TRAFFIC SIGN BOARD RECOGNITION AND VOICE ASSISTED SYSTEM: USING CONVOLUTIONAL NEURAL NETWORK

Bindu Rani, Prateek Singh

Department of Computer Science and Engineering (AIML) MIT Moradabad

binducs026@gmail.com[1], prateeksingh0809@gmail.com[2]

ABSTRACT

Traffic sign recognition is necessary for intelligent transportation systems to offer road safety and traffic control. In this paper, we propose a Convolutional Neural Networks (CNN) based traffic sign board recognition system with a voice assisted interface. The CNN architecture for image classification in the system is trained and evaluated using the German Traffic Sign Benchmarks Dataset. Additionally, a voice-assisted system is put into place to provide vocal input in accordance with the recognized traffic sign. The proposed method exhibits promising results in terms of accurately recognizing and comprehending traffic signs, which might enhance road safety and driver assistance.

KEYWORDS: *Traffic sign recognition, Convolutional Neural Networks, Voice assistance, Intelligent transportation systems*

1. Introduction

The safety of both drivers and pedestrians on the road is of utmost importance. In order to control traffic flow, maintain safety, and alert drivers to changing road conditions, traffic signs are essential. In 2002, traffic collisions claimed the lives of 1.2 million people worldwide, accounting for 2.1% of all fatalities and ranking as the 11th leading cause of death [4]. Users can learn about the condition of the road, cautions, restrictions, guidelines, etc. via traffic signs Important details about the message the traffic sign is trying to convey are conveyed by its shape.

The manually performed feature extraction and hand-crafted algorithms used by traditional traffic sign recognition systems may not be strong enough to withstand changes in lighting, weather, and other environmental conditions. A CNN-SVM hybrid system, with CNN being utilized for feature extraction and SVM for classification, is presented by Sunitha A in [5]. In [6], Ying Sun presented a method that creates an area of interest and identifies traffic sign areas by applying Hough Transformation to input photos. CNN is utilized for picture identification and classification. Rebai Karima has shown a system in [7] that takes picture data and organizes it according to certain criteria. To conduct recognition, the data representation of traffic signs is extracted using Lenet-5. In [8], Mohit Singh suggested a system that classifies traffic signs using CNN and color-based segmentation, alerting drivers with a beep sound when the sign is identified.

Every day, there are a huge number of more and more traffic accidents worldwide. This might be the result of not understanding traffic indicators. The proposed system's primary objective is to alert drivers of impending traffic signs and provide audio assistance to them.

2. Literature Review

Several studies have explored the use of CNNs for traffic sign board recognition, leveraging datasets such as the German Traffic Sign Benchmarks Dataset. These studies have shown that CNNs can achieve high accuracy in classifying traffic signs across different categories and environmental conditions. For example, the effectiveness of CNNs in identifying traffic signs under actual driving circumstances was shown by Sermanet et al.'s study [1]. Similar to this, Ranneberger et al. [2] suggested the CNN architecture U-net for biomedical picture segmentation; this design might be used to do tasks related to traffic sign identification.

Additionally, the integration of voice-assisted systems in intelligent transportation systems has gained momentum, as it can provide real-time auditory feedback to drivers, enhancing their situational awareness and decision-making capabilities [3].

Due to their busy lives, people in this day and age frequently fail to see traffic signs and break the law. Numerous studies have been conducted in this field to lessen the quantity of mishaps.

Among them, the detection and identification of traffic signs has emerged as a significant area of study. Three widely used techniques for detecting traffic signs are being used: machine learning-based [11], color-based [12], and shape-based techniques. One of the most often used color-based detection techniques is the HSV (Hue, Saturation, and Value) transformation [13].

To categorize traffic signs and alert drivers, the researchers employed several CNN architectures and a range of classification techniques. Our method seeks to improve identification while offering additional advantages like early driver warning.

Numerous research has used a range of strategies for traffic sign detection.[9] A procedure that uses the Support Vector Machine method. The dataset was bifurcated into 80/20 segments for training and testing, then linear classification was applied. To get the desired result, a series of actions referred to as Colour Segmentation, Shape Classification, and Recognition were performed.

Taking pictures of traffic signs seriously is another method of appreciation [10]. We obtain a video and extract its outlines. Preprocessing of images is completed. This involves separating the foreground from the foundation and reducing and differentiating the improvements. Following these exercises, the signs are grouped according to their shape—circular, hexagonal, or triangular—and submitted to be matched with templates. The items with a few positive forms are coordinated by the pre-trained approach.

Detection and Recognition for traffic signs is done by Raspberry Pi with a lot less code [19]. However, the implementation requires the Raspberry Pi board, which is very costly, to be present.

Caffe is an open-source framework that aids in the highly accurate and efficient detection and recognition of traffic signs on roads [14]. A CNN strategy is suggested for generating a model capable of categorizing traffic sign training sets.

Another way to use the CNN system is for traffic signs, as suggested in [15], where the objective sign's real boundary is determined by projecting the input picture plane's border of a matching template sign image. The method becomes end-to-end trainable when we translate the boundary estimation problem into a CNN-based pose and shape prediction job. Deep learning algorithms can be broadly classified into two types of methods: one-stage detectors and two-stage detectors [16]. The region of interests (ROI) is the basis for the two-stage detectors. Compared to other boundary estimating methods that concentrate on contour estimates or picture segmentation, it is more resilient to obstruction and has a greater number of benchmarks.

In the last step of the network, it predicts the classes of objects using a Support Vector Machine (SVM) classifier. In order to maximize efficiency, it employs linear regression to adjust the bounding box sizes and locations. Other CNN designs that have developed into two-stage detectors include the spatial pyramid pooling network (SPP NET) [17], Faster R-CNN [18], and 13 2021 International Conference on Advanced Research in Computing (ICARC 2021). The two-stage detectors were able to achieve a comparatively low frame rate.

3. Methodology

A voice-assisted interface and a traffic sign recognition model based on CNN make up the two primary parts of the suggested system. Preprocessing data, training models, and interaction with the voice-assisted system are all part of the technique.

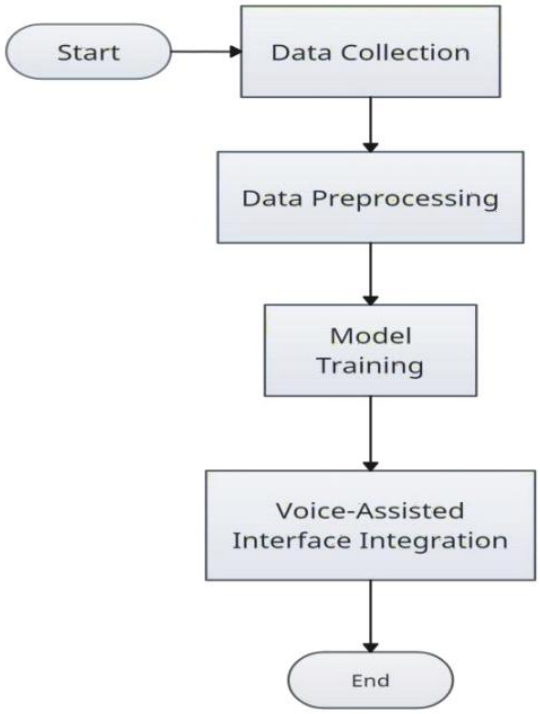

Figure1: Flowchart for Data Methodology

• Data preprocessing:

For the training of CNN model, the Traffic Sign Dataset is used. The collection consists of pictures of many types of traffic signals. The pictures go through a number of preprocessing stages to improve their quality and get them ready for input into the CNN before the model is trained.
In order to simplify calculation and eliminate color information that might not be needed for traffic sign identification, the photos are first converted to grayscale. Histogram equalization is then used to boost the contrast and make the signs more visible. This stage aids in uniformizing the lighting across several photos.
The photos are normalized once the histograms have been equalized to make sure that the pixel values are between 0 and 1. Normalization is an important step since it speeds up the model's convergence and stabilizes the training process. The resulting pre-processed pictures have dimensions of 32x32x1, where 1 denotes the single grayscale channel and 32x32 indicates the image size.

• Model Training:

The CNN model architecture manages tasks related to traffic sign identification. The model consists of many completely linked layers, max-pooling layers, and convolutional layers. While convolutional layers gather characteristics from the input images, max-pooling layers down sample the feature maps to reduce computational complexity and boost the model's ability to generalize. The completely linked layers at the end of the model classify the data based on the obtained attributes. The quantity of neurons in the output layer is equal to the quantity of traffic sign classifications, and a SoftMax activation function provides the probabilities for each class. The model is trained using the Adam optimizer with categorical cross entropy loss for multi-class classification problems.
To keep track of its development and avoid overfitting, the model's performance is assessed using validation datasets during training. When the model performs well enough on the validation datasets, it is tested on test datasets to see how well it can generalize.

• Integration of Voice-Assisted Interfaces:

Real-time feedback to the user is provided by combining the voice-assisted interface with the CNN model. Using text to-speech technology, the interface translates the detected traffic sign into audible speech that the driver or pedestrian may hear. Accessibility and user engagement are enhanced by this voice-assisted technology, particularly for blind or visually impaired people. Accessibility and user engagement are enhanced by this voice-assisted technology, particularly for blind or visually impaired people.

4. Implementation

Python is a well-liked programming language for machine learning and deep learning applications, and it is used to create the suggested system. Various libraries and frameworks like as Keras, Tensor Flow, Open CV, and pyttsx3 are used for different parts of the implementation.

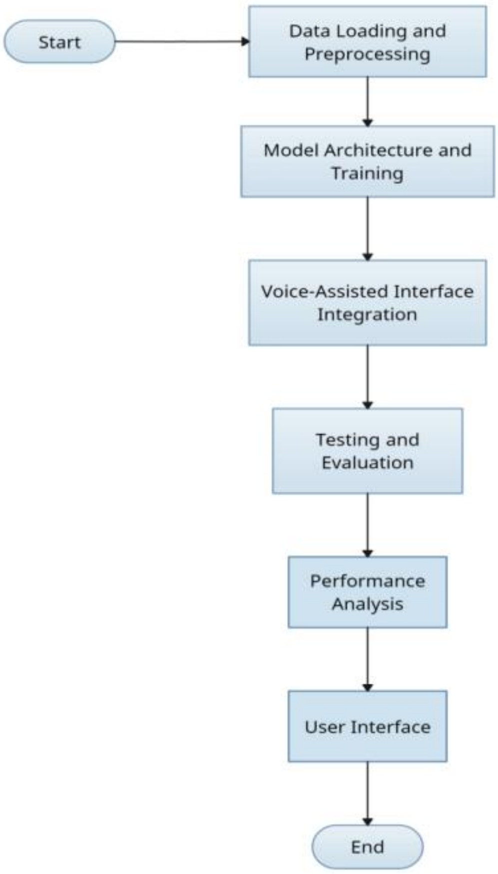

Figure2: Flowchart of Implementation

• Data Loading and Preprocessing:

The Traffic Sign Dataset, which includes pictures of traffic signs in various categories, is loaded as the initial stage of the implementation. Each folder in the collection represents a distinct type of traffic signs. Using the Open CV library, which offers effective image processing capabilities, the pictures are loaded into memory.

The pictures go through a number of preprocessing stages once they are imported in order to get them ready for input into the CNN model. First, in order to decrease computational complexity, the photos are shrunk to a standard size of 32 by 32 pixels. This is a typical approach in deep learning projects. To exclude color information that might not be necessary for traffic sign identification, the photos are then transformed to grayscale. Additionally, grayscale conversion lowers the number of channels in the photos, simplifying processing.

Histogram equalization is used to boost the contrast and visibility of the photos after they have been converted to grayscale. In image processing, histogram equalization is a frequently employed approach to equalize illumination across various photographs. Lastly, the photos' pixel values are normalized to make sure they lie between 0 and 1. In order to stabilize the training process and speed up the model's convergence, normalization is crucial.

Figure3: Data preprocessing

• Model Architecture and Training:

Tasks involving the identification of traffic signs are handled by the CNN model architecture. Multiple convolutional layers, max-pooling layers, and fully linked layers make up the model. In order to lower computational complexity and increase the generalization capacity of the model, max-pooling layers down sample the feature maps, whilst convolutional layers collect features from the input pictures.

Based on the retrieved characteristics, the fully linked layers at the conclusion of the model classify data. The output layer has as many neurons as there are categories of traffic signs, and each class's probabilities are provided by a SoftMax activation function. For multi-class classification problems, the model is trained with the Adam optimizer with categorical cross entropy loss.

To keep track of its development and avoid overfitting, the model's performance is assessed using validation datasets during training. When the model performs well enough on the validation datasets, it is tested on test datasets to see how well it can generalize.

• Voice-Assisted Interface Integration:

To give the user feedback in real time, the voice-assisted interface is integrated with the CNN model. The pyttsx3 package, which offers a straightforward interface for text-to-speech conversion, is used by the interface. The CNN model's recognition of traffic signs informs the interface's design, which is intended to deliver real-time feedback.

• Testing and Evaluation:

To assess the implementation's performance in real-world circumstances, it is put through a thorough testing process on a variety of datasets. To evaluate the system's accuracy and resilience, it is tested in a variety of lighting, weather,
and traffic sign scenarios. The outcomes show how well the suggested approach works at correctly identifying traffic signs and giving users immediate feedback.

• Performance Analysis:

Accuracy, speed, and resilience of the implemented system's performance are evaluated. Measures like accuracy are common assessment measures used to assess how accurate the CNN model is. The speed of the model is analyze by measuring how long it takes it to recognize and interpret traffic sign board in real time. Testing the system under a variety of difficult circumstances, such as dim light, occlusion, and bad weather, allows for the assessment of its resilience.

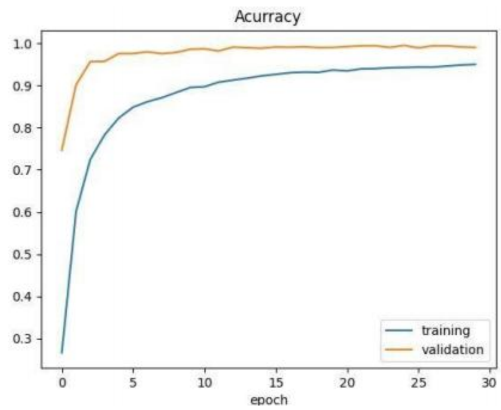

Figure4: Acurracy

• User Interface:

The system's user-friendly interface makes it simple for people to engage with it. By giving users immediate feedback, the voice-assisted interface improves their situational awareness and decision-making skills. Additionally, usability and user satisfaction are taken into account while evaluating the system's performance.

5. Results & Discussion

When it comes to correctly recognizing traffic signs in various categories and environmental situations, the trained CNN model shows encouraging results. Both the training and validation datasets show that the model performs well in terms of accuracy, suggesting that it can effectively generalize to new data. The efficacy of the suggested technique in identifying traffic signals is further validated by evaluation on the test dataset. Furthermore, real-time feedback is effectively provided via the voice-assisted interface, which enhances user awareness and safety.

A voice-assisted interface combined with CNN-based traffic sign recognition has several benefits, such as real-time feedback, intuitive interaction, and accessibility for people with visual impairments. Subsequent research paths might investigate additional improvements to the system, such multi-modal interaction, adaptive learning methods, and real time implementation on embedded systems. In order to determine the system's usefulness and effect on road safety, its performance can also be assessed in actual driving circumstances.

Figure5: Home Page

Figure6: Register Page

Figure7: Login Page

Figure8: Output

6. Conclusion

To put it all up, the suggested method for voice assistance and traffic sign recognition shows promise for improving road safety and user support. The technology contributes to safer road conditions and smarter transportation systems by offering accurate and accessible traffic sign recognition capabilities through the use of voice-assisted interfaces and CNN-based picture categorization.

Several CNN-related models were examined, and the one with Maximum precision on the GTSRB dataset was applied. [20] The precision of the model has increased thanks to the establishment of distinct classes for every traffic sign board.

When driver recognizes the sign board, an audio message is transmitted to notify them. The signs surrounding the vehicle are shown on a map, assisting the driver in making the right choices.

REFERENCES

1. Sermanet, P., LeCun, Y., Bottou, L., Bengio, Y., & Fergus, R. (2011). Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229

2. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham

3. LeCun, Y., Bottou, L., Bengio, Y., &Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324

4. V. Ciuntu and H. Ferdowsi, "Real-Time Traffic Sign Detection and Classification Using Machine Learning and Optical Character Recognition," 2020 IEEE International Conference on Electro Information Technology (EIT), 2020, pp. 480- 486, doi: 10.1109/EIT48999.2020.9208309

5. Sunitha. A, Shanthi. S, "Traffic Sign Classification and Detection Using Deep Learning" International Journal Of Innovative Research In Technology, Volume 6 , Issue 11, ISSN: 2349- 6002, April 2020.

6. Ying Sun, PingshuGe, Dequan Liu, "Traffic Sign Detection and Recognition Based on Convolutional Neual Networks" IEEE, CAC, 978-1-7281-4094-0, Nov 2019.

7. DjebbaraYasmina, RebaiKarima, AzouaouiOuahiba, "Traffic Signs Recognition with Deep Learning" IEEE, ICASS, 978-1-5386-6866-5, November 2018.

8. Mohit Singh, Manish Kumar Pandey, Lakshya Malik, "Traffic Sign Detection and Recognition for Autonomous Vehicles" International Journal of Advance Research, Ideas and Innovations in Technology, Volume 4, Issue 2, ISSN: 2454-132X, 2018.

9. 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.
10. 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.
11. W. Ritter, F. Stein, and R. Janssen, "Traffic sign recognition using colour information," Math. Comput. Model., vol. 22, no. 4–7, pp. 149–161, Aug. 1995, doi: 10.1016/0895-7177(95)00131-K.
12. S. B. Wali, M. A. Hannan, A. Hussain, and S. A. Samad, "An Automatic Traffic Sign Detection and Recognition System Based on Colour Segmentation, Shape Matching, and SVM," Math. Probl. Eng., vol. 2015, pp. 1–11, 2015, doi: 10.1155/2015/250461.
13. H. Fleyeh, "Color detection and segmentation for road and traffic signs," in IEEE Conference on Cybernetics and Intelligent Systems, 2004., Singapore, 2004, vol. 2, pp. 809–814, doi: 10.1109/ICCIS.2004.1460692. 14. 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.
15. 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.
16. P. Soviany and R. T. Ionescu, "Optimizing the Trade-off between Single-Stage and Two-Stage Object Detectors using Image Difficulty Prediction," ArXiv180308707 Cs, Aug. 2018
17. K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," ArXiv14064729 Cs, vol. 8691, pp. 346–361, 2014, doi: 10.1007/978-3-319-10578-9_23. 18. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," ArXiv150601497 Cs, Jan. 2016
19. Anushree.A., S., Kumar, H., Iram, I., & Divyam, K. (2019). Automatic Signboard Detection System by the Vehicles. 20. Krish Sukhani, Dr Radha Shankarmani, Jay Shah 2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456302

AUTHORS

Bindu Rani has received M. Tech degree in Computer Science and Engineering. She is currently working as an Assistant Professor in Computer Science and Engineering (AIML) department at MIT with 8 years of professional experience. She has wide expertise in Artificial Intelligence & Machine Learning.

Prateek Singh is a B.Tech 4th year student in computer science and engineering department at Moradabad Institute of Technology affiliated with Dr. A.P.J Abdul Kalam Technical University his research interest including Machine Learning, Artificial Intelligence and Natural Processing Language, Python and SQL.