a video sensor and deep learning techniques. Based on this classification, the autonomous car will respond as indicated by the observed traffic sign and will activate the actuator.

3.Methodology

Preliminary data processing, feature extraction, and classification are critical aspects in implementation. You Only Look Once, or YOLO for short, is a technique for object recognition and segmentation that may provide very advanced training and performance on a single GPU. YOLOv8 covers a wide range of vision AI tasks, including image classification, segmentation, object identification, and posture estimation. The main purpose of choosing this technology is to build a unique but effective perception platform for self-driving cars that includes enhanced traffic sign recognition and precise road lane detection via the YOLOv8 algorithm. Yolov8 models are pretrained on COCO dataset and re-trained when new data is trained using the model. Yolov8 deployment is ideal for lane and sign recognition since it is capable of identifying numerous elements in a single frame of an image or video frame, and the found objects are classed according on their features. This type of functionality is also appropriate for small-scale applications.
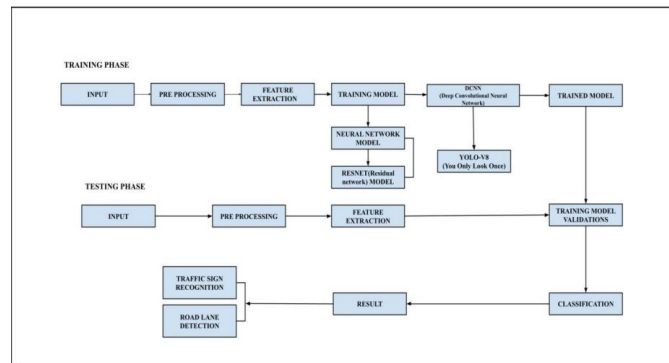
Fig.1 Overall Architecture

A. Dataset Collection and Input

Data collection is a critical step in training and testing models. A massive collection of traffic sign images and footage of road lanes is acquired from the "Kaggle" dataset repository. Collected a large dataset depicting an array of real-life driving scenarios. Captures that depict a wide range of illumination and climate conditions (plain, wet, and fog), as well as diverse road types, are included,ensuring that the collection uses a variety of traffic indications. "PNG" or "JPG" are the image input formats. Similarly, video is available in "MP4" and "AVI" formats. The image is provided as input for the traffic sign, which covers various lighting and weather circumstances such as rain, fog, and so on. More than 1500 image datasets are utilised for sign identification, and a video dataset is uploaded for each traffic lane. To select the input image, the tkinter dialogue box is used. In each image, create boundary boxes for signage and lane markings,indicating accurate coordinates (x, y, width, and height) for each marked object.

B. Preprocessing

Preprocessing is an important step in deep learning processes because it improves model performance and minimises the likelihood of overfitting. Preprocessing methods are often employed to clean, modify and prepare data for analysis, making it more suitable for deep learning model training and evaluation. The image is scaled to 224×224 in this process, and the video is then converted into frames, which are then resized to 512 x 512. The initial image is an RGB image and a frame is converted grayscale image. It does not alter the original picture during conversion, instead generating a new image with the updated dimensions. Preprocessing results in a more precise and reliable model.

C. Feature Extraction

The process of choosing and transforming raw data into a constrained and equivalent set of attributes that represent the core aspects of the data is referred to as feature extraction. The pre-processed image is used to extract relevant characteristics. The pre-processed image is used to determine the mean, median, and variance. Local Binary Pattern (LBP) is a useful texture descriptor for images that thresholds nearby pixels based on the value of the current pixel. LBP descriptors can easily capture local spatial patterns and gray-scale contrast in an image. The LBP descriptor's base unit is 3×3 pixel blocks. The difference between the central pixel and its neighbouring 8 pixels is produced as a local texture feature representation. The pro-processed frames are blurred to decrease noise, and canny edge detection is

applied to determine the lanes' boundaries. This approach is commonly used to detect the edges of objects in images or frames.

D. Image Splitting

Image splitting is a machine learning approach that divides a dataset into smaller groups for training and testing. This is done to increase the model's efficiency and accuracy. The Classificationof the dataset that is used to evaluate the model's performance. Because the model was not trained on the testing set, its performance on it is a measure of its ability to generalise to new data. The dataset is randomly divided in an 80:20 ratio, with 80% being training data and 20% being testing data. Training is used to train the model, while testing is used to evaluate the model's performance. Splitting the dataset into training and testing sets is essential for preventing over-fitting, improving model accuracy, and comparing the model's performance on the testing set against other models or various training approaches.

E. Classification

A mixture of deep learning algorithms aids in the identification of traffic signs and road lanes. Resnet-50 is a residual network with 50 layers. A neural network model is trained using resnet by residual blocks in this instance. Resnet connections are skipped or shortcuts are added. These shortcuts allow the gradient to skip over particular layers during training, making it easier to train extremely deep networks. Yolo-v8 is employed for detecting objects since it allocates the image into grids and immediately predicts the bounding boxes and class probabilities for each grid in a single forward pass. A confidence score and class label are assigned to each detection. Non-maximum suppression is used to reduce duplicate bounding boxes, retaining just the most confident boxes. It entails optimising a loss function that makes inaccurate predictions of bounding box coordinates and class probabilities during training. Deep convolutional neural networks (DCNN) are used to classify and find patterns in images and videos. DCNN are made up of numerous convolutional layers that are followed by one or more fully connected layers. Each convolutional layer creates a feature map by applying a series of filters to the input image. The resulting feature maps are then transferred to the next convolutional layer for further processing. The fully connected layers are in charge of classifying the image based on the convolutional layers' retrieved features.

F. Performance Analysis

The practise of analysing how well a system or process works against set criteria is known as performance analysis. The confusion matrix is used to generate performance measures like as accuracy, precision, recall, and f1-score. For traffic signs, the sign is detected and instructions are delivered based on the sign utilising speech conversion and the sign is highlighted using the bounding box. The road lane is recognised by identifying the lanes, and general instructions are provided. Using yolov8, achieved the greatest accuracy of 99.4% for traffic sign recognition and 97.4% for road lane detection for self-driving automobiles.

4. Results and Discussion

When the model is deployed in a streamlit application with a user interface, either traffic sign or road lane detection can be selected. In traffic sign identification, an image with various illumination settings, weather conditions, road surroundings, and so on is uploaded as input Fig.2. After uploading the image,

it proceeds through various processing phases before displaying the performance data, as well as the detected traffic sign and directions for that sign. Signs are also indicated using the bounding box. Furthermore, the command is transformed to speech so that it may be delivered via voice output and a warning sound, making it easier to grasp. It processes the frames at several levels and outputs metrics along with instructions. For detecting the lanes, the result is converted to video. The accuracy of traffic sign recognition was 99.4%, while the accuracy of road lane detection was 97.9%.
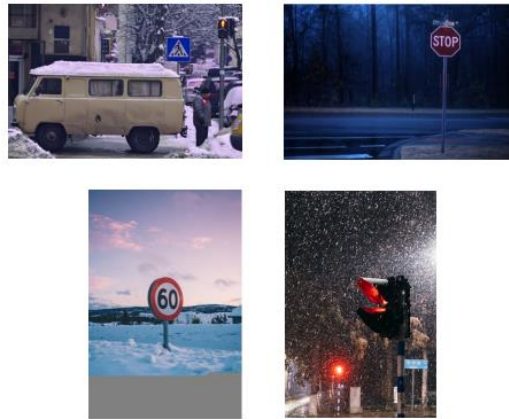
Fig.2 Input Traffic Sign Images
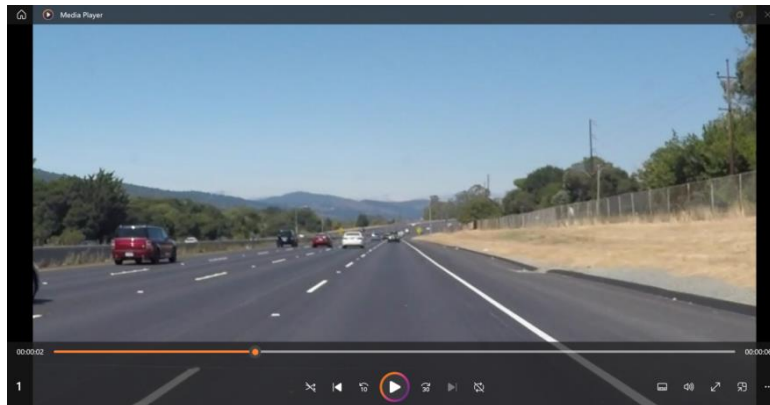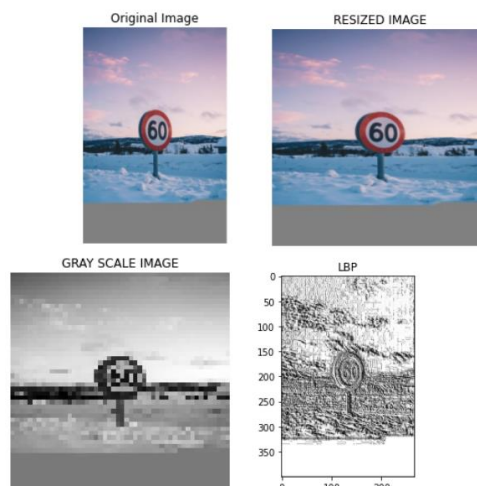

Fig.3 Input Video of Road Lane
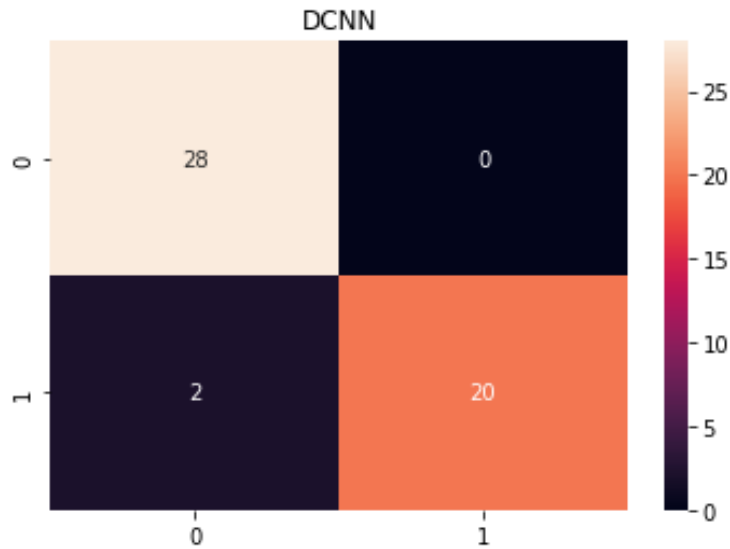

Fig.4 Processed Images

Fig.5 Confusion Matrix

The provided confusion matrix offers insights and summarises the performance of a classification model. The confusion matrix contains several important terms.

True positives: Instances that are appropriately classed as positive.

False positives: Instances are wrongly labelled as positive when they are in fact negative.

False negatives: When an instance is wrongly classified as negative when it is actually positive.

True negatives: Instances are appropriately classed as negative.

Accuracy = TP + TN/TP + TN +FN+FP

Precision = (TP + FP) / (TP + FP)

TP / (TP + FN) = recall

F1 Score = 2 * (Precise + Recall) / (Precise + Recall)

When training a model, it is usual practise to track both training and validation accuracy, as well as training and validation loss. These metrics will demonstrate how effectively the model learns from training data and how well it generalises to new data. These measurements are visualised throughout epochs (training iterations spanning the full dataset). Training and validation accuracy are represented on one graph, and training and validation loss are plotted on another. The percentage of correctly categorised occurrences on the training dataset is referred to as training accuracy. It demonstrates how well the model learns from the training data supplied. The rate of properly categorised occurrences on a different validation dataset is measured by validation accuracy. The training loss is the difference between the actual and predicted outcomes during training. It measures how well the model fits the

training data. The validation loss, like the training loss, reflects the difference between the model's actual and predicted values on the validation dataset.This provides information on how effectively the model generalises to new data. If the model's training accuracy is high but its validation accuracy is poor, it may be overfitting. Overfitting happens when a model gets very specialised in training data and fails to generalise well to new data. If the model's training and validation accuracy are both poor, it is underfitting, which suggests it hasn't learnt the common trends in its training effectively.
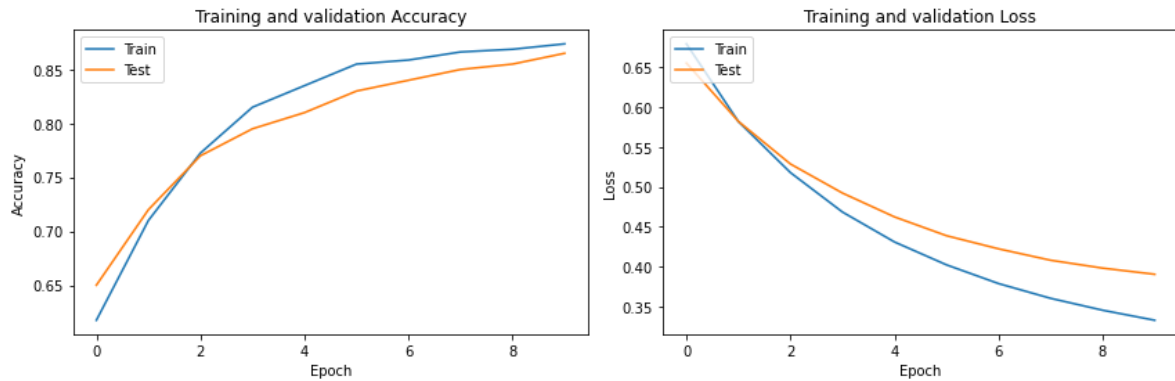

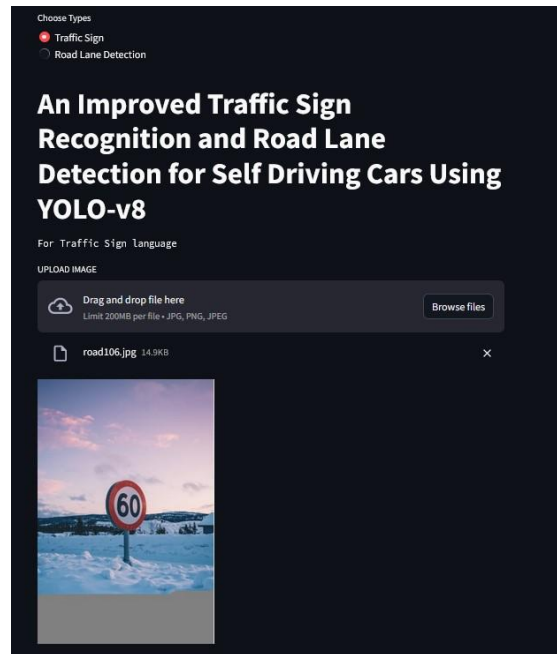Fig.6 Training vs Validation Loss and Accuracy
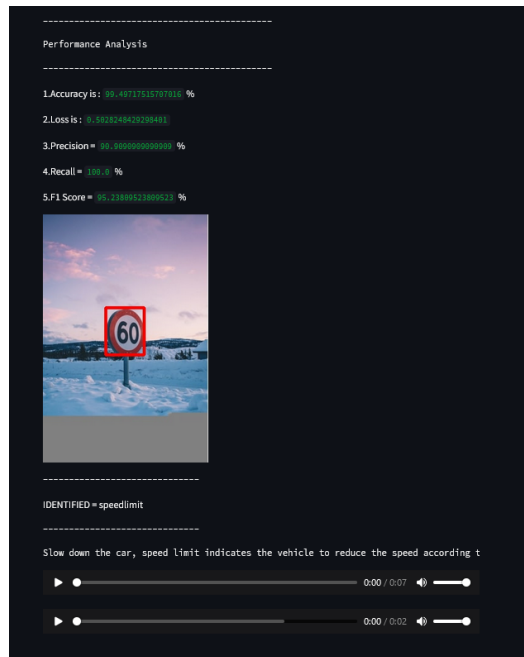

Fig.7 User Interface Output-1

Fig.8 User Interface Output-2

Fig.9 User Interface Output-3

5. Conclusion and Future Works

Furthermore, employing YOLOv8, this study proposes a novel approach for sign recognition and lane detection, displaying considerable increases in both accuracy and processing time. For sign and lane detection, the integration of yolo-v8 and resnet algorithms beat the other models in terms of accuracy and speed. These studies demonstrate that advances in deep learning and computer vision can give trustworthy perception for self-driving cars. The proposed technique offers a high level of assurance for increasing the security and dependability of self-driving automobiles, which will aid in the recognition of autonomous vehicles in a wide range of driving scenarios.The results show considerable improvements in both traffic sign recognition and road lane detection. According on the current study, several future upgrades for improved performance may be explored. Performing data augmentation, which might be done by flipping, rotating, scaling, and rotating to artificially increase the size of the

training dataset. This would most likely allow the sign to be recognised even if it was oriented differently. Using a pre-trained model and the transfer learning approach on a huge dataset. In the future, the attention mechanism might be used to create the desired consequences. Using online learning techniques to progressively update the model as new data becomes available, reducing the requirement to retrain the entire model from scratch.To improve model performance, hyper-parameter modifications can be applied. To assure the reliability and safety of the self-driving car system, use a realistic test dataset and real-world situations to thoroughly validate any upgrades. In order to boost performance, any additional algorithms and distinct models might be merged.

References

1. Lee, H.S. and Kim, K.,2023.Simultaneous traffic sign detection and boundary estimation using convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, *19*(5), pp.1652-1663.

2. Tang, J., Li, S. and Liu, P.,2022. A review of lane detection methods based on deep learning. *Pattern Recognition*, *111*, p.107623.

3. Li, J. and Wang, Z., 2022. Real-time traffic sign recognition based on efficient CNNs in the wild. *IEEE Transactions on Intelligent Transportation Systems*, *20*(3), pp.975-984.

4. Gad, G.M., Annaby, A.M., Negied, N.K. and Darweesh, M.S., 2022. Real-time lane instance segmentation using segnet and image processing. In *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)* (pp. 253-258). IEEE.

5. Ikhlayel, M., Iswara, A.J., Kurniawan, A., Zaini, A. and Yuniarno, E.M., 2022, November. Traffic sign detection for navigation of autonomous car prototype using convolutional neural network. In *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)* (pp. 205-210). IEEE.

6. Triki, N., Karray, M. and Ksantini, M., 2023. A real-time traffic sign recognition method using a new attention-based deep convolutional neural network for smart vehicles. *Applied Sciences*, *13*(8), p.4793.

7. Singal, G., Singhal, H., Kushwaha, R., Veeramsetty, V., Badal, T. and Lamba, S., 2023. RoadWay: lane detection for autonomous driving vehicles via deep learning. *Multimedia Tools and Applications*, *82*(4), pp.4965-4978.

8. Fredj, H.B., Chabbah, A., Baili, J., Faiedh, H. and Souani, C., 2023. An efficient implementation of traffic signs recognition system using CNN. *Microprocessors and Microsystems*, *98*, p.104791.

9. Zhang, Y., Lu, Y., Zhu, W., Wei, X. and Wei, Z., 2023. Traffic sign detection based on multi-scale feature extraction and cascade feature fusion. *The Journal of Supercomputing*, *79*(2), pp.2137-2152.

10. Sharma, V.K., Dhiman, P. and Rout, R.K., 2023. Improved traffic sign recognition algorithm based on YOLOv4-tiny. *Journal of Visual Communication and Image Representation*, *91*, p.1037