International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Traffic Sign Board Recognition and Voice AlertSystem using Convolutional Neural Network

[1] Dr M Prasad, [2]Darshan L
[1]Associate Professor, Department of MCA, BIET, Davanagere
[2]Final Year Student, Department of MCA, BIET, Davanagere
[1]Prasadmarpalli52@gmail.com
[2] ldarshan1999@gmail.com

***Abstract*—To ensure a smooth and secure flow of traffic, road signs are essential. A major cause of road accidents is negligence in viewing the Traffic signboards and interpreting them incorrectly. The proposed system helps in recognizing the Traffic sign and sending a voice alert through the speaker to the driver so that he/ she may take necessary decisions. The proposed system is trained using Convolutional Neural Network (CNN) which helps in traffic sign image recognition and classification. A set of classes are defined and trained on a particular dataset to make it more accurate. The German Traffic Sign Benchmarks Dataset was used, which contains approximately 43 categories and 51,900 images of traffic signs. The accuracy of the execution is about 98.52 percent. Following the detection of the sign by the system, a voice alert is sent through the speaker which notifies the driver. The proposed system also contains a section where the vehicle driver is alerted about the traffic signs in the near proximity which helps them to be aware of what rules to follow on the route. The aim of this system is to ensure the safety of the vehicle's driver, passengers, and pedestrians.**

Keywords—Convolutional Neural Network, GTSRB Dataset, Object Detection, Object Classification Traffic, Traffic Signs, Voice Alert

1 INTRODUCTION

There have been a lot of technological advancements and cars with auto-pilot mode have come up. Autonomous vehicles have come into existence. There has been a boom in the self driving car industry. However, these features are available only in some high end cars which are not affordable to the masses. We wanted to devise a system which helps in easing the job of driving to some extent.

On conducting a survey we found that the magnitude of road accidents in India is alarming. Reports suggest that every hour there are about 53 mishaps taking place on the roads. Moreover, every hour more than 16 deaths occur due to these mishaps [18]. When someone neglects to obey traffic signs while driving, they are putting their life as well the life of the other drivers, their passengers and those on the road at risk. Hence, we came up with this system in which traffic signs are automatically detected using the live video stream and are read out aloud to the driver who may then take the required decision. Another area of focus in our system is the idea of getting the location of the user using GPS. Also, all the traffic signs will be stored in a database along with their location so that the driver will be notified in advance regarding the next approaching Traffic Sign.

The following is a breakdown of the paper's structure: Section II summarises the literature review; Section III explains the technique and how the models work; and Section IV displays the Results and Analysis. The paper's conclusion is in Section V, while the Future Scope is in Section VI.

I. LITERATURE SURVEY

In this era of a fast paced life, people generally tend to miss out on recognizing the traffic sign and hence break the rules. A lot of research has been done in this domain in order to reduce the number of accidents. Researchers have used a variety of classification algorithms and a number of CNN architectures to classify the traffic signs and alert the driver. Our system aims to optimize the process of recognition and at the same time provide other benefits such as early alert to the driver.

The detection of traffic signs has been done in a variety of techniques in numerous studies. [1]One of the processes employs the Support Vector Machine technique. The dataset was divided into 90/10 for training and testing purposes, and it employs linear classification. To achieve the desired result, a series of phases called Color Segmentation, Shape Classification, and Recognition were followed.

Raspberry Pi is used in detecting and recognizing Traffic Signs with much less coding [2]. However, it requires the Raspberry Pi board at one's discourse for implementation which is quite costly. Another way of

Traffic sign recognition is picture intensive [3]. A video is acquired and broken down into frames. Image preprocessing is done which includes separating the foreground and the background, thinning and contrast enhancement.The signs are then categorised as hexagonal, triangular, or circular in shape and transmitted for template matching after these operations. The objects with some definite shape are matched from the pretrained algorithm.

Caffe, an open source system, that helps to detect and recognise road traffic signs with high accuracy and efficiency[4]. A CNN approach is proposed for training traffic sign training sets and obtaining a model that can categorise traffic signs. Another method for using the CNN scheme is proposed in [11], in which the actual border of the goal sign is estimated by projecting the boundary of a corresponding template sign image into the input picture plane. The method advances to become end-to-end trainable when we transform the boundary estimation problem into a pose and shape prediction job based on CNN. It is more resistant to occlusion and narrow objectives than other boundary estimating techniques that focus on contour estimation or image segmentation.

[6] proposes a multi-resolution feature fusion network architecture for sign detection, which aids in the separation of numerous small objects from sign boards. A vertical spatial sequence attention (VSSA) module can also be used to gather extra context information for improved detection. Using GPS-based tracking, Augmented Reality technology is incorporated in mobile apps [5]. It uses the coordinates of a user's smartphone as a pointer to assist people in dynamically and simply locating possible resources in the immediate vicinity based on the direction of the user's camera view.

In [7] AlexNet structure of CNN is used in which the architecture contains eight layers. The first five layers are convolutional layers and the latter three are all connected layers. The accuracy of this architecture comes out to be 92.63%. Also, the GoogleNet architecture is implemented in [7] which helps in working with large data and a high number of parameters. However it comes with an issue that the large data causes network overfitting decreasing the accuracy to just 80.5%. VGG CNN is proposed in [8] which has a significantly superior performance as compared to other available architectures. The number of parameters in this approach is considerably reduced in order to optimise and speed up the calculation. The network also includes the BN (batch normalisation) and GAP (global average pooling) layers, which help to improve accuracy without increasing the number of parameters. However, we discovered in [10] that by deleting the pool4 layer of VGG16 and using dilation for ResNet, we can combine the improved architecture of Faster-RCNN with Online Hard Examples Mining (OHEM), making the system more resilient and assisting in the detection of minor traffic signs.

Chuanwei Zhang et al. [9] suggested a traffic sign recognition method based on an improved Lenet-5 network. In this method, the Lenet-5 CNN model is used, which allows for overall network improvements. The improved Lenet-5 classifier outperforms the convolutional neural network and the classic Gabor and SVM classifier in terms of accuracy and real-time performance. The developers of [12] propose a traffic sign identification system based on CNN. They used CNN as a feature extractor and MPPs as an effective classifier to predict category codes. Using MPPs greatly increased the precision of recognition.

We find almost a gist of all the above papers in as it presents a mini-batch proposal selection mechanism in combination with a deep hierarchical architecture that allows the neural network to detect the traffic signs as well as the traffic lights by training them on separate datasets. The method resolves the problem of instances from one dataset not being labelled in the other. The system helps in giving a new dimension to our project by providing the idea of traffic sign localization for driver assistance [15]. The position of the traffic sign can be determined with a 1 metre accuracy using a single colour camera and a high precision GNSS (global navigation satellite systems) receiver. Another use of GPS as suggested in [16], is determining the style of driving where GPS data is collected from the mobile phone of a person while also detecting the Traffic Signs in the locality. It helps in classifying the style of driving as safe or aggressive.

II. METHODOLOGY

A. *Dataset*

In the proposed system, the German Traffic Sign Benchmarks (GTSRB) Dataset is used. Fig. 1 shows the 43 different traffic signs that are considered to train the model. It has 51,900 single images distributed among the 43 classes including the training and the test dataset. The count of the number of photos per class is shown in Fig. 2. There is no ambiguity as the images are just focussed on the traffic signs and each of them is unique. The training dataset has different folders for each of the present classes. A CSV file is also present wherein the path of each image and its class and other details such as width, height is mentioned.

Fig. 1. Traffic Signs Taken into consideration

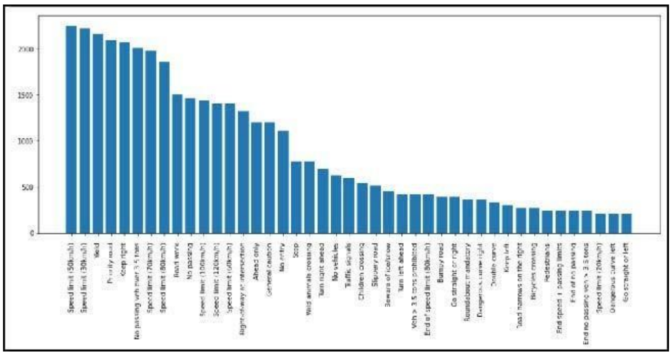

Fig. 2. Number of images per class in the dataset

B. *Data Preprocessing*

To perform image processing, images need to be converted into numpy arrays (i.e. numeric values). After loading the images, they are resized to 30*30 pixels. Post this, the labels of the image are mapped with the image and hence the dataset is ready to be trained.

C. *Model*

Convolutional Neural Network (CNN) is an algorithm falling in the domain of Deep Learning. CNN can take a picture as input, assign priority to different items in the picture, and distinguish them from one another. It requires much less preprocessing as compared to other classification algorithms. Convolutional Network has the ability to learn the

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

filters or characteristics in the images as opposed to the primitive methods filters where they are done manually.

The architecture of a Convolutional Network can be compared to the connectivity pattern of Neurons in the Human Brain. The design itself was inspired by the organization of neurons as present in the Visual Cortex of the human brain. The neurons respond to stimuli only in a certain region of the field of view which is known as the Receptive Field. The visual area is a collection of a number of such receptive fields which help us in viewing objects. Once the model is trained over a series of epochs i.e. iterations, it develops the ability to distinguish between the dominating features and certain low level features in the images. Based on this training, the model classifies them using the Softmax Classification technique.

Fig. 3 represents the number of layers used in the model. There are 4 convolution layers and 2 max pooling layers along with dropout, flatten and dense layers. Adam optimizer is used in the neural network. The input size of the image is 30*30*1. The model employs the ReLU activation function. We obtain a fully connected layer after the Flatten layer. and finally the output is determined by using the softmax activation function.

D. *Proposed Solution*

Fig. 4 demonstrates the accuracy of the trained network. This model turned out to give the best accuracy as compared to the other models that we analysed.

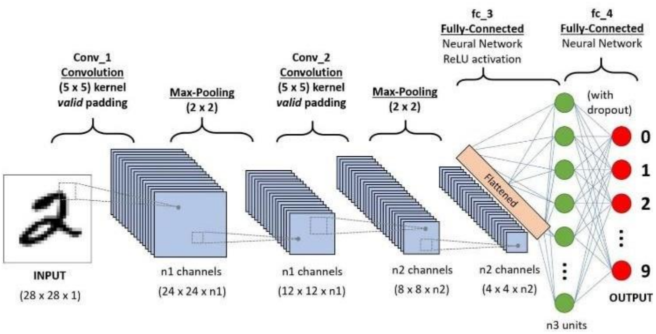

Fig. 3. Neural Network Representation

```
Epoch 18/20
981/981 [==============================] - 135s 137ms/step - loss: 0.1820 - accuracy: 0.9542
 - val_loss: 0.0652 - val_accuracy: 0.9838
Epoch 19/20
981/981 [==============================] - 105s 107ms/step - loss: 0.1819 - accuracy: 0.9558
 - val_loss: 0.1869 - val_accuracy: 0.9484
Epoch 20/20
981/981 [==============================] - 104s 106ms/step - loss: 0.1947 - accuracy: 0.9530
 - val_loss: 0.0573 - val_accuracy: 0.9852
```

Fig. 4. Accuracy of the model on running for 20 epochs

E. *Implementation*

After training the model, it is saved and then the saved model is used for prediction. A full stack web application with NodeJs and Express Handlebars has been developed using this model for prediction. It incorporates different logics to make it a product which can be used with certain improvements in place. Fig. 5 depicts the Flow The suggested system is depicted in this diagram.

The CNN model is applied in the first part wherein the input is an image. After being processed, one of the classes out of the 43 classes is obtained as the output. If a certain image is not containing a traffic sign, then the user gets a prompt of "No Sign Detected". This is done by analysing the output array of the "model.predict" function in python. The "model.predict" function returns an array of values representing how closely the image falls under each of the 43 classes and finally predicts the class based on the highest value.

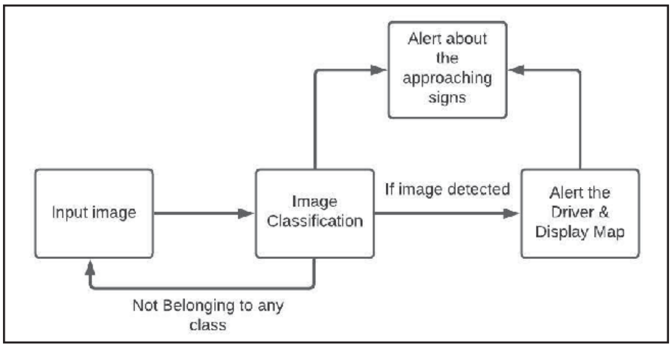

Fig. 5. Flow Diagram

After multiple iterations, it was found that even if an image does not fall in any of the given classes, the model, not being trained for an extra other class, classifies it into one of the 43 classes, but the value predicted by the "model.predict" function is pretty low. So the threshold value for separating the images which actually do not have a traffic sign in them but are predicted as one is taken as 0.68.

The value of the classes in "model.predict" is in the range 0 to 1 and hence if the model classifies it in a particular class with a value less than 0.68 it will be identified as none of the above, else it will be assigned a class. Once the image has been classified, the meta data is fetched from the image using "exif-parser" and then the sign text along with the GPS coordinates are stored in the database.

All this data is then available to the user in a map. The map has markers containing the latitude and longitude along with the name of the traffic sign. Another important feature that needs to be highlighted here is that the aim of the proposed system is to alert the drivers. Therefore, rather than just alerting about the sign which the car is approaching i.e. the sign which has been detected, an algorithm in which the traffic signs that are in proximity i.e. the ones that will be approached within the next 5 minutes (or 1 km) are also to be alerted to the driver, is implemented. The computation of this is done by taking into account the locations of the signs which was stored by extracting the metadata.

Fig. 6 is a sample test case given to the model and Fig. 7 represents the predicted output which will be voiced out to the driver. It also contains a map depicting the location of the various traffic signs in the database.

Fig. 6. No Entry Sign (Input)

With this algorithm the signs which are in the vicinity of the user are found and by using google text to speech (gTTs) a voice alert is sent. With this the user will keep in mind thathe is going to approach certain signs and hence follow the rules. Along with this, there is a section for the user to learn about traffic signs and a section for helpline too. The driver will be verified initially before using our system. This is currently being done by having a database of license numbers and information of the license holder.

III. RESULTS AND ANALYSIS

The trained neural network which contains 4 convolution layers and 2 max pooling layers along with dropout, flatten and dense layers proved to give a better result as compared to the other CNN Architectures used in AlexNet, GoogleNet,

TABLE I. ACCURACY OF VARIOUS MODELS AVAILABLE

Method	Accuracy
AlexNet	92.63%
Trained Neural Network	98.52%

V. CONCLUSION

The Traffic Sign Board Detection and Voice Alert Systemis implemented using Convolutional Neural Network. Various models under the CNN heading were studied and the one withhighest accuracy on the GTSRB dataset was implemented. The creation of different classes for each Traffic sign has helped in increasing the accuracy of the model. A voice message is sent after recognition of the sign which alerts the driver. A map is displayed on which the signs in the vicinity ofthe driver are displayed thus helping him/her take appropriate decisions. This paper is a significant advancement in the field of driving as it would ease the job of the driver withoutcompromising on the safety aspect. Also this system can easilybe implemented without the need of much hardware thusincreasing its reach.

VI. FUTURE SCOPE

The prototype can be expanded to include an inbuilt alert system with a camera in the vehicle's centre. Also, the feature of getting the estimated time for reaching that particular trafficsign can be added. This system can also be expanded for identification of traffic signals and hence prompt the user about the time to reach that particular signal and its status as well. The user can accordingly plan their trip start time and hence cross all signals without having to wait. Also the driver verification will be done with the help of an API providing theinformation about the license holder and the license number

REFERENCES

[1] Yadav, Shubham & Patwa, Anuj & Rane, Saiprasad & Narvekar, Chhaya. (2019). Indian Traffic Sign Board Recognition and Driver Alert System Using Machine Learning. International Journal of Applied Sciences and Smart Technologies. 1. 1-10. 10.24071/ijasst.v1i1.1843.

[2] Anushree.A., S., Kumar, H., Iram, I., & Divyam, K. (2019). Automatic Signboard Detection System by the Vehicles.

[3] S. Harini, V. Abhiram, R. Hegde, B. D. D. Samarth, S. A. Shreyas and K. H. Gowranga, "A smart driver alert system for vehicle traffic using image detection and recognition technique," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2017, pp. 1540-1543, doi: 10.1109/RTEICT.2017.8256856.

[4] C. Wang, "Research and Application of Traffic Sign Detection and Recognition Based on Deep Learning," 2018 International Conference on Robots & Intelligent System (ICRIS), Changsha, China, 2018, pp. 150-152, doi: 10.1109/ICRIS.2018.00047.

[5] M A Muchtar et al 2017 J. Phys.: Conf. Ser. 801 012010

[6] Y. Yuan, Z. Xiong and Q. Wang, "VSSA-NET: Vertical Spatial Sequence Attention Network for Traffic Sign Detection," in IEEE Transactions on Image Processing, vol. 28, no. 7, pp. 3423-3434, July 2019, doi: 10.1109/TIP.2019.2896952.

[7] S. Huang, H. Lin and C. Chang, "An in-car camera system for traffic sign detection and recognition," 2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS), Otsu, Japan, 2017, pp. 1-6, doi: 10.1109/IFSA-SCIS.2017.8023239.

[8] Bi, Z., Yu, L., Gao, H. et al. Improved VGG model-based efficient traffic sign recognition for safe driving in 5G scenarios. Int. J. Mach. Learn. & Cyber. (2020).

[9] Chuanwei Zhang et al., Study on Traffic Sign Recognition by Optimized Lenet-5 Algorithm, International Journal of PatternRecognition and Artificial Intelligence, doi:0.1142/S0218001420550034

[10] Han, C., Gao, G. & Zhang, Y. Real-time small traffic sign detection with revised faster-RCNN. Multimed Tools Appl 78, 13263–13278 (2019). https://doi.org/10.1007/s11042-018-6428-0

[11] H. S. Lee and K. Kim, "Simultaneous Traffic Sign Detection and Boundary Estimation Using Convolutional Neural Network," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 5, pp. 1652-1663, May 2018, doi: 10.1109/TITS.2018.2801560.

[12] R. Qian, Y. Yue, F. Coenen and B. Zhang, "Traffic sign recognition with convolutional neural network based on max pooling positions," 2016 12th International Conference on Natural Computation, Fuzzy