iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** I **Month of publication:** January 2023

DOI: https://doi.org/10.22214/ijraset.2023.48555

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue I Jan 2023- Available at www.ijraset.com

Traffic Sign Identification Using Modified LeNet-5 CNN

Nataraja Chandra Shekhar Somavarapu[1], S Shreekar Goud Bomma[2], Neha Reddy Kaluvala[3], Sriharsha Vardhan Puchakayala[4]
[1, 2, 3, 4]*Electronics and Communication Engineering, Sreenidhi Institute of Science and Technology, India*

Abstract: ***In 21st century, it is obvious that automobiles have become the standard mode of transportation. With the spread of their usage, road traffic becomes more intricate; hence road safety must be promising enough avoiding accidents. Drivers may sometimes overlook the sign boards along the roads which might lead to undesirable casualties. This could be helped by a system that could assist the driver in keeping an eye on the traffic sign boards on the way. It can be achieved using CNN i.e. Convolutional Neural Networks. CNNs are preferred when the applications are concerned with computer vision or image recognition. A deep learning CNN model is built in the LeNet-5 architecture with some modifications and is trained with traffic signs images from GTSRB dataset which stands for German traffic sign recognition benchmark with the help of OpenCV, Tensorflow, Keras and other libraries. The trained model is tested with test data and is observed to be recognizing any of the traffic signs which were already learnt by the network. Vehicle cameras come handy to capture the real-time sign boards. Implementation of this supervised learning model in real-time would aid road safety through vehicle control.***

Keywords: Convolutional Neural Networks, Supervised learning, Deep-Learning, Traffic signs, GTSRB, Computer vision, LeNet-5 Architecture

I. INTRODUCTION

Transportation has been holding a prominent place in the chapter of human civilisation from centuries. The advancements can be witnessed with automobiles being introduced with new features every now and then. With wide spread of usage of vehicles, it is necessary to maintain public transport and road safety measures to the best. Often it is evident that neglecting the traffic sign boards by the roads while driving results in many fatalities. It could be helped if a driver support system is installed in the vehicles. This system should be able to help the person who is driving by identifying the sign boards in the journey and intimating the same to him/her so that the driver gets appropriate alerts and drives cautiously. The paper explains how it can be achieved in real-time. To serve the purpose, a prototype is built where standard traffic sign dataset by Kaggle is collected to train a modified LeNet-5 convolutional neural network model. Pycharm, a Python environment is availed of for building, training and testing the model. Libraries like OpenCV which is an open source computer vision (CV) library, Tensorflow, Keras etc. were helpful in the process. When a trained model encounters a sign board, it detects and recognizes if it belongs to any of the trained classes of traffic signs based on supervised learning and sends corresponding message-alert to the user in response.

II. BACKGROUND THEORY

A. *Supervised Learning*

In the era of artificial intelligence where machine mimic human characteristics to make our lives simpler and comfortable, it is important for the model to adopt a suitable machine learning technique. A model is nothing but an artificial neural network which is a mathematical representation human neuron. The most widely known machine learning techniques include the following.

1) Supervised learning
2) Unsupervised learning
3) Semi-Supervised learning
4) Reinforcement learning

Supervised learning, also known as inductive learning deals with the training data to which labels are already provided. In other words, when a model trained with supervised learning technique is tested with test data, we get desired outputs or labels i.e. known results are obtained. So when unknown/ unseen data is given, the model would have the knowledge of what class or category it would fall into from the training labels. The algorithms like logistic regression and back propagation neural network could be the examples of this kind of learning method. It is best suited for classification and regression problems.

©IJRASET: All Rights are Reserved

B. Computer Vision

As the name suggests, computer vision is an integrative field that handles the way a computer system acquired top notch compression from digital media such as photos and videos. It can be framed as mimicking the human visionary system in the following steps.

1) Obtaining or acquiring the images
2) Pre-Processing them
3) Analysing the features or parameters in order to understand them and extraction
4) Object detection or segmentation
5) Reconstruction and Interpretation
6) Decision making

Since the data dealt with is mostly digital images and videos, computer vision is interrelated to the concepts like digital image and video processing, object detection and recognition etc. the application areas include medicine, military, autonomous vehicles etc., especially where inspection and identification tasks are predominant.

III.LITERATURE SURVEY

It might be quite unbelievable but is a fact that the death toll due to road accidents worldwide is around 13 lakhs every year. And the major causes that account to this misfortune are over speeding, absent-minded driving, drunken driving, signal jumping etc. Thus, it is a necessity to make sure rules and regulations are strictly followed. However, to err is human; sometimes signs may get overlooked by the drivers. This issue has been being addressed by many researchers with driver-assistance systems in vehicles which can capture sign boards and make drivers alert.

Many algorithms can serve the purpose of TSR (Traffic Sign Recognition) in practice. But what makes difference is the efficiency with which it is done. Earlier, template matching algorithms such as GHT which stands for Generalized Hough Transform were in use but it was not efficient when it came to real-time testing where sign boards captured are at wide range of illuminations. But the main challenge in TSR is the accuracy in identifying the signs. Also, trade-off has to be maintained between the model depth and dataset size. Early detection of objects mostly depends on the prominent features like edges and the separable regions like HoG (Histogram of Gradient) [1]. Different networks result in different forms of outputs which include boundary estimation, pose estimation, considering it as a pose and shape prediction problem etc. [2].

In general, there are two sub-tasks in TSR [3]: localization and classification. Localization refers to locating the objects in an image, in other terms, detecting the objects whereas classification is affirming the category or class into which an image falls. Since the whole operation is on images in real-time, interdisciplinary knowledge is required such as computer vision (CV) research area in order to suggest new methods and more efficient solutions [4]. CV is the field of study on how a system understands and interprets a digital image or video in real-time [5].

Reference [6] shows how CNN has recently come into practice to eliminate the features or elements, ultimately the images which do not belong to Traffic signs. HOG (Histogram of Oriented Gradients) feature extraction process helps in determining the basic shapes like circles, triangles etc. in the sign boards. It is efficient in object recognition tasks with a higher rate [7], [8].

Since the objective is to identify a traffic sign, the datasets GTSRB and GTSDB (German Traffic Sign Detection Benchmark) are most popularly utilized. Another option available is the Belgium TS (Traffic Sign) dataset. CNN is the widely used deep neural network [9] which can be constructed in any of the architectural frameworks such as LeNet, AlexNet, ResNet, CapsNet, MobileNet, SqueezeNet [10] etc. studies show that LeNet-5 architecture results better than CapsNet CNN. There is no perfect model yet; the research is still going on to obtain 100 per cent accuracy in any sort of challenging conditions.

IV.PROPOSED METHODOLOGY

A neural network model which can be first trained with a set of images of different traffic signs present in the dataset is proposed. The images are in different alignments and orientations. This model can be implemented in vehicles in order to identify the sign boards with prior knowledge of supervised learning, thus being an aid to the drivers in cautious driving. A convolutional neural network model is built using LeNet-5 architecture with some changes. There are two major tasks as shown- training and testing. At first, a neural network is built and is trained with GTSRB dataset which is in other words called a model. This trained model then undergoes testing phase where the results are expected.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

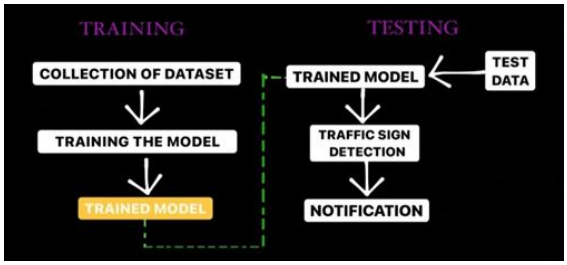


Fig. 1 Implementation flow

A. Collection of Dataset

The CNN model takes image dataset as input and performs myriad mathematical operations and transforms. As the concern is traffic signs here, we use a standard classification dataset of images which is known to be GTSRB. Figure 2 shows forty three classes of training data. Each traffic sign is called a class with several images each at different alignments and illuminations. It is evident from the figure that the classes have unequal number of images corresponding to distinct signs.



Fig. 2 Traffic signs in GTSRB

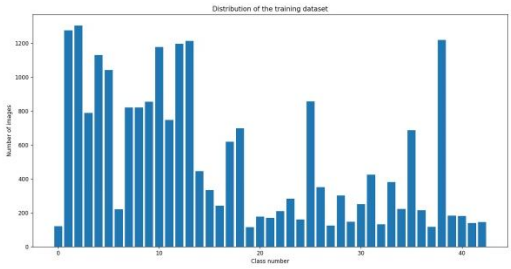


Fig. 3 Unequal image data

And to make use of available images effectively, we apply some mathematical operations on them in order to fetch different versions of them which would help train the model better.

B. Creating and training a model

Creating a model is nothing but building a neural network with specifications. We build a convolutional neural network in LeNet-5 architecture with some modifications in it. A CNN comprises the following.

1) Feature Learning: This is a training stage and comprises convolution layers and pooling layers.

A convolutional layer can be considered the basic functional unit or stage in a CNN. Its function is extracting the key features in a picture. The more the convolutional layers, the more are the important and complex features. A convolutional layer has an activation function, often ReLu in order to deal with the nonlinearities in the data and number of filters which are applied on an image to result in feature maps. Hence these features are also called feature detectors or kernels. In any CNN architecture it can be observed that the convolutional layers are followed by respective pooling layers. Pooling is the operation where the size of feature map is brought down. Hence, the number of convolutional and pooling layers determines how deep a neural network is.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Pooling basically replaces the whole region covered by a filter by a single value. The value which replaces the region is dependent on the operation or pooling technique used. Out of many pooling techniques available, two are widely used and are known as max pooling and average pooling. In max pooling, the replacing value is the maximum value of the feature map i.e. the most prominent feature of a feature map gets reflected in the next feature map when max pooling is applied. Likewise, in average pooling, the average value of the region of feature map gets reflected at the output of pooling layer.
With the increase in depth, the receptive field gets enlarged, thus reducing the computational costs.

2) Classification: This step is also called as inference and is application-oriented.
After all the convolution and pooling operations are done, the matrix or pixel form in of images is converted into one-dimension, in other terms, vector form. This process, called flattening, enables further classification.
Fully connected layer, in short FC plays a role of a medium between humans and network language significantly when the network is deep and it becomes difficult for the human comprehension of the features extracted since they are considerably abstract. There are several nodes in these layers. Each node in the previous layer is connected to every node in an FC.
Next to the last FC is the output layer which is a classifier. Softmax or Logistic layer is used in general. A logistic layer is used when the classification is binary i.e. into two classes; when there are multiple classes, softmax is preferred. This layer has as many nodes as the number of classes into which classification has to be done. Since there are 43 different signs here, we have 43 nodes in the Softmax layer as in [11].

$$\begin{matrix}2.0\\1.0\\0.1\end{matrix} \rightarrow \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} \rightarrow \begin{matrix}0.7\\0.2\\0.1\end{matrix} \quad (1)$$

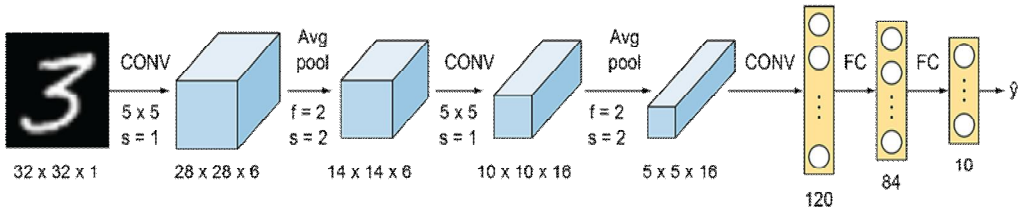


Fig. 4 LeNet-5 CNN model

Normalisation of values is done here so that they fall between 0 and 1 thus giving probabilities for each class. The same is depicted in Eq. (1). The class with the highest probability will be the final output. The normalisation feature is useful when the network has to be appended to another network.
The current architecture comprises three convolution, two pooling and two FC layers. A softmax output layer is the classifier. Max pooling technique is made use of in the pooling layers and to handle the non-linear behaviour of the input images, tanh function is used in convolutional layers for activation. Figure 6 shows the LeNet architecture of a CNN.
The working model is a primary LeNet-5 CNN model other than the fact that the activation function used alongside convolution is ReLU function instead of Tanh function. All the seven layers, activation function and the desired pooling technique can be imported with the help of python libraries such as openCV, Tensorflow, Keras, numpy etc.
After the model is ready with the required specifications, it has to be trained with the selected dataset. Training process includes building correspondence between the signs and their meanings and making the model learn the same. It is done for all available signs.

C. Testing and Notification
After a model gets trained with certain data, in specific called training data, it is expected to show desirable results on unknown or unseen data which is referred to as the test data. The whole data in this regard is in the form of images. The GTSRB dataset with about 40,000 images is divided into two portions: the larger for training and the other portion of images for self-testing. This is when the model challenges itself to recognize test images and the efficiency at this test defines the accuracy of the model.
Then, it can be tested in real-time with problem images. When such an image gets captured by the interfaced camera module, the model detects and recognizes the traffic sign embedded in it. Fig.5 shows a sign board which is known to the trained model.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



Fig. 5 Road works sign board

When a particular sign board is identified by the model, it indicates the driver about the same. Each traffic sign is nothing but a pictorial representation of a caution and the same gets reflected as an alert message. Figure 8 depicts how the sign when placed in front of the camera gets detected and recognised and prints the corresponding meaning in statement.

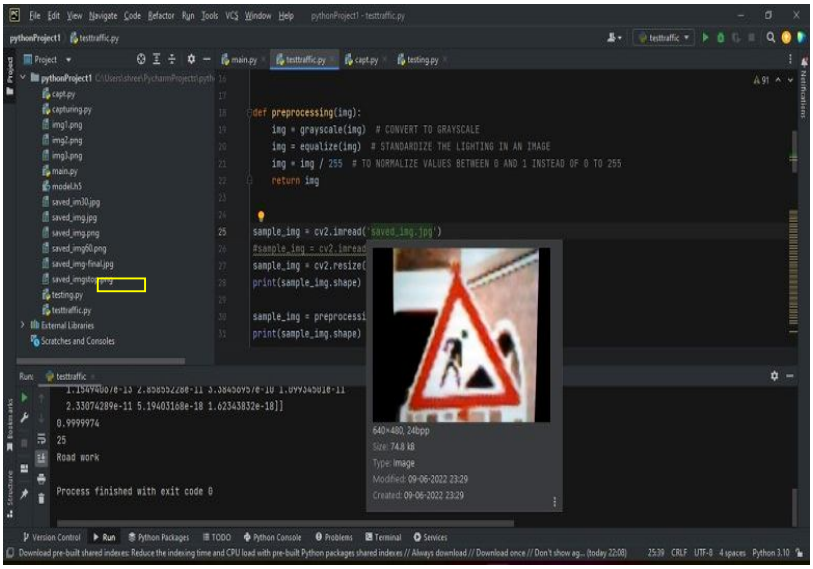


Fig. 6 Message in console

V. TEST RESULTS

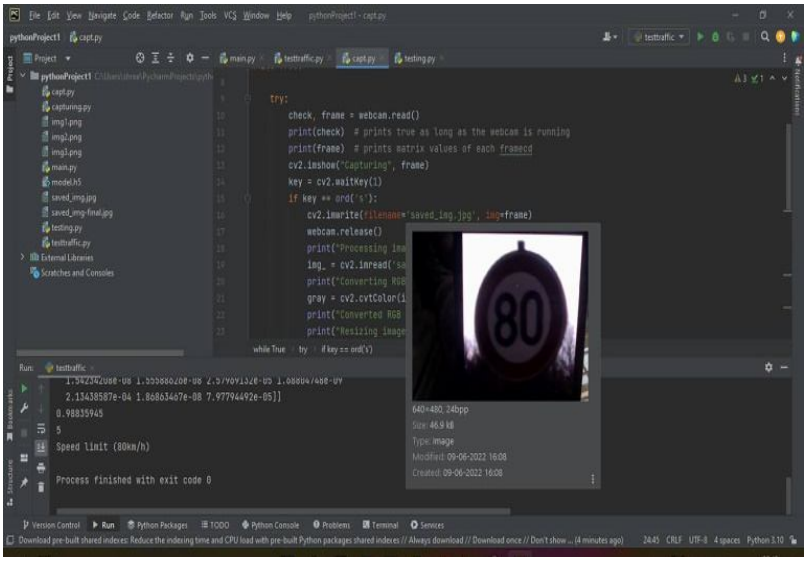


Fig. 7 Test result for 80kmph sign

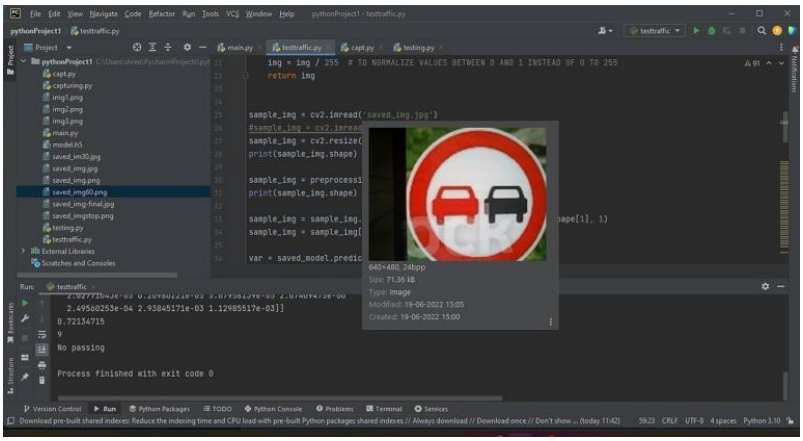


Fig. 8 Test result for No passing sign

The model trained with 43 different traffic sign images is tested with unseen images which contribute to test data to obtain desired results. As soon as the python code is run, all the classes of traffic signs get installed. Since the dataset is huge, it is given in 10 epochs i.e. the whole set is divided into ten portions to train the model. Once they are all loaded and ten epochs get completed, we run the camera module program in order to activate the integrated camera of the system. Camera turns on and detects if any traffic sign is present in its scope.

One such sign of 80kmph speed is used to demonstrate the performance of the model. It is brought into the range of the camera so that the sign gets detected. The digits eight and zero together bound by a circle refers to the speed limit indication board. Since 80kmph sign is one of the trained classes i.e. the model has learnt it before as a part of the supervised learning, the image gets recognised as a distinct class with the corresponding label and the same message, here 'Speed Limit (80km/h)' is displayed in the console and in the same way, the 'No passing' sign.

VI. CONCLUSIONS

A supervised machine learning model is built in python language. Pycharm environment has been helpful since most of the required libraries could be installed and accessed through it. A convolutional neural network in LeNet-5 architecture is constructed with minor changes; rectified linear units are added to the convolution layers instead of Tanh functions as in regular LeNet-5 structure. The model has seven layers in total. GTSRB is an open source dataset with about forty thousand images of forty three distinct traffic signs. Pre-built model is trained with these images as 43 separate classes and the corresponding notations.

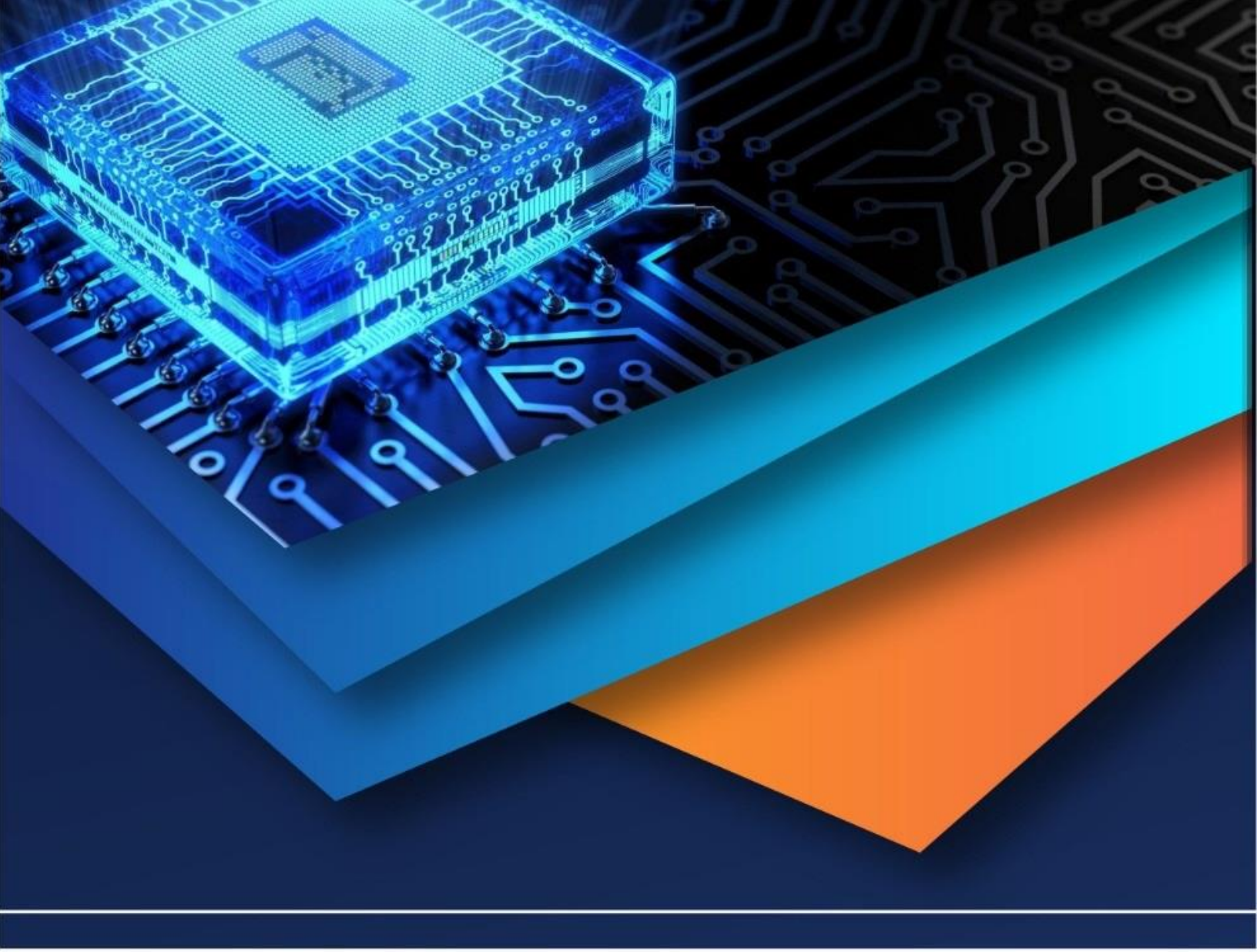
With computer vision, it has become possible to work on the image dataset of standard signs. A distinct traffic sign is recognised with the help of HSV parameter (HSV stands for hue saturation-value) and it is all a part of learning. After training gets completed, model is tested with test images. In order to capture them, a camera module is required. Integrated camera of a laptop is used. A python code is written for it to accept images and those images to be tested by the trained model. Hence, necessary interfacing is done between camera, testing and training modules and it is observed that the trained signs are recognised by the model correctly with an accuracy of about 98 per cent. Based on the results obtained from the prototype model, it is expected that the performance would be even better in real-time application. An efficient model is expected to identify the signs at different illuminations placed at different angles and even in case of disturbed backgrounds. The proposed model can be considered to be modest. Accuracy in classification is appreciable on standard traffic sign dataset- GTSRB as well as the new or unseen data of the mentioned forty-three signs in image form. Implementation of such a system in automobiles would be a smart step towards road safety.

REFERENCES

[1] CNN Design for Real-Time Traffic Sign Recognition Alexander Shustanov, Pavel Yakimov 3rd International Conference, ITNT-2017, Samara, Russia
[2] Simultaneous Traffic Sign Detection and Boundary Estimation using Convolutional Neural Network Hee Seok Lee and Kang Kim IEEE Transactions On Intelligent Transportation Systems 27 February 2018
[3] A Novel Neural Network Model for Traffic Sign Detection and Recognition under Extreme Conditions Haifeng Wan, Lei Gao, Hindawi, Journal of Sensors, Volume 2021.
[4] Robust Chinese Traffic Sign Detection and Recognition with Deep Convolutional Neural Network Rongqiang Qian, Bailing Zhang, Yong Yue, Zhao Wang, Frans Coenen 2015 11th International Conference on Natural Computation (ICNC)

[5] Road Safety Awareness and Comprehension of Road Signs from International Tourist's Perspectives: A Case Study of Thailand Kasem Choocharukul, Kerkritt Sriroongvikrai Elsevier World Conference On Transport Research Society 2017
[6] Transfer learning based hybrid 2D-3D CNN for traffic sign recognition and semantic road detection applied in advanced driver assistance systems Khaled Bayoudh, Faycal Hamdaoui, Abdellatif Mtibaa Springer Science+Business Media, LLC, part of Springer Nature 2020
[7] Traffic Sign Recognition with Convolutional Neural Network Based on Max Pooling Positions Rongqiang Qian, Yong Yue, Frans Coenen and Bailing Zhang 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)
[8] Two-stage traffic sign detection and recognition based on SVM and convolutional neural networks Ahmed Hechri, Abdellatif Mtibaa The Institution of Engineering and Technology 2019
[9] Manman Su, Qinglong You, Hui Qu and Qirun Sun Hindawi Journal of Sensors Volume 2021
[10] Traffic sign recognition using weighted multiconvolutional neural network Sudha Natarajan, Abhishek Kumar Annamraju, Chaitree Sham Baradkar The Institution of Engineering and Technology 2018
[11] CNN based Traffic Sign Detection and Recognition on Real Time Video Mrs. Deepali Patil, Ashika Poojari, Jayesh Choudhary, Siddhath Gaglani International Journal of Engineering Research & Technology (IJERT) Special Issue - 2021

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)