IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Traffic Sign Recognition System

[1]PV Jayachandra Reddy, [2]N Vishal, [3]Rickl Pamula, [4]Shaik Badil Hassan, [5]Prabhakar K

[1-4]Department of Computer Science, CMR University, Bangalore, India

[5]Professor Department of Computer Science, CMR University, Bangalore, India

Abstract—The Traffic signs are important in communicating information to drivers. Thus, comprehension of traffic signs is essential for road safety and ignorance may result in road accidents. Traffic sign detection has been a research spotlight over the past few decades. Real- time and accurate detections are the preliminaries of robust traffic sign detection system which is yet to be achieved. This study presents a voice-assisted real-time traffic sign recognition system which can assist drivers. This system functions under two subsystems. Initially, the detection and recognition of the traffic signs are carried out using a trained Convolutional Neural Network (CNN). After recognizing the specific traffic sign, it is narrated to the driver as a voice message using a text-to-speech engine. An efficient CNN model for a benchmark dataset is developed for real-time detection and recognition using Deep Learning techniques. The advantage of this system is that even if the driver misses a traffic sign, or does not look at the traffic sign, or is unable to comprehend the sign, the system detects it and narrates it to the driver. A system of this type is also important in the development of autonomous vehicles. Initially, the detection and recognition of the traffic signs are carried out using a trained Convolutional Neural Network (CNN). After recognizing the specific traffic sign, it is narrated to the driver as a voice message using a text-to-speech engine. An efficient CNN model for a benchmark dataset is developed for real-time detection and recognition using Deep Learning techniques. The advantage of this system is that even if the driver misses a traffic sign, or does not look at the traffic sign, or is unable to comprehend the sign, the system detects it and narrates it to the driver. A system of this type is also important in the development of autonomous vehicles.

Index Terms— Convolutional Neural Network (CNN), Dataset, Image Processing

1. INTRODUCTION

Since automobiles have become an indispensable medium of transportation, assurance of safety has been implemented in every country through proper road rules and regulations. Among them, traffic signs provide valuable information to the drivers and help to communicate the rules to be followed in that specific area. The purpose of a traffic sign is to convey a message quickly and accurately with minimum reading skills. Negligence, lack of attention, lack of familiarity, accidentally or deliberately not noticing traffic signs, distracting driving behaviours have been discovered as major reasons for the ignorance of road signs among the drivers which eventually lead to road accidents. Furthermore, drivers in unurbanized communities may find it difficult in decoding the message conveyed by a specific road sign due to a lack of familiarity with the plenty of road signs in urbanized areas. Some drivers tend to ignore certain traffic signs believing that they are not necessary. Different attitudes of the drivers are also a reason for this ignorance. Ignorance or unfamiliarity with traffic signs could result in severe accidents and may even cost lives.

1.1 Motivation and Significance

The automobiles have become an indispensable medium of transportation, assurance of safety has been implemented in every country through proper road rules and regulations. Among them, traffic signs provide valuable information to the drivers and help to communicate the rules to be followed in that specific area. The purpose of a traffic sign is to convey a message quickly and accurately with minimum reading skills. Negligence, lack of attention, lack of familiarity, accidentally or deliberately not noticing traffic signs, distracting driving behaviours have been discovered as major reasons for the ignorance of road signs among the drivers which eventually lead to road accidents.

1.2 Problem statement

The current landscape of traffic sign recognition systems faces notable challenges that hinder their effectiveness in providing timely and user-friendly information to drivers. The absence of a seamless integration between real-time visual processing and user interaction results in inefficiencies and potential safety concerns. Existing systems often lack voice assistance, requiring drivers to divert their attention from the road to interpret visual cues. This project addresses the need for a more efficient and user-friendly solution by developing a Voice-Assisted Real-Time Traffic Sign Recognition System using Convolutional Neural Network (CNN). The primary challenges include inefficient communication channels, potential safety risks associated with visual distractions, limited adaptability to diverse environmental conditions, integration challenges with navigation systems, the need for comprehensive multi-modal interactions, privacy and security concerns related to video processing, and the absence of a continuous improvement mechanism. Through this project, we aim to enhance the adaptability, safety, and overall user experience of traffic sign recognition systems by incorporating voice assistance and addressing the identified challenges.

1.3 Objectives

The primary objectives of this study include:

- Real-Time Traffic Sign Recognition: Develop a system capable of recognizing traffic signs in real-time using computer vision techniques and a Convolutional Neural Network (CNN).
- Voice Assistance Integration: Implement voice assistance to enhance the user interface and user experience.
- Driver Awareness and Safety: Improve driver awareness and safety by delivering timely information about road signs.
- Multi-Modal Interaction: Facilitate a multi-modal interaction system where both visual and auditory cues work together seamlessly.

2. LITERATURE REVIEW

The first research on traffic sign recognition can be traced back to 1987; Akatsuki and Imai attempted to make an early traffic sign recognition system. A system capable of automatic recognition of traffic sign could be used as assistance for drivers, alerting them about the presence of some specific sign (eg, a one-way street) or some risky situation (eg, driving at a higher speed than the maximum speed allowed). It also can be used to provide the autonomous unmanned some specific designed signs. Generally, the procedure of a traffic sign recognition system can be roughly divided of two stages namely detection and classification.

2.1 Traffic Sign Recognition with Multi-Scale Convolutional Networks

This seminal work introduced the use of deep learning for traffic sign recognition, employing multi-scale convolutional networks. The study demonstrated the effectiveness of CNNs in handling scale variations and achieved state-of-the-art results on the German Traffic Sign Recognition Benchmark (GTSRB)

2.2 Deep Traffic: A Framework for Efficient and Real-Time Traffic Sign Recognition

This research focused on developing an efficient and real-time traffic sign recognition system using a CNN-based approach. The study addressed the challenges of real-time processing, proposing a lightweight network architecture suitable for deployment on embedded systems.

2.3 Enhanced Traffic Sign Recognition with AlexNet-based Feature Fusion

This study explored the use of the AlexNet architecture for traffic sign recognition and introduced a feature fusion mechanism to enhance performance. The research highlighted the importance of feature extraction and fusion for improving the accuracy of traffic sign recognition systems.

2.4 Transfer Learning for Traffic Sign Recognition: A Systematic Investigation

Focusing on transfer learning, this research systematically investigated the application of pre-trained CNNs for traffic sign recognition. The study evaluated various pre-trained models and transfer learning strategies, providing insights into the benefits of leveraging knowledge from large-scale datasets.

2.5 Robust Traffic Sign Recognition under Adverse Conditions via CNNs

This study addressed the robustness of traffic sign recognition under adverse conditions, including low lighting and challenging weather. The research proposed a CNN-based approach that incorporated data augmentation techniques and highlighted the importance of training datasets that simulate real-world variations.

3 METHODOLOGY

In developing a Traffic Signs Classification system utilizing Convolutional Neural Networks (CNNs), a systematic methodology is crucial. Initially, the problem is clearly defined, outlining the system's goals and specific challenges, such as real-time processing and environmental robustness. A diverse dataset of annotated traffic sign images is then assembled and divided for training, validation, and testing. Subsequent preprocessing involves standardizing image sizes, normalizing pixel values, and applying data augmentation techniques. The choice of a CNN architecture, possibly leveraging transfer learning, is a pivotal step. Model development follows, incorporating fine-tuning of hyperparameters and addressing class imbalances. Training is conducted on the training dataset, with ongoing monitoring and adjustments based on validation set performance. Evaluation metrics, including accuracy and precision, assess the model's proficiency on the test set. Optimization strategies, such as quantization and pruning. enhance real-time processing capabilities. A user-friendly interface is designed, providing clear outputs and mechanisms for user feedback. The system is developed with adaptability and scalability in mind, accommodating updates and scalability for an increasing number of sign classes. Security measures, including adversarial robustness and anomaly detection, are implemented. Deployment involves optimizing the model for edge devices and integration with existing systems, followed by thorough real-world testing. Documentation of the entire process is emphasized, facilitating continuous improvement and updates based on user feedback and evolving road infrastructure.

Convolutional Neural Networks CNN is the state-of-the-art deep learning technique used in computer vision. Neural Network is a mathematical model which is modelled based on the primitives of neurons. Many artificial neurons are networked into layers to build a deep neural network. It accepts vectors as inputs and passes through the layers of the network and predicts the output. CNN is a type of deep neural network which consists of three types of layers namely convolution, pooling, and fully connected layers. Out of these, the first two types are involved with the extraction of characteristics while the fully connected layers map the extracted features into classification. Several convolutional neural network architectures have been adapted in the process of image detection and image recognition. The CNN that is used in this paper is the YOLO which was proposed by Joseph Redmon et al

YOLO Architecture Several versions of YOLO architectures have been incorporated in the training and testing phase in this paper. Here the focus is set on YOLOv4 which yielded the best results, YOLOv4 was developed by Alexy Bukovsky et al. Instead of selecting the Regions of Interest (ROI) as in two-stage detectors like RCNN, the YOLO algorithm predicts classes and bounding boxes from the whole image in just one run in the network. YOLOv4 outperforms the other members in the YOLO family with an Average Precision of 43.5% on COCO dataset with 65 FPS in a Tesla V100. Also, YOLOv4 addresses the need for multiple GPUs by employing an object detector which can be trained on a single GPU. The top-level architecture of YOLOv4 is shown, which is extracted from the YOLOv4 paper

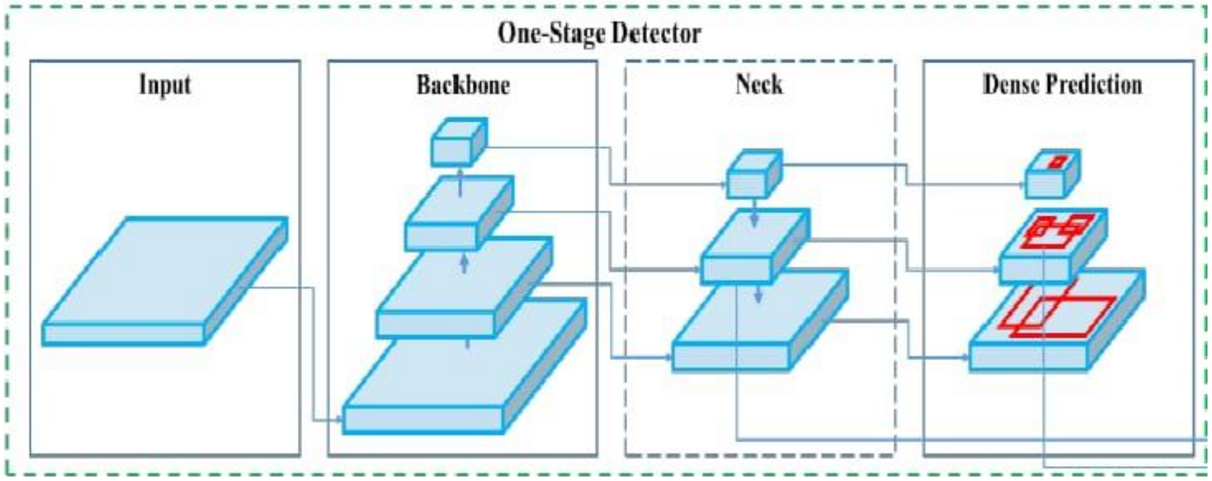

Fig: The high-level architecture of YOLOv4 network.

As shown in Fig. the backbone is used as a feature extractor. Authors have set focus on CSPResNext50, CSPDarknet53 and Efficient Net-b3 as backbones for the YOLOv4 object detector. Based on the experiments on ImageNet and MSCOCO datasets, CSPDarknet53 is taken as the backbone for the YOLOv4 detector. The Neck is the set of extra layers that connects the backbone with the head. Neck layers are used to extract different feature mappings at different levels of the backbone. YOLOv4 uses Spatial Pyramid Pooling (SPP) and a modified version Path Aggregation Network (PAN) for the detector. The head part or the dense prediction is the network which is used to carry out the detection parts. Specifically, it carries out the detection and regression of the bounding boxes. YOLOv4 uses the same head as in YOLOv3 [14]. As in Fig. 1, the detections are carried out at 3 YOLO layers. But, for this research, the tiny version of YOLOv4 architecture is used where the network size is dramatically reduced. YOLOv4 architecture uses only 2 YOLO layers and the convolutional layers in the CSP backbone are compressed. Thus, it makes the detections at faster rates.

The ultimate target of the YOLO detection layer is to predict the class of an object and locate the bounding box related to that image. As shown in Fig. 2 the bounding box has four parameters describing it: centre coordinates represented by (bx, by), width (bw) and height (bh),

As mentioned earlier, there is no selection of ROI like in two stage detectors, Instead, the input image is split into SS squares. Each square in the grid predicts B number of bounding boxes and their confidence values along with the classes C. Confidence values measure whether the square is consisting of objects and if there is any object the accuracy of the bounding box is predicted.

Confidence pr(Object) x loU

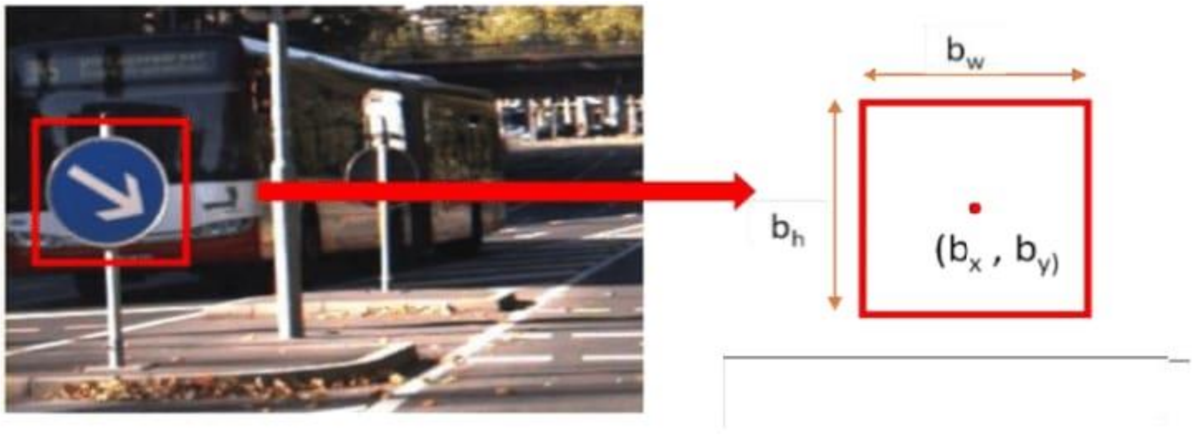

Fig: Format of the bounding box in YOLO network

When a square in the grid contains a part of the ground truth box of an object, the value of pr(Object) becomes I and 0 if there is no ground truth box. IoU (Intersection over Union) indicates the intersection over union values between the predicted bounding box, and the ground truth box. As each bounding box is represented by five values: bx, by, bw, bh, and confidence, the output is a tensor of shape SS (S*) B + C) When multiple frames are predicting the same object, YOLO uses a non-maximum suppression technique to select the most suitable frame.

Audio Narration The system is integrated with an audio feedback system to narrate the detected traffic signs. As the detections are happening real time, the audio outputs should also be provided in real time. For this, the detections and audio outputs are allowed to run parallelly in such a way that when an object is detected, the voice feedback of the sign is provided simultaneously. The gates (Google Text to Speech) library is used for audio narrations. The language can be customized as per user preference. Whenever a new detection is observed in the frame, the detected sign is fed into the algorithm where the voice is played. Fig. illustrates the high-level architecture including the audio feedback system.

Fig: sample images for mapillary traffic sign dataset

Dataset Initially, the system was developed using the German Traffic Sign Detection Benchmark (075DB) dataset for the CNN model. The dataset contains 900 images in which 600 are training images and 300 are validation images, zero to six traffic signs are included per image. The traffic signs in this dataset appear in every perspective and under many lighting conditions. Few example images are given it Fig. 3.4 The traffic sign instances are divided into four categories as danger, prohibitor, mandatory, and others. The dataset contains annotations in CSV format, and it is converted to YOLO format by developing an algorithm using python. LabelMe tool was used to test the annotations which were converted to YOLO format.

The subsequent models were tested using the images extracted from the Maxillary Traffic Sign Dataset Which Fascinations over 100,000 high-resolution images. There are nearly 300 classes of traffic signs covering almost all the continents. The dataset contains images varying under different environmental conditions like rain, sun, snow, dawns, daylight, night etc. Few example images from the dataset are given in Fig. 3.5 Due to a large number of data and classes, only selected classes were used for the experiment to cope with the system configurations given in TABLE 1. As the dataset is annotated into

300 classes, the number of images available for some classes were not enough for the training purpose. Thus, the classes were integrated in such a way that both regulatory type and warning type of that particular sign is grouped to represent the same sign. Since the annotations were done in JSON format, a separate algorithm was developed to select the desired classes and to convert the JSON annotations into YOLO format. Labelling tool was used to confirm the correctness of the conversion. Fig. Shows the distribution of a filtered dataset.

4 IMPLEMENTATION

4.1 Deep Learning-Based Approach

- **Use a Convolutional Neural Network (CNN) for end-to-end feature extraction and classification.**
- LeNet: Lightweight, suitable for small datasets.
- VGGNet or ResNet: Deeper models for better performance on large datasets.
- YOLO (You Only Look Once): For real-time detection and recognition.

4.2 Real-Time Recognition

Use frameworks like TensorFlow Lite, ONNX, or PyTorch Mobile for deployment on embedded systems.

Use OpenCV for image acquisition and preprocessing.

4.3 Convolutional Neural Networks (CNNs)

CNNs are a class of deep learning models specifically designed to process and analyze visual data. They are highly effective for traffic sign recognition due to their ability to automatically extract spatial features from images.

4.4 Challenges and Considerations

- The Environmental Factors: Poor lighting or adverse weather conditions can affect recognition accuracy.
- Occlusion: Signs may be partially blocked by objects (e.g., trees or vehicles).
- Real-Time Constraints: Ensure low latency for processing and inference.
- Dataset Imbalance: Some traffic signs are less common, leading to an imbalance in the dataset.

5 RESULTS & DISCUSSION

Step 1: The user can upload the image using upload image option as shown in fig1.

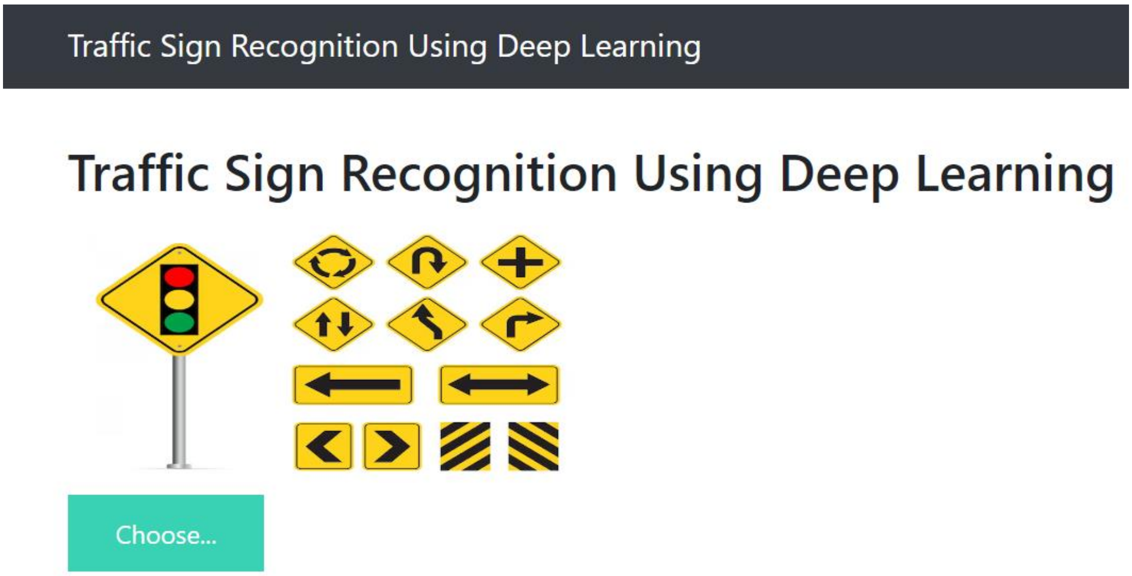

Fig1: upload image

Step 2: We must select the image which must be uploaded which is present in the dataset as shown in fig 2.

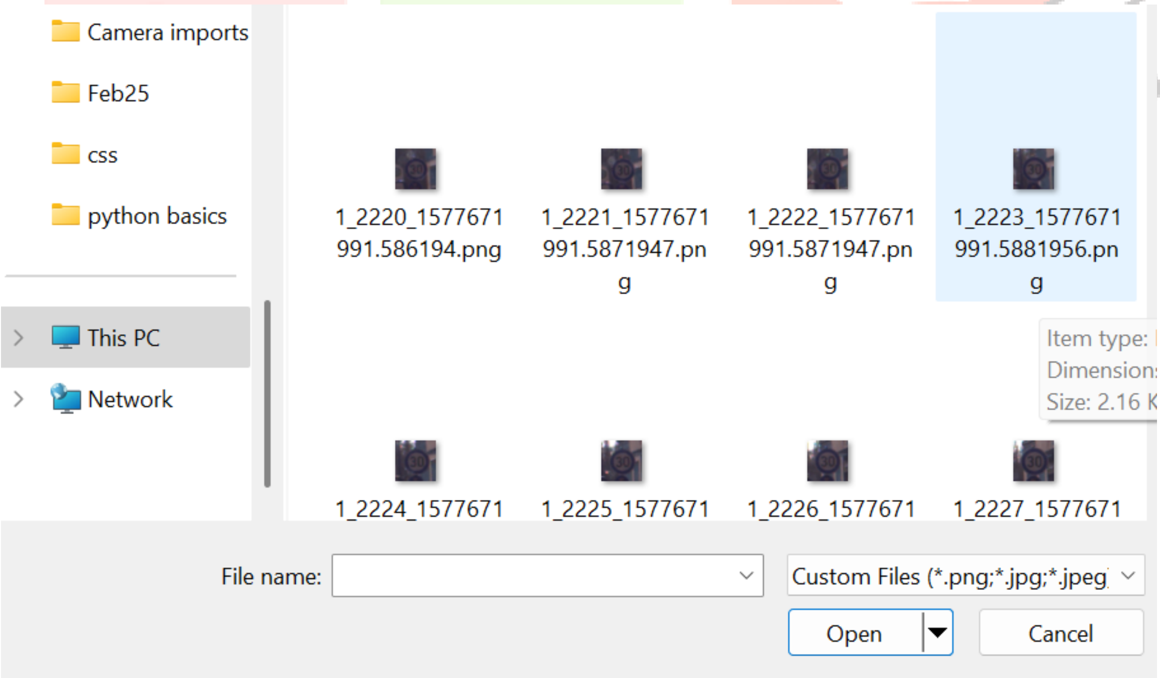

Fig2: select image from the dataset

Step 3: As shown in fig 3 we should upload the image and classify it so the user can read the sign.

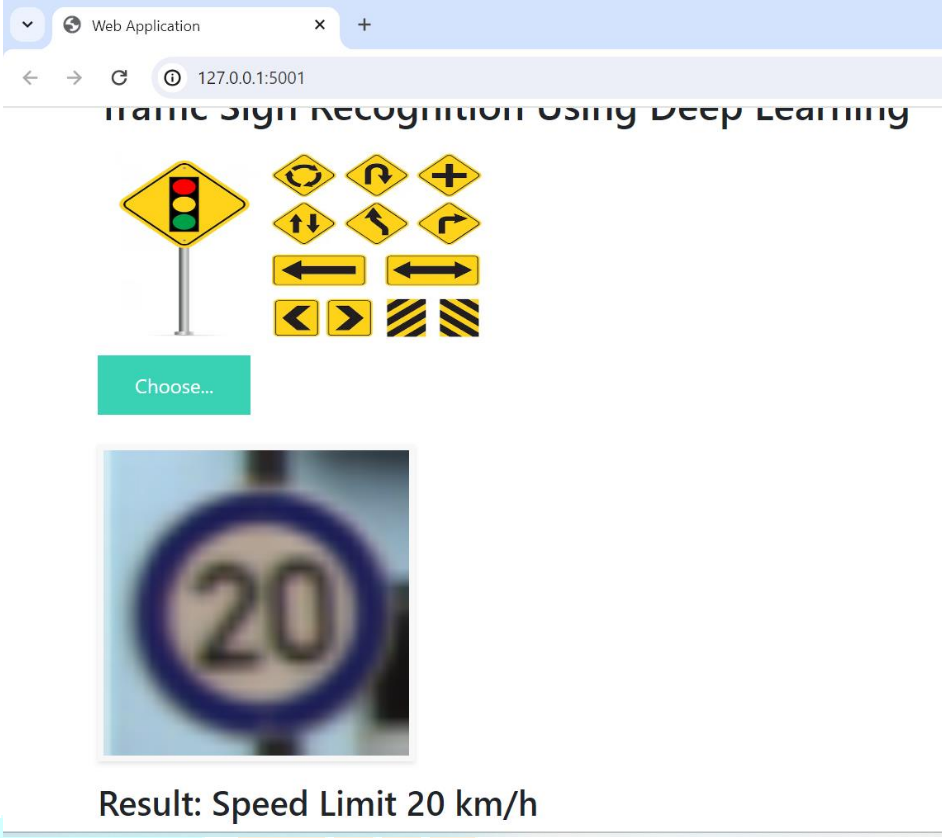

Fig3:image classification

Step 4: Repeat the same as step1, the user can upload the image in dataset again as shown in fig 4.

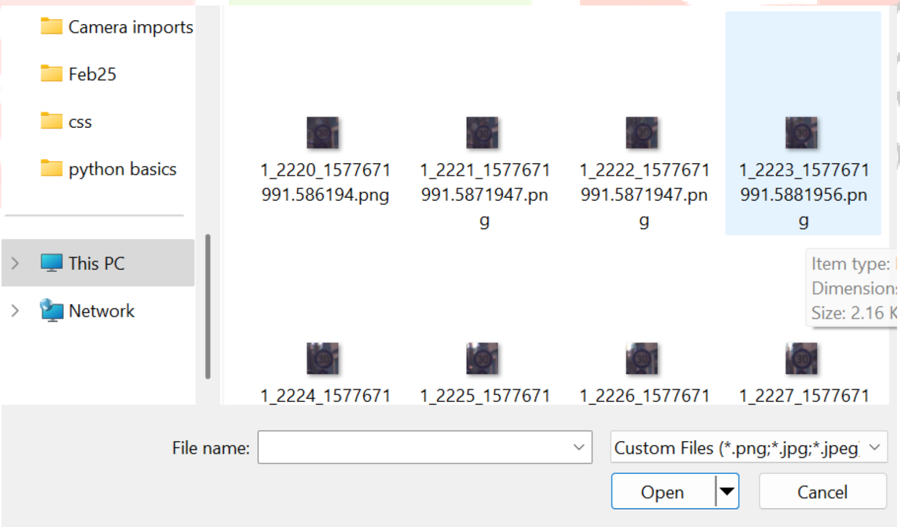

Fig4: upload image

Step 5: As shown in fig 5 the upload image is scanned and classified so the user can read the sign again, which is necessary

Fig5: sign prediction

6 CONCLUSION

In conclusion, a Traffic Sign Recognition (TSR) system using Convolutional Neural Networks (CNNs) proves to be an effective and efficient solution for recognizing and interpreting traffic signs in real-time. CNNs excel at capturing visual features, making them highly suitable for distinguishing among various traffic sign shapes, colors, and symbols. By leveraging CNNs, TSR systems achieve high accuracy and robustness, handling diverse conditions such as different lighting, weather, and partial occlusions.

However, while CNNs offer promising results, challenges remain, particularly in optimizing these models to work efficiently on embedded hardware and to generalize across diverse sign types and regions. Advances in model optimization, transfer learning, and domain adaptation are continuously improving TSR systems, making them more suitable for practical deployment in autonomous vehicles and driver assistance technologies.

Overall, CNN-based TSR systems represent a significant advancement toward safer, more autonomous driving environments, contributing to improved road safety and efficient navigation by enabling vehicles to recognize and respond to traffic signs accurately.

7 ACKNOWELDGEMENT

Authors acknowledge the support from CMR University for the facilities provided to write this article. We also extend our gratitude to the reviewers for their valuable suggestions and constructive feedback.

REFERENCES

[1] Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 779-788, Doi: 10.1109/CVPR.2016.91.

[2] W. Ritter, F. Stein, and R. Janssen, "Traffic sign recognition using colour information," Math. Compute. Model., vol. 22, no. 4-7, pp. 149-161, Aug. 1995, Doi: 10.1016/0895-7177(95)00131- K.

[3] 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. Problem. Eng., vol. 2015, pp. 1-11, 2015, Doi: 10.1155/2015/250461.

[4] S. S. Md Sallah, F. A. Hussin, and M. Z. Yusoff, "Shape-based Road sign detection and recognition for embedded application using MATLAB," in 2010 International Conference on Intelligent and Advanced Systems, Kuala Lumpur, Malaysia, Jun. 2010, pp. 1-5, Doi: 10.1109/ICIAS.2010.5716193.

[5] A. Shustanov and P. Yakimov, "A Method for Traffic Sign Recognition with CNN using GPU:" [in Proceedings of the 14th International Joint Conference on e-Business and Telecommunications, Madrid, Spain, 2017, pp. 42-47, Doi: 10.5220/0006436100420047.

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," IEEE Trans. Pattern Anal. Mach. Intel., vol. 38,no. 1, pp. 142-158, Jan. 2016, Doi: 10.1109/TPAMI.2015.2437384.

[7] H. Fleych, "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.