Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, May : 2024

UTILIZING CNN FOR VOICE-ENABLED REAL-TIME TRAFFIC SIGN RECOGNITION SYSTEM

Dr. E. Balakrishna, *Associate Professor CSE, Vaagdevi College of Engineering (Autonomous), India*
M. A. Touseef Ajmal Khan, *UG Student, CSE, Vaagdevi College of Engineering (Autonomous), India*
Afra Naaz, *UG Student, CSE, Vaagdevi College of Engineering (Autonomous), India*
B. Sahithya Reddy, *UG Student, CSE, Vaagdevi College of Engineering (Autonomous), India*
B. Sindhura, *UG Student, CSE, Vaagdevi College of Engineering (Autonomous), India*

ABSTRACT
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 is capable of assisting 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.

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[1]. 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 behaviors 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[2] 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[3]. 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[4].
To address the above problems, this paper provides a method to detect and recognize traffic signs in real-time with higher accuracy and narrating the signs to the drivers. A system of this type can be used in both vehicle assistive systems and autonomous vehicles[5]. The system is implemented using the Convolutional Neural Network (CNN) model architecture of YOLO. With the faster detection rates and optimized accuracy of the model, the system can be used as a real-time traffic sign detection system. The narration of the message given by a particular traffic sign can assist the drivers while driving. By the voice narration, the issues like missing the traffic signs, lack of familiarity, and the complexity of the traffic signs can be solved.
Our system capitalizes on CNNs' strengths to achieve accurate and swift detection and classification of traffic signs in real-time. Furthermore, by integrating voice commands into the interface, we enhance user experience and facilitate intuitive interaction[6]. Drivers can receive timely auditory notifications about detected signs, allowing them to maintain focus on the road while staying informed

about crucial traffic regulations and warnings[7]. This paper provides an in-depth exploration of the underlying architecture of the CNN model utilized, detailing its design considerations and optimization techniques. Additionally, it discusses the dataset employed for training and evaluation, highlighting the importance of diverse and representative data in achieving robust performance[8].

2. LITERATURE SURVEY

We present YOLO, a new approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, we frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance. Our unified architecture is extremely fast. Our base YOLO[9] model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is less likely to predict false positives on background. Finally, YOLO learns very general representations of objects. It outperforms other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains like artwork.

In this paper, we present a novel approach for the detection and recognition of traffic signs. Colour images are acquired by a camera mounted in a car. In the first step, these images are colour segmented with a pixel classifier. Colour combinations which are characteristic for traffic signs generate hypotheses. These hypotheses are verified using a pictogram classifier.
Our system has been successfully tested on thousands of traffic scenes. The processing of a 512 by 512 frame takes approximately 1 second on a Sparc-10.
This project is part of the European research programme PROMETHEUS and is being developed by Daimler-Benz in collaboration with various university labs[10].

The main objective of this study is to develop an efficient TSDR system which contains an enriched dataset of Malaysian traffic signs. The developed technique is invariant in variable lighting, rotation, translation, and viewing angle and has a low computational time with low false positive rate. The development of the system has three working stages: image preprocessing, detection, and recognition. The system demonstration using a RGB colour segmentation and shape matching followed by support vector machine (SVM) classifier led to promising results with respect to the accuracy of 95.71%, false positive rate (0.9%), and processing time (0.43 s). The area under the receiver operating characteristic (ROC) curves was introduced to statistically evaluate the recognition performance. The accuracy of the developed system is relatively high and the computational time is relatively low which will be helpful for classifying traffic signs especially on high ways around Malaysia. The low false positive rate will increase the system stability and reliability on real-time application.

This paper aims to present three new methods for color detection and segmentation of road signs. The images are taken by a digital camera mounted in a car. The RGB images are converted into IHLS color space, and new methods are applied to extract the colors of the road signs under consideration. The methods are tested on hundreds of outdoor images in different light conditions, and they show high robustness. This project is part of the research taking place in Dalarna University/Sweden in the field of the ITS[11]

In this paper, we propose a road sign detection and recognition algorithm for an embedded application. The algorithm is based on the Hough transform method to detect lines in order to identify and determine the shape of the road sign. Shape measurements are currently employed to identify the road sign ratios of area and perimeter. The variables are compared with the template library which we have developed. In order to evaluate the accuracy, the proposed algorithm was tested on most of the

Malaysian road signs, and the results are reported in this paper. Successful detection and recognition rate is about 83.67%.

In recent years, the deep learning methods for solving classification problem have become extremely popular. Due to its high recognition rate and fast execution, the convolutional neural networks have enhanced most of computer vision tasks, both existing and new ones. In this article, we propose an implementation of traffic signs recognition algorithm using a convolution neural network. Training of the neural network is implemented using the TensorFlow library and massively parallel architecture for multithreaded programming CUDA[12]. The entire procedure for traffic sign detection and recognition is executed in real time on a mobile GPU. The experimental results confirmed high efficiency of the developed computer vision system.

Object detection performance, as measured on the canonical PASCAL VOC Challenge datasets, plateaued in the final years of the competition. The best-performing methods were complex ensemble systems that typically combined multiple low-level image features with high-level context. In this paper, we propose a simple and scalable detection algorithm that improves mean average precision (mAP) by more than 50 percent relative to the previous best result on VOC 2012-achieving a mAP of 62.4 percent. Our approach combines two ideas: (1) one can apply high-capacity convolutional networks (CNNs) to bottom-up region proposals in order to localize and segment objects and (2) when labeled training data are scarce, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, boosts performance significantly. Since we combine region proposals with CNNs, we call the resulting model an R-CNN or Region-based Convolutional Network. Source code for the complete system is available at http://www.cs.berkeley.edu/~rbg/rcnn.

The traffic sign recognition process includes two parts: detection and classification. In this paper, we use an object detection algorithm called SSD to detect the traffic signs. This convolutional neural network uses multiple feature maps to detect objects[13]. For the traffic sign is very small to the whole picture, the SSD model has been improved to have a better detection result of traffic signs. In the experiments, the model has been simplified and the size of the prior box has been modified. The improved network has a good detection effect on small targets. The results on the test data set show that the proposed algorithm performs well for single-target, multi-target and dark-light images. The precision and recall on the test data set are 91.09%, and 88.06%.

There are mainly two types of state-of-the-art object detectors. On one hand, we have two-stage detectors, such as Faster R-CNN (Region-based Convolutional Neural Networks) or Mask R-CNN, that (i) use a Region Proposal Network to generate regions of interests in the first stage and (ii) send the region proposals down the pipeline for object classification and bounding-box regression. Such models reach the highest accuracy rates, but are typically slower. On the other hand, we have single-stage detectors, such as YOLO (You Only Look Once) and SSD (Singe Shot MultiBox Detector), that treat object detection as a simple regression problem by taking an input image and learning the class probabilities and bounding box coordinates. Such models reach lower accuracy rates, but are much faster than two-stage object detectors. In this paper, we propose to use an image difficulty predictor to achieve an optimal trade-off between accuracy and speed in object detection. The image difficulty predictor is applied on the test images to split them into easy versus hard images. Once separated, the easy images are sent to the faster single-stage detector, while the hard images are sent to the more accurate two-stage detector. Our experiments on PASCAL VOC 2007 show that using image difficulty compares favorably to a random split of the images. Our method is flexible, in that it allows to choose a desired threshold for splitting the images into easy versus hard.

Existing deep convolutional neural networks (CNNs) require a fixed-size (e.g., 224x224) input image. This requirement is "artificial" and may reduce the recognition accuracy for the images or sub-images of an arbitrary size/scale. In this work, we equip the networks with another pooling strategy, "spatial pyramid pooling", to eliminate the above requirement. The new network structure, called SPP-net, can generate a fixed-length representation regardless of image size/scale. Pyramid pooling is also robust to object deformations. With these advantages, SPP-net should in general improve all CNN-based image classification methods. On the ImageNet 2012 dataset, we demonstrate that SPP-net boosts the accuracy of a variety of CNN architectures despite their different designs. On the Pascal VOC 2007 and Caltech101 datasets, SPP-net achieves state-of-the-art classification results using a single full-image representation and no fine-tuning. The power of SPP-net is also significant in object detection. Using SPP-net, we compute the feature maps from the entire image only once, and then pool features in arbitrary regions (sub-images) to generate fixed-length representations for training the detectors. This method avoids repeatedly computing the convolutional features. In processing test images, our method is 24-102x faster than the R-CNN method, while achieving better or comparable accuracy on Pascal VOC 2007. In ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014, our methods rank #2 in object detection and #3 in image classification among all 38 teams[14]. This manuscript also introduces the improvement made for this competition.

3. PROBLEM STATEMENT

Because there are so many traffic signs in urbanized regions, drivers in less urbanized neighborhoods could have trouble understanding what a particular sign is trying to say. Certain traffic signs are often ignored by drivers who think they are unnecessary. This ignorance is also caused by the drivers' disparate views.

Few Disadvantages of the Existing System are:

- Some driver my not familiar with the traffic sings could result in severe accidents.
- There is a chance of death.
- There is No voice assisted system with the traditional system.

4. PROPOSED SYSTEM

We are working on a technique that will allow us to more accurately detect and interpret traffic signs in real time while also explaining them to drivers. This kind of system can be utilized in autonomous automobiles as well as vehicle assisting systems. Convolutional Neural Network (CNN) model[15], which has faster detection rates, is used to create the system.

Few Advantages of Proposed System are:

- The system can be used as a real-time traffic sign detection system
- The narration of the message given by a particular traffic sign can assist the drivers while driving. By the voice narration.
- The narration of the message given by a particular traffic sign can assist the drivers while driving. By the voice narration.

5. IMPLEMENTATION

5.1.i Upload Dataset

Using this module, we can load traffic signs dataset from the location of the project to train the CNN Algorithm.

5.1.ii Generate Training And Testing Images

- ImageDataGenerator: that rescales the image, applies shear in some range, zooms the image and does horizontal flipping with the image. This ImageDataGenerator includes all possible orientation of the image.
- train_datagen.flow_from_directory is the function that is used to prepare data from the train_dataset directory Target_size specifies the target size of the image.
- test_datagen.flow_from_directory is used to prepare test data for the model and all is similar as above.
- fit_generator is used to fit the data into the model made above, other factors used are steps_per_epochs tell us about the number of times the model will execute for the training data.
- epochs tell us the number of times model will be trained in forward and backward pass.

5.1.iii Generate CNN Model

In this module we are generating CNN Model with train_datagen and test_datagen generated by ImageDataGenerator class.

Here we have trained this CNN algorithm multiple time to get the better accuracy using epochs. Finally, we will get the best CNN model with average accuracy 99%.

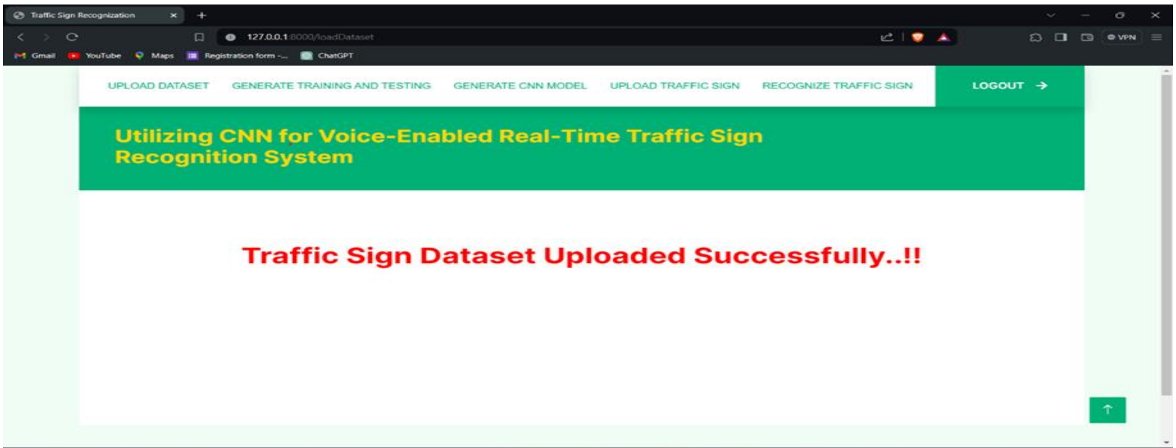
5.1.iv Upload Traffic Sign

Using this module, we can upload test image and pass the test image to the model to recognize traffic sign.

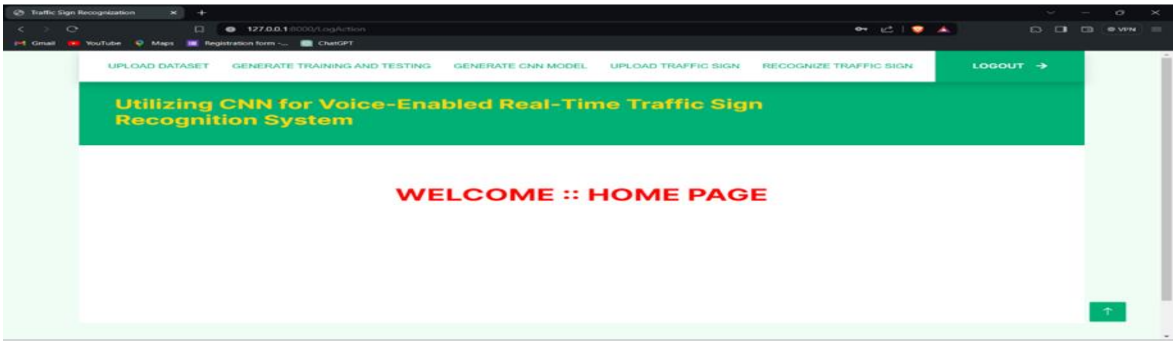
5.1.v Recognize Traffic Sign

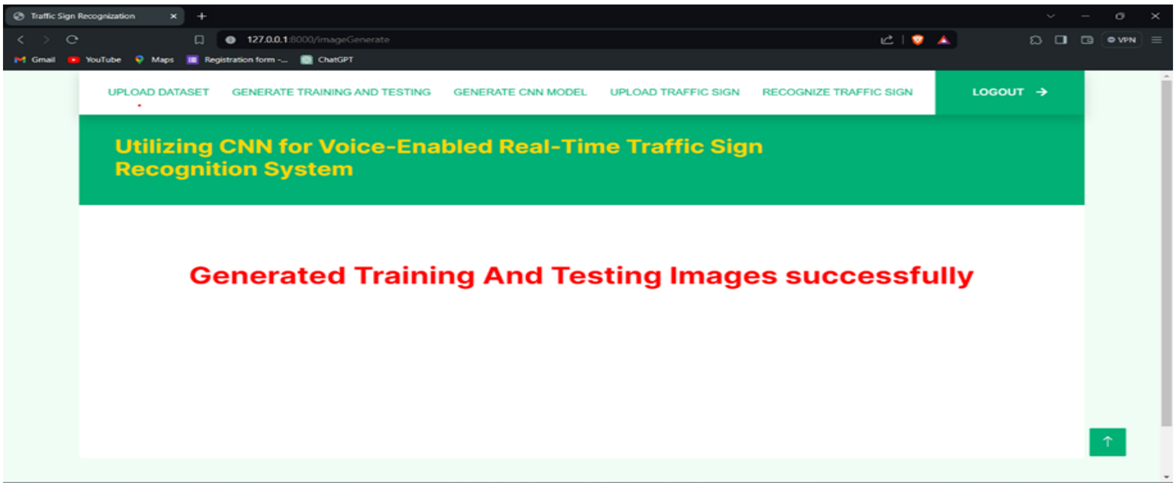
Using this module, we will call the CNN Model which is already generated and take the image from the 4th step and pass to model.Then the model will identify the traffic sign and gives voice alert message.

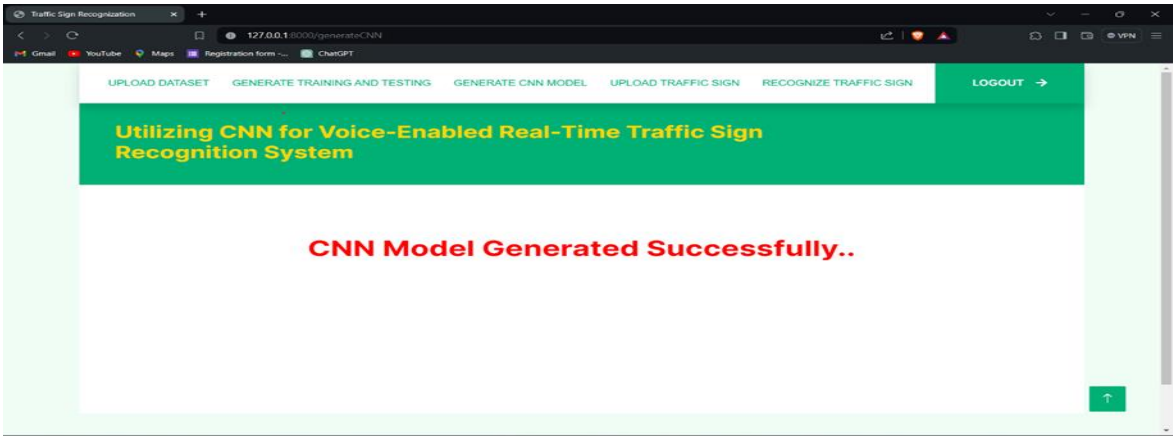
6. RESULTS/DISCUSSION

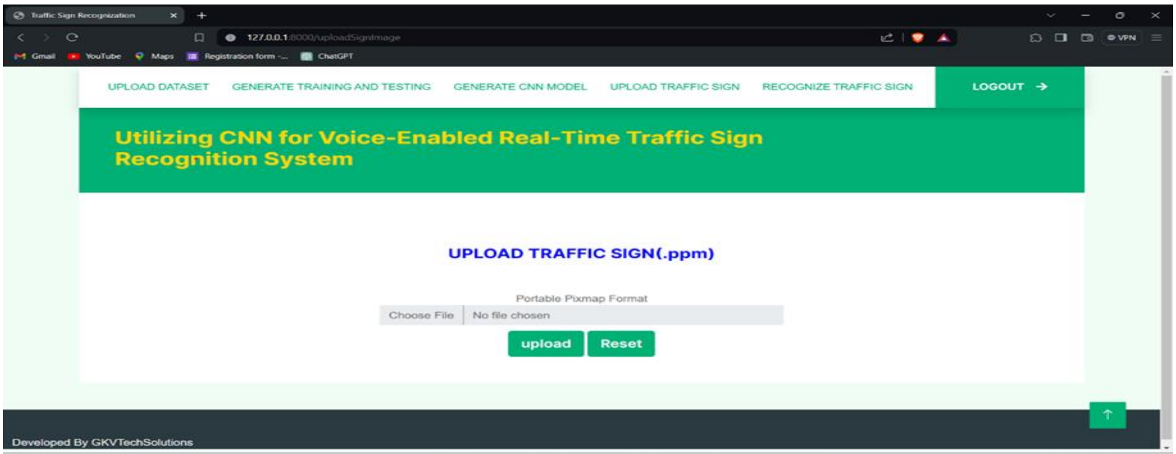


Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, May : 2024

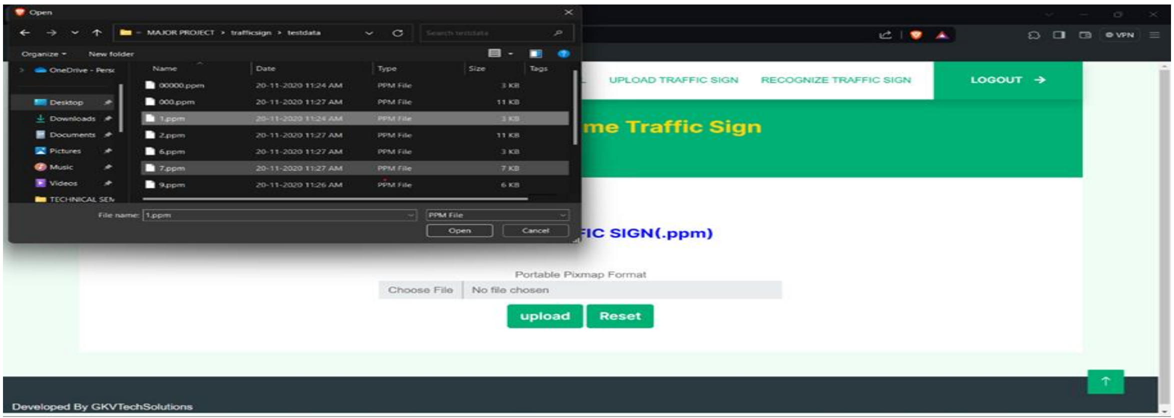


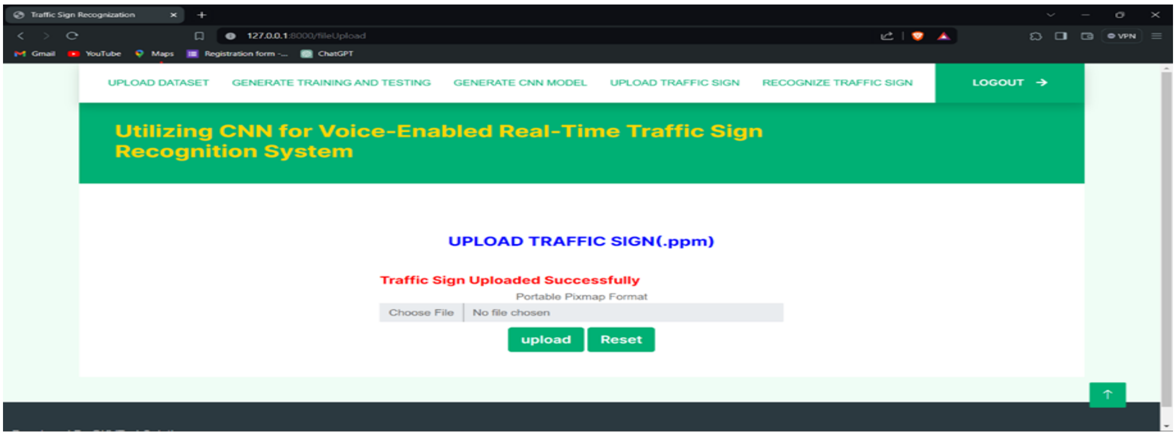


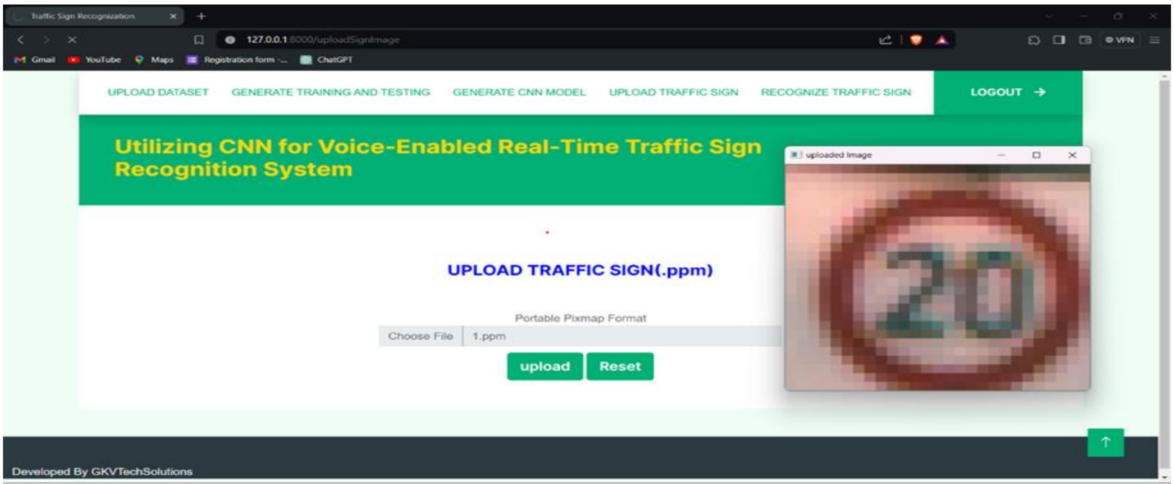


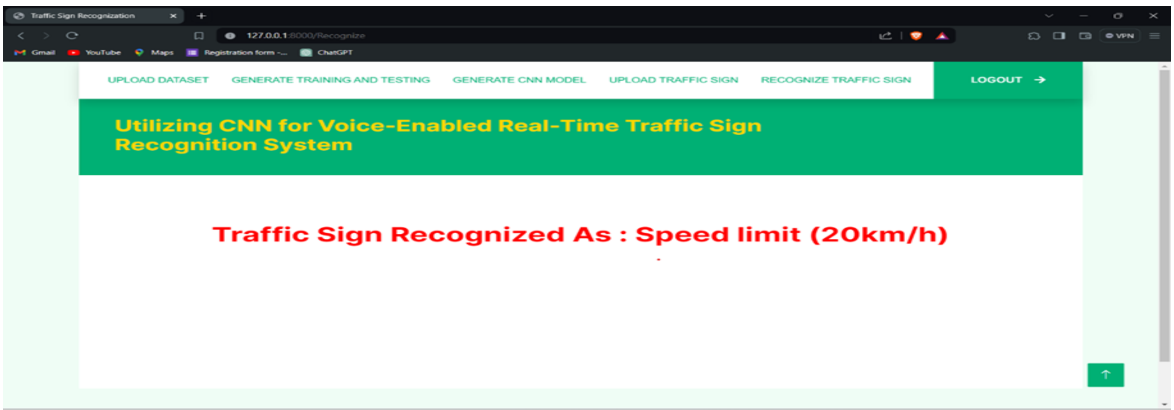


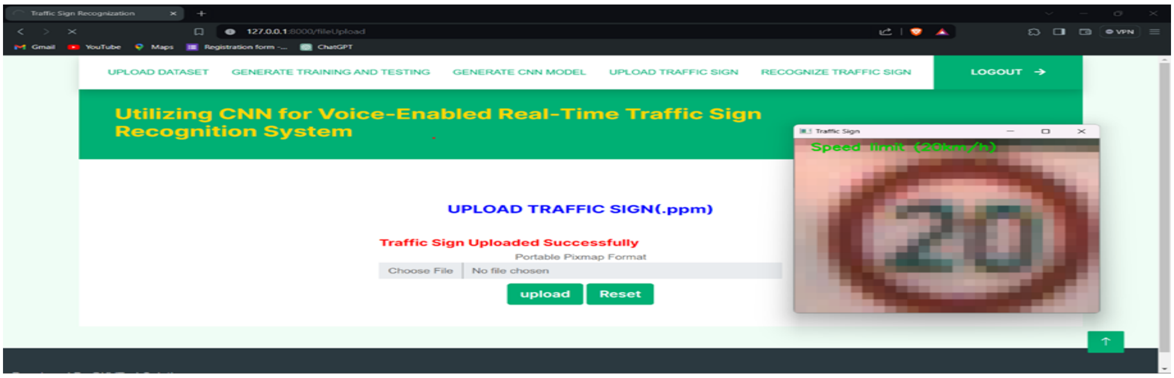
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, May : 2024











7. CONCLUSION

In this work, we have presented a robust real-time traffic sign detection approach with the detection speed of 55 FPS. We have also achieved a mean average precision of 64.71%. The accuracy of the presented approach can be increased further by adjusting the configurations of the YOLO architecture and tuning the hyper-parameters while maintaining the detection speed consistently. The adverse effect of partially occluded traffic signs, damaged traffic signs, and extreme weather conditions can further be decreased by applying techniques such as presenting the CNN with partly visible signs, applying 3D reconstruction algorithms and fuzzy C-means clustering.

The model we presented in this paper can detect traffic signs at a very high frame rate of 55 FPS and could achieve mean average precision of 64.71%. Having a frame rate of over 30 FPS guarantees the real-time performance of the system. Further, the voice assistant feature along with accurate detection can solve most of the problems which are caused due to the missing or not being aware of the traffic signs.

8. FUTURE SCOPE

In future, we would like to extend the training of the model to the whole dataset with high-end GPU devices. To increase the accuracy further, we see the potential of modifying the architecture of YOLO detector. With the accuracy increased, we can embed the system into a single board PC within the vehicles to assist the drivers.

9.REFERENCES

[1] J. 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. Comput. 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. Probl. Eng., vol. 2015, pp. 1–11, 2015, doi: 10.1155/2015/250461.
[4] 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.
[5] 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.
[6] 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.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 1, pp. 142–158, Jan. 2016, doi: 10.1109/TPAMI.2015.2437384.
[8] B. Gao, Z. Jiang, and J. zhang, "Traffic Sign Detection based on SSD," in Proceedings of the 2019 4th International Conference on Automation, Control and Robotics Engineering - CACRE2019, Shenzhen, China, 2019, pp. 1–6, doi: 10.1145/3351917.3351988.
[9] 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.
[10] 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.
[11] 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.
[12] A. Shustanov and P. Yakimov, "CNN Design for Real-Time Traffic Sign Recognition," Procedia Eng., vol. 201, pp. 718–725, 2017, doi: 10.1016/j.proeng.2017.09.594.
[13] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020.
[14] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," ArXiv180402767 Cs, Apr. 2018, Accessed: Feb. 07, 2021. [Online]. Available: http://arxiv.org/abs/1804.02767.
[15] C. Ertler, J. Mislej, T. Ollmann, L. Porzi, G. Neuhold, and Y. Kuang, "The Mapillary Traffic Sign Dataset for Detection and Classification on a Global Scale," ArXiv190904422 Cs, May 2020.