isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 6, Issue: 3,Special Issue: 2 ,Apr,2023 ,ISSN_NO: 2321-3337

Voice-Assisted Traffic Sign Board Detection Using Convolutional Neural Network

Sowmiya.G[1], Abirami.V[2], Sneha.B.A[3], Suresh S[4]

UG Scholar[1, 2, 3]–Department of Computer Science and Engineering,
GRT Institute of Engineering and Technology, Tiruttani, India.
Assistant Professor[4]- Department of Computer Science and Engineering,
GRT Institute of Engineering and Technology, Tiruttani, India.

gm2206683@gmail.com , abiramiselvi1@gmail.com, epsisneha486@gmail.com, surevitcahcet@gmail.com

Abstract - Traffic sign detection has great achievable for smart vehicles. In current years, traffic sign detection has made huge progress with the growth of deep learning. A primary reason of road accidents is negligence in viewing the Traffic signboards and interpreting them incorrectly. The proposed system is trained using a Convolutional Neural Network (CNN), which helps in the detection and classification of traffic signs. The proposed system helps in recognizing the Traffic sign and produce a voice assist to the driver so that he/ she might also take necessary decisions. To enhance the accuracy, a set of classes is described and trained on a particular dataset. The German Traffic Sign Benchmarks Dataset (GTSRB), which includes approximately 43 classes and 50,000 traffic sign images, was used. The accuracy of the execution is about 0.969. The proposed system additionally includes a section where the vehicle driver is alerted about the traffic signs in the close to proximity which helps them to be aware of what policies to follow on the route. The aim of this system is to make sure the protection of the vehicle's driver, passengers, and pedestrians.

I. INTRODUCTION

There have been many advances in technology and cars with autonomous driving modes are already available. Self-driving cars have become a reality. The self-driving car industry is booming. However, these features are only available in high-end cars that are not suitable for the masses. We wanted to create a system that eases the burden of driving.

We found out that the scale of road accidents in India is alarming after conducting a survey. According to reports, about 53 accidents happen on the road every hour. Furthermore, more than 16 people die every hour due to these accidents [18]. When a person disobeys traffic signs while driving, they endanger their own life as well as the lives of other drivers, passengers and road users. Accordingly, we have developed this system in which traffic signs are detected using a images and read aloud to the driver, who can then give necessary decisions. Also, all the traffic signs will be stored in a database 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 summarizes 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.

II. LITERATURE SURVEY

In this era of hectic life, people often tend not to recognize traffic signs and thus violate the rules. A lot of research has been done in this area to reduce the number of accidents. The researchers used a variety of classification algorithms and several CNN architectures to classify traffic signs and warn drivers. Our system aims to optimize the identification process while providing other benefits such as early warning for drivers.

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 categorized 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 pre trained algorithm.

Caffe, an open source system that helps to detect and recognize 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 categories.

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 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 smart phone 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 help in working with large data and a high number of parameters. However it comes with an issue that the large data causes network over fitting decreasing the accuracy to just 80.5%. VGG CNN is proposed in [8] which have a significantly superior performance as compared to other available architectures. The number of parameters in this approach is considerably reduced in order to optimize and speed up the calculation. The network also includes the BN (batch normalization) 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] proposed 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 an overview of all the above-mentioned papers, because it proposes a mechanism for selecting a mini-set of proposals together with a deep hierarchical architecture that allows a neural network to recognize road signs by learning them in separate sets. This method fixes an issue where instances of one record are not marked in another. The system helps to add a new dimension to our project by presenting the idea of traffic sign recognition with voice alert.

III. METHODOLOGY

A. Dataset

In the proposed system, the German Traffic Sign Benchmarks (GTSRB) Dataset is used. Fig 1 shows the 43 different traffic signs considered for training the model. It contains 51,900 frames spread over 43 classes, including training and test data. The number of images per class is shown in the Fig 2. There is no ambiguity because the images are focused only on road signs and each one is unique. The training data has different folders for each available class. There is also a CSV file that mentions the path of each image and its class and other details like width, height.

Fig.1. Traffic Signs Taken into consideration

Volume: 6, Issue: 3,Special Issue: 2 ,Apr,2023 ,ISSN_NO: 2321-3337

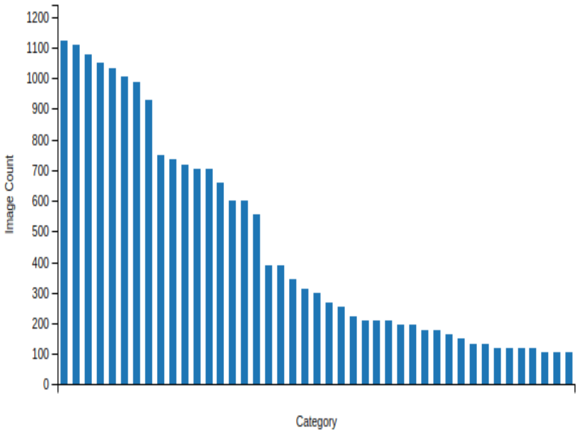


Fig.2. Number of images per class in the dataset

B. Data Processing

The image processing, images must be converted into numpy arrays (i.e., numeric values). After uploading the images, the image captions are associated with the image and thus the dataset is ready for training.

C. Model

A Convolutional Neural Network (CNN) is an algorithm that falls under the scope of deep learning. A CNN can take an image as input, prioritize different elements in the image, and distinguish them from each other. It requires less processing than other classification algorithms. A convolutional network has the ability to learn filters or features in images, compared to the original method filters, which are performed manually.

The architecture of a convolutional network can be compared to the pattern of connections of neurons in the human brain. The design itself is inspired by the arrangement of neurons found in the visual cortex of the human brain. Neurons respond to stimuli only in a specific area of the visual field known as the receptive field. The visual field is a collection of several such receptive fields that help us perceive objects. After the model has gone through a series of epochs i.e. iterations, the ability to distinguish dominant features and some low-level features in images. Based on this training, the model classifies them using the Softmax classification technique.

Fig. 3 shows the number of layers used in the model. There are 4 Convolutional layers and 2 maximum lumped layers along with smoothing, smoothing and density layers. Adam optimizer is used in neural network. The input image size is 30*30*1. The model uses the ReLU activation function. We connect the full layer after the Flatten layer. and finally the result is determined using the softmax activation function.

D. Proposed Solution

Fig. 4 shows the correctness of the training network. This model provided the best accuracy compared to the other models we analyzed.

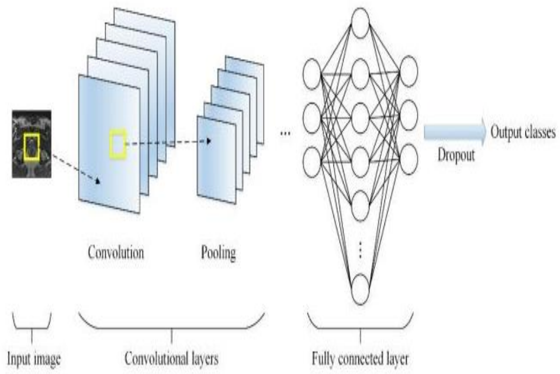
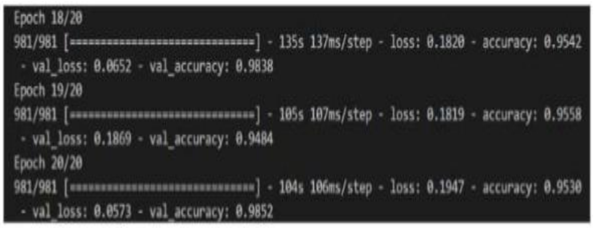


Fig.3. Neural Network Representation



```
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 the model is trained, it is saved and then the saved model is used for prediction. Developed a complete web application with NodeJs and Express Handlebars using this model for prediction. It includes various logics to make it a usable product with certain improvements. Fig. 5 shows the process. The proposed system is shown in this diagram.

The CNN model is used in the first part where the input is an image. After processing, one of the 43 classes is taken as the result. If a particular image does not contain a motion token, the user will receive a "No token detected" message. This is done by parsing the output table of the "model.predict" function in Python. The "model.Predict" function returns an array of values indicating how accurately the image fits each of the 43 classes, and finally predicts the class based on the highest value.

isrjournals
International Journal of Advanced Research in Computer Science Engineering and Information Technology
Volume: 6, Issue: 3,Special Issue: 2 ,Apr,2023 ,ISSN_NO: 2321-3337

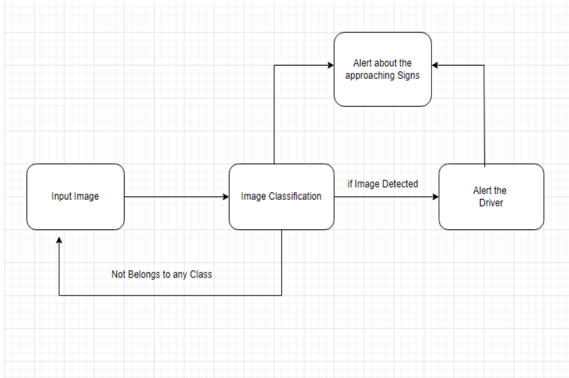


Fig.5. Flow Diagram

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 are stored in the database.

Important feature that needs to be highlighted here is that the aim of the proposed system is to alert the drivers. So, instead of just warning the sign of the approaching car, i.e. the recognized sign uses the algorithm of nearby traffic signs.

Fig. 6 is a sample test case given to the model and Fig. 7 and Fig. 8 represents the predicted output which will be voiced out to the driver.

Fig.6. Yield (Input)

Fig.7. Sign Prediction

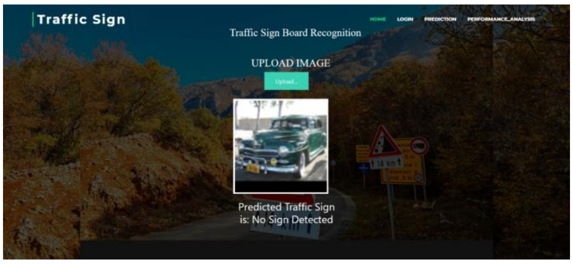


Fig.8. Sign Prediction

IV. RESULTS AND ANALYSIS

A trained neural network with 4 convolutional layers and 2 maximum pooling layers along with deletion, smoothing and density layers gave better results compared to other CNN architectures used in AlexNet, GoogleNet, VSSANet. As mentioned in Table 1, the accuracy of the trained network is 98.52%. Table 2, shows the performance analysis.

TABLE I. ACCURACY OF VARIOUS MODELS AVAILABLE

Method	Accuracy
AlexNet	92.63%
GoogleNet	80.5%
VSSANet	94.42%
Our Trained Neural Network	97.00%

TABLE II. PERFORMANCE ANALYSIS

Accuracy	0.969
Precision	0.952
Recall	0.948
F-Measure	0.959

V. CONCLUSION

Voice-Assisted Traffic Sign Board Detection is implemented using Convolutional Neural Network. Different models were examined under the CNN heading and the model with the highest accuracy was applied to the GTSRB dataset. For each traffic sign, the creation of different classes helped to improve the accuracy of the model. When a sign is recognized, a voice message is alert the driver. This document is an important advancement in the field of driving, as it makes the work of the driver easier without compromising the safety aspect. Again, this system can be easily implemented without requiring a lot of hardware, increasing its scope

VI. FUTURE ENHANCEMENT

The prototype can be expanded to include a built-in alert system with a camera in the center of the car. Also, a function can be added to get the estimated time to reach this road sign. This system can also be extended to detect traffic signals and thus inform the user about the arrival time of this signal and its status. The user can plan the start time of his journey according to the plan and thus cross all the signals without waiting. Driver verification is also done through an API that provides information about the license holder and license number.

VII. 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 etal2017J. 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 Pattern Recognition and ArtificialIntelligence, 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 Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 2016, pp. 578-582, doi: 10.1109/FSKD.2016.7603237.

[13] A. Pon, O. Adrienko, A. Harakeh and S. L. Waslander, "A Hierarchical Deep Architecture and Mini-batch Selection Method for Joint Traffic Sign and Light Detection," 2018 15th Conference on Computer and Robot Vision (CRV), Toronto, ON, Canada, 2018, pp. 102-109, doi: 10.1109/CRV.2018.00024.

[14] A. Welzel, A. Auerswald and G. Wanielik, "Accurate camera-based traffic sign localization," 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 2014, pp. 445-450, doi: 10.1109/ITSC.2014.6957730.