

Traffic Sign Recognition using Machine Learning: A Review

Vaibhavi Golgire

Department of Computer Engineering,
Pimpri Chinchwad College of Engineering,
Savitribai phule pune university, Pune, Maharashtra, India

Abstract:- A series of warnings about the route are conveyed by traffic signs. They keep traffic going by aiding travelers in reaching their destinations and providing them with advance notice of arrival, exit, and turn points. Road signs are placed in specific positions to ensure the safety of travelers. They also have guidance for when and where drivers can turn or not turn. In this paper, we proposed a system for traffic sign detection and recognition, as well as a method for extracting a road sign from a natural complex image, processing it, and alerting the driver through voice command. It is applied in such a way that it helps drivers make fast decisions. In real-time situations, factors like shifting weather conditions, changing light directions, and varying light intensity make traffic sign identification challenging. The reliability of the machine is influenced by a number of factors such as noise, partial or absolute underexposure, partial or complete overexposure, and significant variations in color saturation, wide variety of viewing angles, view depth, and shape/color deformations of traffic signs (due to light intensity). The proposed architecture is sectioned into three phases. The first of which is image pre-processing, in which we quantify the dataset's input files, determine the input size for learning purposes, and resize the information for the learning step. The proposed algorithm categorizes the observed symbol during the recognition process. A Convolutional Neural Network is used to do this in the second phase, and the third phase deals with text-to-speech translation, with the detected sign from the second phase being presented in audio format.

Keywords- *Convolution Neural Network, Machine Learning, Image Preprocessing, Feature Extraction, Segmentation, Data Augmentation, Text to speech conversion.*

INTRODUCTION

According to official statistics, about 400 road accidents occur in India every day. Road signs help to avoid accidents on the road, ensuring the safety of both drivers and pedestrians. Additionally, traffic signals guarantee that road users adhere to specific laws, minimizing the likelihood of traffic violations. Route navigation is also made easier by the use of traffic signals. Road signals should be prioritized by all road users, whether they are drivers or pedestrians. We overlook traffic signs for a variety of reasons such as problems with concentration, exhaustion, and sleep

deprivation. Other causes that contribute to missing the signs include poor vision, the influence of the external world, and environmental circumstances. It is much more important to use a system that can recognize traffic signals and advise and warn the driver. Image-based traffic-sign recognition technologies analyze images captured by a car's front-facing camera in real time to recognize signals. They help the driver by giving him or her warnings. The identification and recognition modules are the key components of a vision-based traffic sign recognition system. The detection module locates the sign area in the image/video, while the recognition module recognizes the sign. The sign regions with the highest probability are selected and fed into the recognition system to classify the sign during the detection process. For traffic sign recognition, various machine learning algorithms such as SVM, KNN, and Random Forest can be used [6]. However, the key disadvantage of these algorithms is that feature extraction must be done separately; on the other hand, CNN will do feature extraction on its own [1]. As a result, the proposed system employs a convolutional neural network. Input preprocessing module will prepare image captured with the help of vehicle camera for recognition stage before that. The driver will get a voice warning message after recognition.

RELATED WORK

In any kind of study, the most critical move is to do a literature review. This move would allow us to identify any gaps or flaws in the current structure which will attempt to find a way to get around the limitations of the current method. We briefly discuss similar work on traffic sign detection identification and recognition in this segment. Comparative analysis of reference articles is shown below in Table 1.

1. Wasif Arman Haquea ,SaminArefin b , A.S.M.

they considered small input images, a small

Paper	Technology/Algorithm	Advantages	Limitation
DeepThin: A novel lightweight CNN architecture for traffic sign recognition without GPU requirements Author- Wasif Arman Haquea ,SaminArefin b , A.S.M. Shihavuddin c , Muhammad Abul Hasan Year- 2021	Authors proposed DeepThin architecture which is divided into 3 modules input processing, learning, and prediction. Image resizing is done in preprocessing. four convolutional layers, two overlapping max-pooling layers followed by a single fully connected hidden layer is used for learning ,class prediction is done with the help of CNN	Because of light weight architecture it can be used on a low-end personal computer even without GPUs. Such network optimization lowers the energy usage criteria for deep learning testing, allowing for environmentally sustainable characteristics in the solution	Only the color characteristic of the sign is considered during the detection process. They concentrated on the RGB and grayscale values of signs.
An efficient convolutional neural network for small traffic sign detection Author- Shijin Songa ,Zhiquang Que b, JunjieHoua , Sen Dua , YuefengSonga Year- 2019	Author focused on issues of small object detection and compared accuracy against R-CNN and Faster R-CNN.CNN Model is optimized using convolution factorization, redundant layer cropping and fully connected transformation	The model has been optimized to use less GPU memory and reduce computing costs	Image preprocessing details are missing
Traffic Sign Detection and Recognition using a CNN Ensemble Author- AashrithVennelakanti, Smriti Shreya, ResmiRajendran, Debasis Sarkar, Deepak Muddegowda, PhanishHanagal Year- 2019	Hue Saturation Value(HSV) color space is used instead of RGB for color based detection and Douglas Peucker algorithm is then used for shape based detection	Two data sets used for evaluation and CNN Ensembles are used to improve accuracy	Good accuracy is achieved but only triangular and circular shapes are considered for detection
Deep Learning for Large Scale Traffic-Sign Detection and Recognition Author- DomenTabernik; DanijelSkočaj Year- 2020	CNN, the mask R-CNN is used for traffic sign detection and recognition. To have low inter-class and high intra-class variability they produced new data set called DFG traffic-sign	Data augmentation has been done and By changing segmented, real-world training samples, more synthetic traffic-sign instances are developed. There were two kinds of distortions used: geometric/shape distortions (perspective shifts, color shifts) and appearance distortions (brightness shifts)	Miss detection of traffic signs scenarios not considered
The Speed Limit Road Signs Recognition Using Hough Transformation and Multi-Class Svm Author- Ivona Matoš; Zdravko Krpić; Krešimir Romić Year- 2019	SVM is used for classification and HOG descriptor for feature extraction	Images with a lot of noise were treated well and up to 95% performance was achieved	Proposed system scope is limited to only circular signs

Shihavuddin c , Muhammad Abul Hasan [1] describe the “A novel lightweight CNN architecture for traffic sign recognition without GPU requirements”. Author focused on Main challenges in detecting traffic signs in real time scenarios includes distortion of images, speed factor, motion effect, noise, faded color of signs. Training only on grayscale images gives average accuracy. So authors proposed DeepThin architecture which is divided into 3 modules input processing, learning, and prediction. Architecture is deep and thin at the same time. Thin because they considered small number of feature maps per layer and deep because 4 layers used. And since

number of feature maps, and large convolution strides, it has become possible to train without a GPU. use of overlapping max pooling and sparsely used stride convolution made training faster and reduced overfitting issue. Data augmentation is performed in order to achieve robustness. For augmentation they used operations such as original random shearing of training images, zoomed-in/zoomed-out, horizontally-shifted, vertically-shifted during training. For experimentation German Traffic Sign Recognition Benchmark and Belgian Traffic Sign Classification dataset is used. hyper parameter tuning is done for kernel size and feature map and

During training phase CNN model is used with backpropagation learning algorithm, cross-entropy, stochastic gradient descent (SGD) as the optimizer.

2. **Shijin Songa ,Zhiqiang Que b, JunjieHoua , Sen Dua , YuefengSonga [2]** describe the “An efficient convolutional neural network for small traffic sign detection”. In this paper, researcher focused on issues for small object detection and proposed efficient convolutional neural network for small traffic sign detection and compared accuracy against R-CNN and Faster R-CNN. CNN model is explained in detail along with forward propagation, back word propagation, loss functions. Authors increased the number of convolutional kernels per Conv layer from the start and implemented Max-pooling layers with a stride of 2 to down-sample the network in the feature extraction phase. To optimize this model further three strategies used convolution factorization, redundant layer cropping and fully connected transformation. The Tsinghua-Tencent data set is used for evaluation. Proposed model is not only efficient but also consumed less GPU memory and save the computation cost.
3. **AashrithVennelakanti, Smriti Shreya, ResmiRajendran, Debasis Sarkar, Deepak Muddegowda, PhanishHanagal [3]** describe the “Traffic Sign Detection and Recognition using a CNN Ensemble”. Proposed system in this paper is divided into two modules detection and recognition and it is evaluated on Belgium Data Set and the German Traffic Sign Benchmark. Detection involves capturing images of traffic sign and locating object from image and in recognition stage convolutional neural network ensemble is used which will assign label to detected sign. In first phase Hue Saturation Value(HSV) color space is used instead of RGB because HSV model is more similar to the way human eye process image and it has wide range of colors. After that color based detection and shape based detection is implemented, in color based detection red values of sign are checked if they fall under particular threshold then that part is examined to see if sign is present or not. Douglas Peucker algorithm is then used for shape based detection. Authors focused on only 2 shapes circle and triangle. This algorithm found area from no of edges detected in image and bounding boxes are used to separate ROI. Now sign inside bounding box is validated by applying image thresholding and inversion filter. In the second phase detected sign is classified using feed-forward CNN network with six convolutional layers and As they used ensemble method, aggregated result of 3 CNN is a final output. They achieved 98.11% accuracy for triangular traffic signs and 99.18% for circles.
4. **DomenTabernik; DanijelSkočaj [4]** describe the “Deep Learning for Large-Scale Traffic-Sign Detection and Recognition”. In this paper convolutional neural network (CNN), the mask R-CNN is used for traffic sign detection and recognition. Authors used CNN for full feature extraction rather than Hough transform, scale invariant feature transform, local binary patterns. In order to solve real time problems of traffic sign appearance and distortion they also implemented data augmentation method. Swedish traffic-sign dataset (STSD) is used for evaluation of Faster R-CNN and Mask R-CNN. To have low inter-class and high intra-class variability they produced new data set called DFG traffic-sign. To improve the overall recall, average precision modification has been done in Mask R-CNN.
5. **Ivona Matoš; Zdravko Krpić; Krešimir Romić [5]** describe the “The Speed Limit Road Signs Recognition Using Hough Transformation and Multi-Class Svm”. In this paper preprocessing step, hue, saturation, and lightness (HSL) values are used to improve the contrast in dataset images, making detection simpler. The Hough Circle feature was used in the detection process. It uses the Hough transformation to locate circles inside pictures. HOG descriptor is used for edge detection and at the end SVM classifier is used to train and test model and proposed model is tested on MASTIF and GTSRB data set.

EXISTING SYSTEM

In the area of traffic sign detection and recognition, a considerable amount of work has been put forward. As two global characteristics of traffic signs, several authors concentrated on the color and shape attributes of image for detection. These features can be used to detect and trace a moving object in a series of frames. This approach is helpful when the target to be identified is a special color that is distinct from the background color. To detect an object with a certain shape, object borders, corners, and contours may be used. However authors only focused on the detection and recognition measures, ignoring the voice feature, which is an essential driver warning system. In addition, hyper parameter tuning has received less attention. As a result, the proposed system would concentrate on different parameters of the CNN algorithm in order to improve accuracy without requiring additional computing resources.

PROPOSED SOLUTION

In the proposed system, Traffic sign detection and recognition is achieved by CNN algorithm. Before classification input preprocessing is done in order to remove noise, reduce the complexity and improve the precision of the implemented algorithm. Since we can't write a special algorithm for each condition under which an image is taken, we tend to transform images into a format

that can be solved by a general algorithm. At the end voice alert message will be given to driver.

Image Preprocessing :

1. **Gray Scale Conversion:** To save space or reduce computing complexity, we can find it helpful to remove redundant details from images in some situations. Converting colorful images to grayscale images, for example. This is because color isn't always used to identify and perceive an image in several objects. Grayscale may be sufficient for identifying such artefacts [1][3]. Color images can add needless complexity and take up more memory space because they hold more detail than black and white images color images are represented in three channels, which means that converting it to grayscale reduces the number of pixels that need to be processed. For traffic signs gray values are sufficient for recognition
2. **Thresholding and Segmentation:** Segmentation is the method of partitioning a visual image into different subgroups (of pixels) called Image Objects, which reduces the image's complexity and makes image analysis easier. Thresholding is the method of using an optimal threshold to transform a grayscale input image to a bi-level image [4].

Traffic sign recognition:

Deep Learning is a subdomain of Machine Learning that includes Convolutional Neural Networks. Deep Learning algorithms store information in the same manner as the human brain does, but on a much smaller scale. Image classification entails extracting features from an image in order to identify trends in a dataset. We are using CNN for traffic sign recognition as it is very good at feature extraction [1][2]. In CNN, we use filters. Filters come in a variety of shapes and sizes, depending on their intended use. Filters allow us to take advantage of a specific image's spatial localization by imposing a local communication pattern between neurons. Convolution is the process of multiplying two variables pointwise to create a new feature. Our image pixels matrix is one function and our filter is another. The dot product of the two matrices is obtained by sliding the filter over the image. Matrix called "Activation Map" or "Feature Map". The output layer is made up of several convolutional layers that extract features from the image. CNN can be optimized with the help of hyper parameter optimization. It finds hyper parameters of a given machine learning algorithm that deliver the best performance as measured on a validation set. Hyper parameters must be set before the learning process can begin [1]. The learning rate and the number of units in a dense layer are provided by it. In our system will consider dropout rate, learning rate, kernel size and optimizer hyper parameter.

Convolutional Neural Network Architecture

1. Convolution Layer

This layer is major building block in convolution process. It performs convolution operation to identify various features from given image[1]. It basically scans entire pixel grid and perform dot product. Filter or kernel is nothing but a feature from multiple features which we want to identify from input image. For example in case of edge detection we may have separate filter for curves, blur, sharpen image etc. As we go deeper in the network, more complex features can be identified

2. Pooling Layer

This layer is used for down sampling of the features. It reduces dimensionality of large image but still retains important features. It helps to reduce amount of computation and weights. One can choose Max pooling or Average pooling depending on requirement. Max pooling takes maximum value from feature map while average takes average of all pixels.

3. Activation Function

This layer introduces non linear properties to network. It helps in making decision about which information should be processed further and which not. Weighted sum of input becomes input signal to activation function to give one output signal

This step is crucial because without activation function output signal would be simple linear function which has limited complex learning capabilities. Types of activation function includes Sigmoid function, Tan H, ReLU, Identity, Binary Step function. Sigmoid function is mostly used in backpropagation its range is 0 to 1 while TanH range is -1 to 0, Optimization is easy in this function. Range for ReLU is 0 to infinity, it's a most popular activation function.

4. Flattening Layer

The output of the pooling layer is in the form of a 3D feature map, and we need to transfer data to the fully connected layer in the form of a 1D feature map. As a result, this layer transforms a 3*3 matrix to a one-dimensional list.

5. Fully connected Layer

Actual classification happens in this layer. It takes end result of convolution or pooling layer by flattened layer and reaches a classification decision. Here every input is connected to every output by weights. It combines the features into more attributes that better predicts the classes

Output of recognized sign in audio format:

At present the driver will have to read the text written on the classified sign, but with the aid of a speech module, more comfort is assured. A text to speech module will alert driver with detected sign. In Python, there are many APIs available for converting text to voice. The Google Text to Speech API, also known as the gTTS API, is one of these APIs. gTTS is a simple application that transforms entered text into audio that can be stored as an mp3 format. The gTTS API supports several languages and audio can be delivered at customized speed

CONCLUSION

We presented a literature review on traffic sign identification using machine learning techniques, as well as a comparative study and analysis of these techniques in this paper. CNN performs well for recognition and with the aid of hyper parameter tuning, accuracy or recognition rate can be improved. As a result, in the proposed scheme to design a warning traffic sign detection system for drivers, we used CNN for traffic sign recognition. The images will be taken with a camera mounted on the car during the image acquisition stage and the recognition process will be done using the CNN algorithm after preprocessing. The machine issues a voice alert when a traffic sign is identified. This model can be used in circumstances requiring precise navigation.

VII. REFERENCES

- [1] W. Haque, S. Arefin, A. Shihavuddin and M. Hasan, "DeepThin: A novel lightweight CNN architecture for traffic sign recognition without GPU requirements", *Expert Systems with Applications*, vol. 168, p. 114481, 2021.
- [2] S. Song, Z. Que, J. Hou, S. Du and Y. Song, "An efficient convolutional neural network for small traffic sign detection", *Journal of Systems Architecture*, vol. 97, pp. 269-277, 2019. Available: 10.1016/j.sysarc.2019.01.012.
- [3] A. Vennelakanti, S. Shreya, R. Rajendran, D. Sarkar, D. Muddegowda and P. Hanagal, "Traffic Sign Detection and Recognition using a CNN Ensemble," 2019 IEEE International Conference on Consumer Electronics (ICCE), 2019, pp. 1-4
- [4] D. Tabernik and D. Skočaj, "Deep Learning for Large-Scale Traffic-Sign Detection and Recognition," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1427-1440, April 2020
- [5] I. Matoš, Z. Krpić and K. Romić, "The Speed Limit Road Signs Recognition Using Hough Transformation and Multi-Class Svm," 2019 International Conference on Systems, Signals and Image Processing (IWSSIP), 2019, pp. 89-94.
- [6] Degui Xiao, Liang Liu, "Super-resolution-based traffic prohibitory sign recognition", 2019.