ARTICLE IN PRESS

International Journal of Transportation Science and Technology xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

journal homepage: www.elsevier.com/locate/ijtst

Indian traffic sign detection and recognition using deep learning

Rajesh Kannan Megalingam *, Kondareddy Thanigundala, Sreevatsava Reddy Musani, Hemanth Nidamanuru, Lokesh Gadde

Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, India

ARTICLE INFO

Article history:
Received 28 August 2021
Received in revised form 23 January 2022
Accepted 16 June 2022
Available online xxxx

Keywords:
Refined Mask R-CNN
Fast R-CNN
Data augmentation
Pre-processing
Custom dataset
Indian Traffic Sign

ABSTRACT

Traffic signs play a crucial role in managing traffic on the road, disciplining the drivers, thereby preventing injury, property damage, and fatalities. Traffic sign management with automatic detection and recognition is very much part of any Intelligent Transportation System (ITS). In this era of self-driving vehicles, calls for automatic detection and recognition of traffic signs cannot be overstated. This paper presents a deep-learning-based autonomous scheme for cognizance of traffic signs in India. The automatic traffic sign detection and recognition was conceived on a Convolutional Neural Network (CNN)- Refined Mask R-CNN (RM R-CNN)-based end-to-end learning. The proffered concept was appraised via an innovative dataset comprised of 6480 images that constituted 7056 instances of Indian traffic signs grouped into 87 categories. We present several refinements to the Mask R-CNN model both in architecture and data augmentation. We have considered highly challenging Indian traffic sign categories which are not yet reported in previous works. The dataset for training and testing of the proposed model is obtained by capturing images in real-time on Indian roads. The evaluation results indicate lower than 3% error. Furthermore, RM R-CNN's performance was compared with the conventional deep neural network architectures such as Fast R-CNN and Mask R-CNN. Our proposed model achieved precision of 97.08% which is higher than precision obtained by Mask R-CNN and Faster R-CNN models.

Introduction

Smart transport vehicles outfitted with intelligence, including driverless vehicles, call for autonomous traffic sign detection and recognition (TSDR). Recent times have witnessed proliferation of diverse schemes, for the purpose of automatic traffic sign (traffic sign and traffic signal are used interchangeably) detection and recognition; reference to this was first reported in 1990 in Kamada et al. (1990). A significant number of investigations have already been carried out with excellent results, using computer vision for automatic TSDR, as shown in Zhu et al. (2016a,b). Majority of the TSDR work, carried out for traffic signs, outside India, i.e., dealt with a limited number of categories associated with Advanced Driver Assistance Systems (ADAS) (Timofte et al., 2011). Consequently, benchmarks related to TSDR have fewer categories of traffic sign, focused only

Peer review under responsibility of Tongji University and Tongji University Press.
* Corresponding author.
E-mail address: rajeshm@am.amrita.edu (R.K. Megalingam).

https://doi.org/10.1016/j.ijtst.2022.06.002
2046-0430/

on recognition of traffic sign (Zaklouta and Stanciulescu, 2012) and (Zhu et al., 2016a,b). It can be very difficult to detect many other categories of traffic signs with higher disparity in appearance, which are not included in the benchmarks.

Several studies on TSDR have been published, primarily for traffic signs in foreign countries outside of India. Yet, only a limited number of benchmark datasets are available for access by the general public, including the Laboratory for Intelligent and Safe Automobiles (LISA) (Mogelmose et al., 2012), Mapping and Assessing the State of Traffic Infrastructure (MASTIF), Croatia (Segvic et al., 2010), German Traffic Sign Detection Benchmark (GTSDB) (Houben et al., 2013), German Traffic Sign Recognition Benchmark (GTSRB) (Stallkamp et al., 2012), and Tsinghua-Tencent 100 K dataset (Zhu et al., 2016a,b). For the detection and recognition of traffic signs, conventional models are based on deep neural networks. Within the class of deep neural networks, Convolution Neural Network (CNN) has achieved landmark advances in image analysis. Deep neural network frameworks like Fast Region-Based Convolutional Neural Networks (F-CNN) and Mask R-CNN, (Aziz et al., 2020) are trained and tested only using the dataset obtained from foreign roads; these models may not show desired results, when directly applied to Indian roads.

This paper proffers the Refined Mask R-CNN (RMR-CNN) optimal model for detection and recognition of traffic signs for Indian roads. Enhanced accuracy (accuracy is the value in percentage of identifying the traffic signs correctly) in the RMR-CNN framework ensues from augmentation of the Mask R-CNN model by the following pre-processing steps: – shape finding, region of interest (ROI), color probability, besides parametric changes, and data manipulation.

The prime objective of this investigation was to identify measures that minimize traffic accidents. We are motivated to decrease the number of accidents happening In India, road accidents do not only result in loss of lives but it also in a large way affects their families. There are number of ways using which we can decrease the number of accidents, but the most effective way is using the Advanced Driver Assistance Systems (ADAS), in which traffic sign recognition is a part. ADAS is the best solution available to counter this problem. Recognition of Indian traffic signs has not been fully explored by humans; the performance of the latest deep learning methods like Mask R-CNN, Fast R-CNN to decode Indian traffic signs have not yet been verified in practice (traffic signs in roads and highways) using the Indian traffic sign dataset. This research exploited a deep learning-based approach with custom dataset for.

- the detection and comprehension of Indian traffic signs,
- appraisal of the accuracy of recognizing the traffic sign.

Unlike decent-sized traffic sign datasets for countries the world over, hardly any reasonable dataset exists for Indian traffic signs. There is a host of research work on traffic signs detection (TSD) and traffic sign recognition (TSR) mostly outside India. Pre-selection of a felicitous method or algorithm for TSDR is intricated by the lack of a standard dataset with an extensive set of comparable traffic sign categories. Many have wide variation in the number of images in the dataset with varying accuracies and time consumption.

An original dataset for Indian traffic signs was created using deep learning techniques. The datasets for the training and testing of the models were developed by capture of images and their segregation into different categories based on the traffic sign instances present in the image. Only manual traffic signs were used during the analysis of performance of RMR-CNN model, electric signs are not part of the dataset. This approach to the formulation of a dataset for Indian traffic signs has been the first of its kind. The prime contribution of this investigation is deployment of proffered deep learning RMR-CNN [which incorporates improvements to the conventional Mask R-CNN system] for a large number of Indian traffic signs. Instances of the Indian traffic signs are subjected to scaling, orientation change, change in lighting conditions and inter and intra category variations. As part of the verification of the customized dataset, traffic sign detection was compared against Fast R-CNN, whereas traffic sign recognition was evaluated against Mask R-CNN.

While exploring the state of art, R-CNN is able to achieve significant results. There are different versions of R-CNN, starting from fast R-CNN to faster R-CNN to Mask R-CNN. The models are an improvement of one over the other. Of all those Mask R-CNN is widely used and also provides reliable accuracy. Our model which is RMR-CNN is able to improve the accuracy of the available Mask R-CNN. The next section explores contemporaneous models for traffic sign detection and recognition. Section III elaborates the custom dataset categories of Indian road traffic signals. The subsequent section presents the proposed model of a redefined mask-R-CNN for Traffic Sign Detection. Following these sections how recognition of traffic signs takes place in our proposed model has been elaborated. The final two sections focus on the evaluation and results obtained by the submitted RMR-CNN model.

Related work

This section provides a brief overview of prior studies related to detection of traffic signs in different parts of the world. Luo et al. (2018) proposed a 3-step data-driven system to recognize both symbol-based and text-based signs, using a camera mounted on the car; the 3 stages were – extraction of region of interest (ROI), refinement-classification of ROIs, and post-processing. The main drawback of their proposal was excessive post processing time. Mammeri et al. (2013) addressed the challenges and undesirable factors of the TSDR system, an essential component of ADAS. However, the system put forth by Mammeri's team worked over only a limited frequency range; besides, the recognition of traffic signs with a low-resolution camera, camera vibrations and oscillations posed tricky challenges for moving vehicles.

The authors in Lee and Kim (2018) developed a novel CNN traffic-sign detection system that simultaneously estimates the precise location and boundary of traffic signs. Albeit the accuracy was good, Lee's team was obligated by images of very high resolution. Hu et al. (2016) focused on three classes of objects: traffic signs, cars, and cycles. Their proposal detected all the three classes, by a single learning-based detection framework. In their model, the traffic sign detector needed the least amount of time as they used a smaller number of sub-detectors. When additional features were added for detection of other objects, the runtimes for the detection noticeably increased. The system proposed in (Greenhalgh and Mirmehdi, 2012) detected the sign containing regions in the image as maximum stable extremal regions, which help in the detection of signs perfectly, under various weather conditions. Recognition of signs is based on Support vector machine (SVM) classifiers, which were trained using Histogram of Oriented Gradient (HOG) features. However, the precision and recall measures were only 86% and 80% respectively. Subsequently, Greenhalgh and Mirmehdi (2015) employed a scene structure, to identify search regions within an image that had a high probability for the existence of a traffic sign. With obviated structural information, the false positives have netted huge losses to the precision parameter, and the frame rate dropped from 14 frames/s to 6 frames/s.

An efficient 2-module method for traffic sign detection, comprised of a HOG extraction feature, and a single classifier, was trained by Extreme Learning Machine (ELM) algorithm (Huang et al., 2017). The downsides of this model were the performance dependency on tuning parameter, and more than 50% misclassification of the images. Huang et al. (2020) proposed an automatic recognition algorithm, based on visual inspection, in place of the conventional CNN algorithm. But the average precision achieved was notably less than the proposed RMR-CNN method. Using a combination of SVM, and CNN, an extremely fast traffic sign detection method was described in Yang et al. (2016); nevertheless, the accuracy is lesser than comparable state-of-the-art methods like Mask R-CNN. Chen and Lu (2016) suggested a traffic sign detection technique, using Adaptive Boosting (Adaboost) and Support Vector Recognition (SVR). A TSR method, using high contrast region extraction and extended sparse representation, which uses color enhancement technique and voting of neighboring features, was presented in Liu et al. (2016). Color of other objects, which are in traffic signs, get enhanced, imposing a delay in TSR is the downside of their model.

Temel et al. (2020) created a model that detected traffic signs in all types of weather like rain, or if the camera lens is dirty, using spectral characteristics. But the accuracy was limited to 80%. A new model, SegU-Net, an amalgam of state-of-the-art segmentation network, SegNet, and U-Net, was proposed for TSDR in Kamal et al. (2019). On the German Traffic Sign Detection Benchmark dataset, this model achieved a precision score of 95.29%, which was less than conventional methods like Deep Neural Network constituted of Convolutional layers and Spatial Transformer Networks. Wong et al., came up with MicroNet, a compact neural network architecture for TSR (2018). Advanced neural network architectures like Mask R-CNN and Faster R-CNN are computationally fast. A CNN-based TSDR using YOLO (You Only Look Once) architecture was proposed in Avramović et al. (2020). This paper mainly aims at increasing the speed and accuracy of detection using high-definition images by focusing on different regions of interest in images. But the approach used in the above paper can lead to selection of regions in an image which may not encompass any traffic signs. Guo et al. (2020) elucidated problems affecting detection and recognition of Chinese texts, in traffic signs. Occlusion affecting the accuracy in recognition of Chinese characters was the drawback of this method; some of the characters were unrecognized while others were incorrectly recognized.

The authors in Zhang et al. (2017) used the seam carving technique to detect malicious forgeries. Seam carving is an image technique for content aware image resizing. The authors used the blind forensics approach to detect resized images using the seam carving. In addition, the authors also used Weber Local Descriptor (WLD) and Local Binary Patterns (LBP) for seam carving. This technique can be used to decrease the size of the image. In Zhou and Qiu (2021) the authors proposed an enhanced Single Shot Multi-Box Detector (SSD) method. In a conventional SSD method, the feature maps are used to individually predict the object and there is also a lack of interaction between high-level feature and low-level feature maps. In their enhanced SSD they have made the low-level and high-level feature maps interactive, they also used parallel structure in their detection process which maintains the detection accuracy. The authors of Zhang et al. (2021) came up with a methodology to improve the accuracy in object tracking in situations of deformation of images and illumination. The authors used the Resnet features and cascaded correlation filters to improve accuracy and precision. The authors of Zhang et al. (2020) proposed a cascaded correlation filter generated by handcraft, high-level and middle-level Resnet features for better efficiency. In this paper the authors pointed out a problem in using seam carving when the scaling ratio is low. Seam carving is used for resizing images which is used to recognise the image authenticity. An image forensic approach was proposed by the author which is based on the coo-concurrence of adjacent local binary patterns (LBPs) to display better texture information. The authors also used SVM after training for feature classification to check whether an image is seam-carved or not. The authors of Lopez-Montiel et al. (2021) pointed that though many works are focused in developing complex algorithms using deep learning for traffic sign detection, there is no proven reliable methodology which helps to choose appropriate hardware and algorithms. The authors came up with a methodology and an evaluation method to measure the performances of GPU and TPU. To evaluate this methodology they have created some combination of deep learning models such as ResNet50 v1 and MobileNet v1 in combination of Single Shot Multibox Detector (SSD) algorithm and also the Feature Pyramid Network(FPN) for the traffic sign detection in standard LISA dataset. Using their methodology they showed that TPU is 16.3 times faster than GPU and also gets better results in terms of precision.

Indian road traffic signal categories

Road traffic signals are of crucial significance for the safety of vehicles, drivers, passengers, and pedestrians, as well as minimal property damage. Road signs in India are categorized into three categories:

a) Mandatory or Regulatory signs
b) Cautionary, Precautionary or Warning signs
c) Informatory signs

Mandatory signs are those for which the users have no other option and compelled to follow the road sign. Failure to observe these signs is considered as illegal, a chargeable offence. Cautionary or Warning signs alert vehicle users of any slowdowns, roadwork, dangers, or accidents ahead. Informatory signs are very useful to the road users, informing them if they are traversing in the right direction, travelling on the right path towards an intended destination, etc. They also furnish information on roadside services for travelers – food/restaurants, lodging, rest areas, lavatories, gas/diesel stations, etc.

The customized dataset crafted for this paper's research has an aggregate of 6480 images, with 7056 instances of traffic signs, arranged in 100 categories based on the traffic sign instances present in the images. Of the total 100 categories, 13 traffic sign categories were obtained from standard datasets like Deutsche Forschungs Gemeinschaft (DFG), where the traffic signs are identical to the Indian traffic signs. 4544 images, from the total of 6480, were obtained from the real world, which are categorized into 87 traffic sign categories; the residual 1936 images were retrieved from the Jharkhand (state in northern India) Police Department, India Mart, and slide share websites. These 3 are public websites; only images unrestrained by copyright were used. Each category has at least 32 images containing traffic signs. Nearly 70.12% of the total images are of high resolution (4128 × 2322 px, 774 × 1032 px, 960 × 1280 px), with the balance of 29.8% images being of low resolution (225 × 225 px, 200 × 200 px). There are 200 images where 2 instances of traffic sign are found, 40 images where 3 traffic sign instances are found, and 32 images where 4 traffic sign instances are found. About 97% of the traffic sign instances are clearly visible, with only 3% feeble visibility. The sample images of the custom dataset are shown in Figs. 1–3.

Images in the customized dataset are divided into 2 categories – one with signs of over 30 pixels resolution, and the other with signs over 50 pixels resolution. There are few instances where the bounding box is less than 30 pixels, which occurred in images with multiple instances of the same traffic sign; these instances are ignored during training as well as testing phases. A substantial majority of all the traffic sign instances are over 30 pixels. The reason for ignoring instances with less than 30 pixels is due to the down sampling in faster R-CNN and Masked R-CNN, where the 32 × 32 pixels are represented using 1 × 1 pixel in the feature map. The standard test split ratio of 80:20 was used in this RMR-CNN model, which ensured that at least 6 images of each category are used in the custom dataset, with 32 as the minimum number of images in each category. The partitions of the dataset into training and test datasets were done in a random fashion. Doing so yielded 5664 images for the training dataset and 1412 images for the test dataset.

Redefined Mask-RCNN for traffic sign detection

This section includes a discussion of the proposed RMR-CNN system, for detection of traffic signs, with several refinements. At the outset, the Mask R-CNN algorithm used for traffic sign detection is presented in brief; next, refinements in the parametrical values to adapt the Mask R-CNN to our requirements are shown, followed by the improvements in architecture and data augmentation of Mask R-CNN. The refined Mask R-CNN model is shown in Fig. 4.

Mask R-CNN

CNNs, from the class of deep neural networks, have introduced revolutionary concepts in image analysis. Deep neural network frameworks like Fast Region-Based CNNs (F-CNN) and Mask R-CNN are trained and tested using only the dataset obtained from foreign roads; these models may not show realistic results when directly applied to Indian roads. Fast R-CNN is an improvement over R-CNN. The main difference is that F-CNN uses a region proposal network whereas Mask R-CNN uses selective search. Mask R-CNN is an improvement over Faster R-CNN. We generally get a bounding box in Faster R-CNN, which has pixels of both the object and background in the image. Mask R-CNN output gives a binary mask that can be used to determine whether the pixel belongs to a particular object or not. The high accuracy in the proposed RMR-CNN is obtained by making several changes in the present architecture of various CNNs and by data augmentation.

Pre-processing for better accuracy

Three pre-processing steps are carried out prior to application of the Mask R-CNN algorithm, for sign detection and recognition: shape detection, ROI, and color probability.

Shape Detection: The first preprocessing step is intended to find the shapes in an image from live video capture. Once the shape is found we get the images both in color and grayscale. Initially, the camera's image data is converted from color to grayscale. Next, a counter detection technique is applied from OpenCV to detect the counter values. With the help of

Fig. 1. First collage of Custom dataset. 54 images, many of them close up images, of various Indian traffic signs in various positions, orientations and light settings are shown.

obtained counter values, the area is calculated. Based on the area of the counters, the shape of the traffic sign is detected with the threshold parameters defined by the user. These images are then sent to the ROI module.

Fig. 5(a) and (b) are sample images used for detection of ROI. Fig. 6 shows samples of Indian traffic signs from our custom dataset and their corresponding annotation masks, concurrent with the shape detection technique. The annotation mask is achieved by first applying the shape detection method to the original images, and then cropping images based on the coordinates of the traffic sign part of the image obtained during the shape detection method.

Region Of Interest (ROI): Region of interest plays an important role in the detection of the exact location of the traffic sign. Most of the traffic signs in India are presented within the shapes of triangles, rectangles, and circles; hence the suggested RMR-CNN model is capable of finding these shapes as ROI in the image. An image cannot be tested properly by the recognition model if it has any unwanted areas. This may also lead to incorrect predictions, or it may need more time to predict the larger area. Three algorithms are employed to find ROI in an image. If the ROI is circular, the Hough Circle package is used; for triangle and square detection Counter and Edge detection algorithms are used. A separate code, deployed to find ROI for each shape, marks the region of the circle, rectangle, or triangle shapes with a color. The image is cropped by application of a threshold of 5%, to retrieve the image from the ROI without any loss. The cropped images are converted to color images, as a precursor to the next step, color probability [subsection 'D']. Color probability is used to find the percentage of RGB pixel

Fig. 2. Second collage of Custom dataset. 49 images of various Indian traffic signs in various positions, orientations, light settings and at various distances are shown in this collage.

values in the image. Fig. 7(a), 7(b) and 7(c) show the ROI regions, from a sample of the images shown in Fig. 5. Fig. 7(a) corresponds to the ROI of sample image 5(a), grayscale image 7(b) corresponds ROI of 5(a) and image 7(c) corresponds to the ROI of 5(b).

Color Probability: In this step, the Red Green Blue (RGB) value of each pixel of the sample image is calculated and stored in a new dataset. Different lighting conditions such as the sun rays or sunlight directly falling on an instance of a traffic sign, sun light falling on the backside of the traffic sign, and during dark light conditions, are tested to compute the range of RGB pixel values. Based on the obtained dataset, the number of pixels that are in red, black, and white are found using the real-time calculation of range from the image dataset. For each pixel value of image, the counter is increased for red, black, and white colors, based on the range in which it falls. We had calculated the range of RGB pixel counts from the dataset while the image is placed towards the sunlight and placed away the sun light. After the color probability the image will be sent to model. For the image to be sent to the model a threshold value was set from the training dataset.

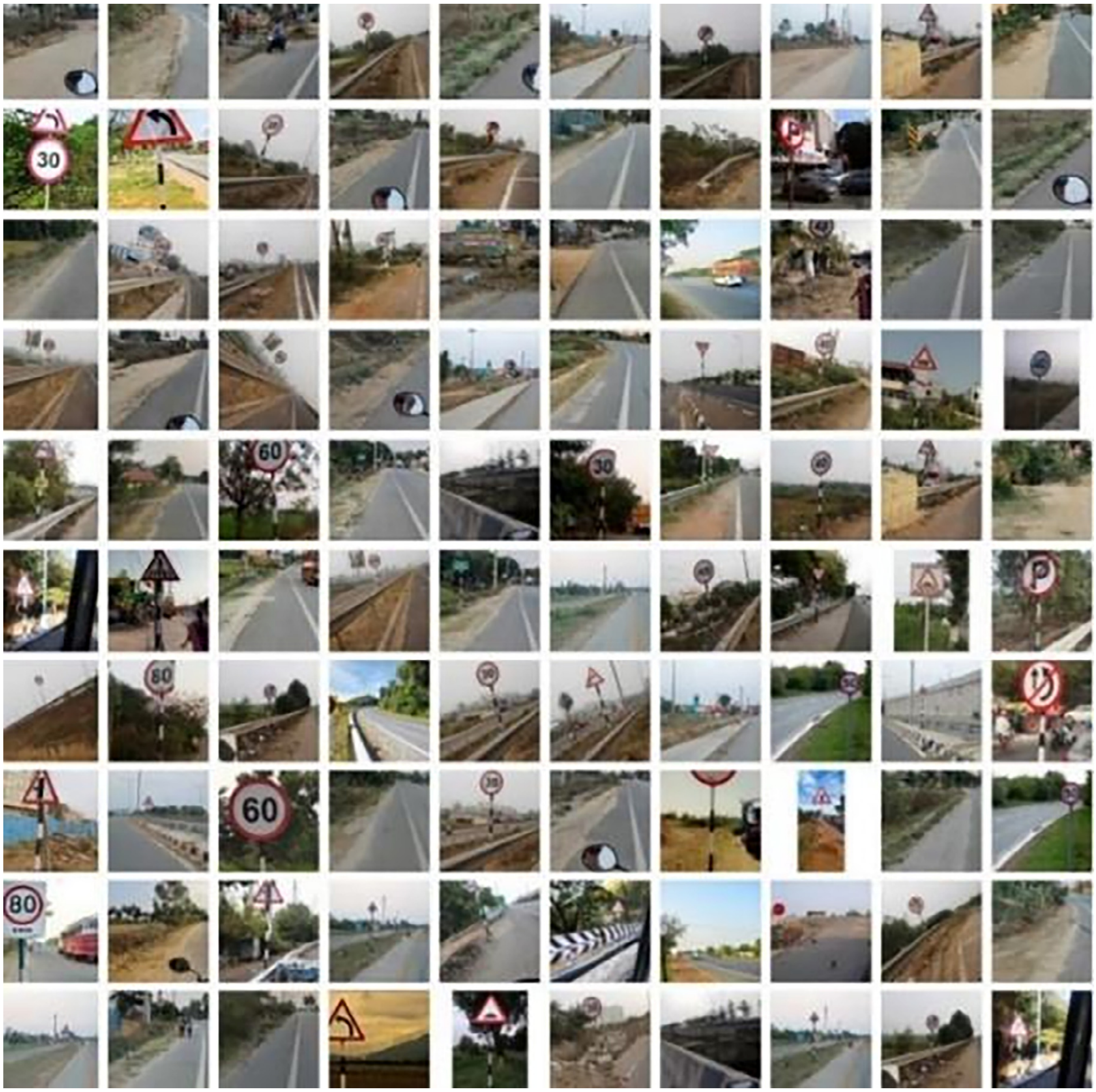

Fig. 3. Third collage of Custom dataset. 100 images of various Indian traffic signs in various positions, orientations, light settings and at various distances are shown in this collage.

Fig. 8(a)	
Color	RGB Values
Red	(218,88,96)
White	(240,232,221)
Black	(67,54,45).

Fig. 8(b)	
Color	RGB Values
Red	(212,32,44)
White	(218,218,216)
Black	(45,45,61)

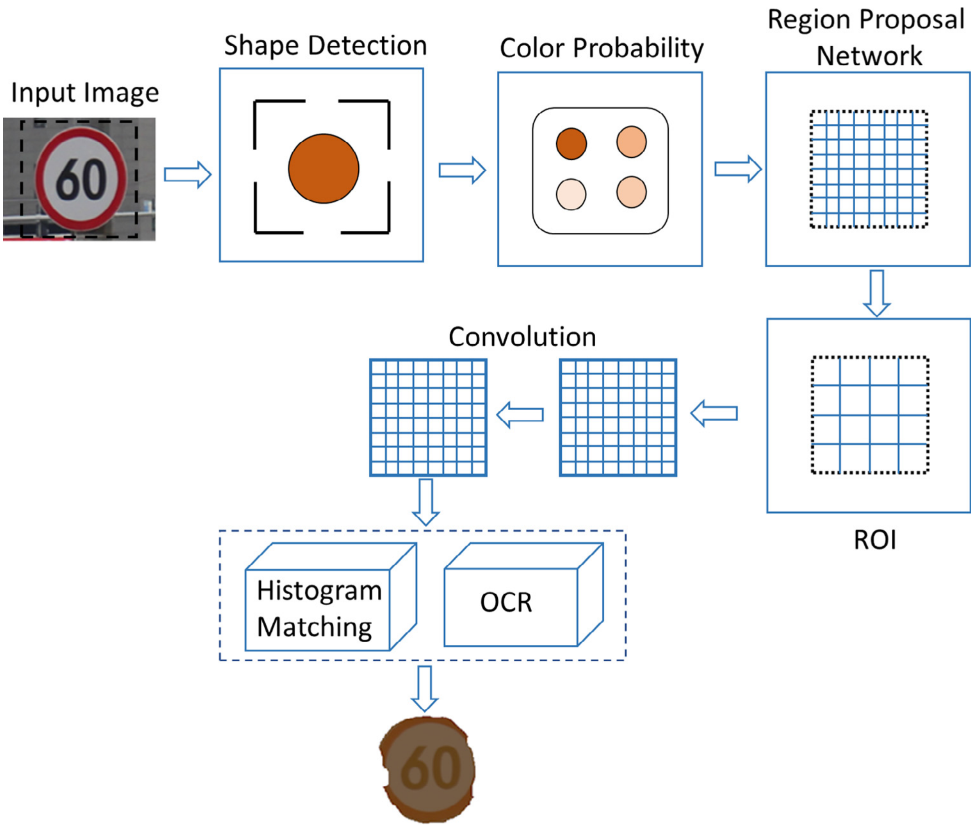

Fig. 4. Refined Mask R-CNN model. Mask R-CNN with shape detection, color probability based on lighting conditions, histogram matching, and optical character recognition (OCR) is proposed for TSD and TSR.

Fig. 5. Sample images used for finding ROI.

By considering more images from dataset, the range for each color was calculated and implemented in color probability.

Fig. 8 show the images of signs where 8(a) is partially exposed to sunlight and the 8(b) is completely exposed to the sun. For the output to be more efficient we calculated the RGB ranges for red, black, white colors, which are explained in next line. On a scale of 256, the RGB values are as noted hereunder. The graphs shown in Fig. 9(a) represents the pixel count of the image partially exposed to sunlight [i.e., Fig. 8(a)] and Fig. 9(b) represents the pixel count of the image completely exposed to sunlight [i.e., Fig. 8(b)]; the pixel counts were obtained by considering the RGB ranges for red, black, white colors (shown in the paragraph before). From the plots 9(a) and 9(b) of RGB pixel count, we calculate the color percentage as follows:

$$\text{color percentage} = \text{count of } (\text{color} - \text{pixel})/(\text{total number of pixels}) \quad (1)$$

Using this approach for this paper's investigation, the threshold values for color percentage were set to 12% for red, 15% for white and 9% for black.

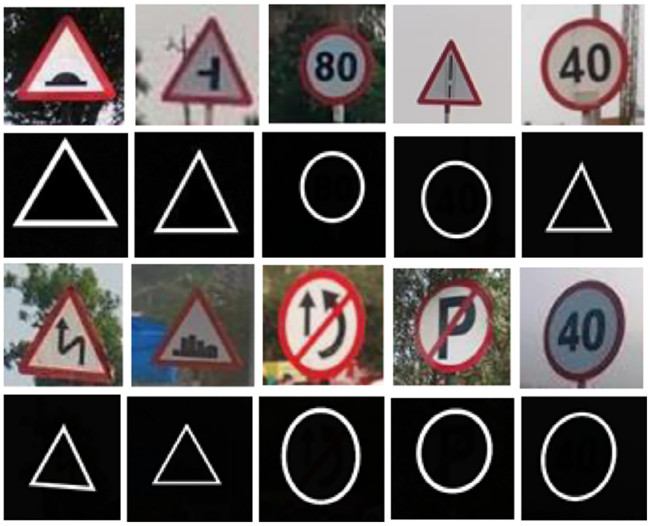

Fig. 6. Sample traffic signs and their corresponding annotation masks.

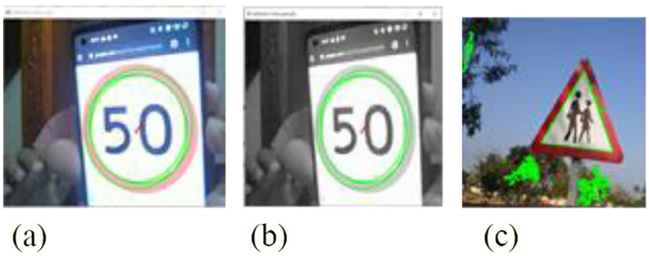

Fig. 7. ROI is marked in green color in all the three images (a), (b) and (c) indicating the outline of traffic sign.

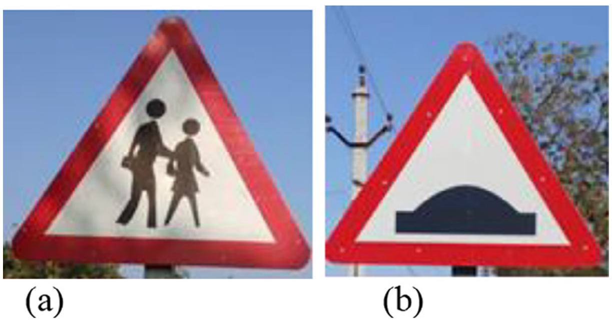

Fig. 8. Two sample images with varying brightness level. (a) Partially exposed to light. (b) Fully exposed to light.

Parametric changes

Even distribution of ROI: Careful selection of the training samples can improve the performance in Mask R-CNN, for Region Proposal Network (RPN). Normally, in Mask R-CNN, the ROIs are selected at random, for the foreground and background. As it is done at random, there is an unequal distribution of ROIs for large objects and small objects present in the image. This unequal distribution is due to the fact that large number of ROIs needed to cover larger objects and small number of ROIs needed to cover smaller object. This increases the accuracy of detection of larger objects but decreases the accuracy for smaller objects. The disparity can be avoided by an even distribution of the training samples for RPN. This is achieved by using the same number of ROIs for each object in the image.

Changing weights: Due to missing region proposal for some cases, the Mask R-CNN cannot achieve 100% recall. During the learning process, the weights of background training regions and foreground training regions are changed. During the

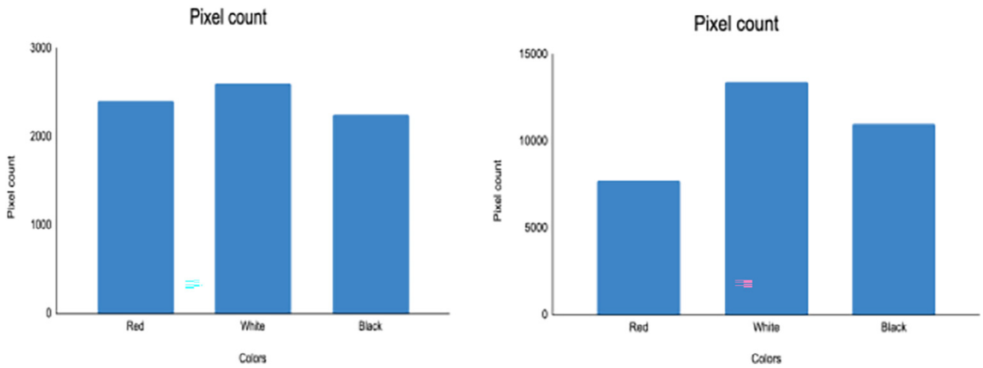

Fig. 9. Results of pixel count of colors. *X*-axis has colors and *Y*-axis has the corresponding pixel count value. (a) Pixel count for 8(a). (b) Pixel count for 8(b).

training, more regions are selected from the background than the foreground, as traffic signs are small in image. Therefore, the training focuses more on learning parameters in the background rather than the foreground region without any weights. This can be avoided using smaller weights for the background regions and larger weights for the foreground regions. The weights are changed for both the RPN and the classification as well. The weights are 0.05 for RPN and 0.2 for the classification network.

Data augmentation

The predominant factor for achievement of high accuracy is the size of the training dataset. As deep learning models have millions of learnable parameters, it is imperative to have a large dataset which helps in precise learning of those parameters. It is impractical to obtain tens of thousands of desired images in real-time. Data Augmentation provides an avenue to increase the size of a dataset with changes in some parameters. A total of 8 data augmentation methods was used which are noted as follows:

Image Rotation: the image is rotated either clockwise or anticlockwise. The resultant empty pixels are assigned pixel values of the adjacent pixel.

Horizontal Flipping: The image is flipped on its *y*-axis.

Brightness Change: The brightness of the image is modified, usually decreased, by subtracting a constant from all the pixel values.

Zoom In: The image is zoomed in to an extra 25%, making its appearance as a cropped image. This can also be accompanied by a decrease in brightness, making it more distinct from the original image.

Right Shift: The image is moved right, by the removal of a portion of pixel values at the right end and shifting the image to the right; the ensuing empty pixels on the left side are assigned the adjacent pixel's values.

Left Shift: The image is moved left, by the removal of a portion of pixel values on the left side and shifting the image to the left side; the resultant empty pixels on the right side are allocated the adjacent pixel values.

Top Shifting: The position of the objects in the image is moved up. This is done by the removal of a portion of pixel values on the top and moving the lower pixel values to the top, and the resulting empty pixels at the bottom are filled with the adjacent pixel values, which are on the top side of the empty pixels.

Bottom Shifting: The position of the objects in the image is moved down. This is done by the removal of a portion of pixel values on the bottom and moving the top pixel values to the bottom, and the resulting empty pixels at the top are filled with the adjacent pixels values which are on the lower or bottom side of the empty pixels.

Fig. 10 shows the data augmentation methods applied on sample images from this study's custom dataset, which facilitate injection of more variations into the custom dataset.

Traffic sign detection using Fast R-CNN

Fast R-CNN and Masked R-CNN were applied to the images for estimation of the accuracy, which are then used for comparing the accuracy of the proposed RMR-CNN. The key disadvantage of using R-CNN is that it takes more time for training. R-CNN takes approximately 47 s for testing each image. So, Faster RCNN was introduced to overcome this issue. The architecture of Fast R-CNN (Girshick, 2015) is comprised of a CNN with an ROI pooling layer, instead of a normal pooling layer. The output from the ROI Pooling layer is fed as input to the SoftMax and BB-regression branches of the Fast R-CNN model. The SoftMax branch helps in derivation of the probability values of each ROI belonging to different categories, besides all background categories. The Fast R-CNN model, trained and tested with the custom dataset, yielded an accuracy of 0.93. Fig. 11 shows the epoch vs. loss values, where the loss of the training set was more than the loss of the validation dataset.

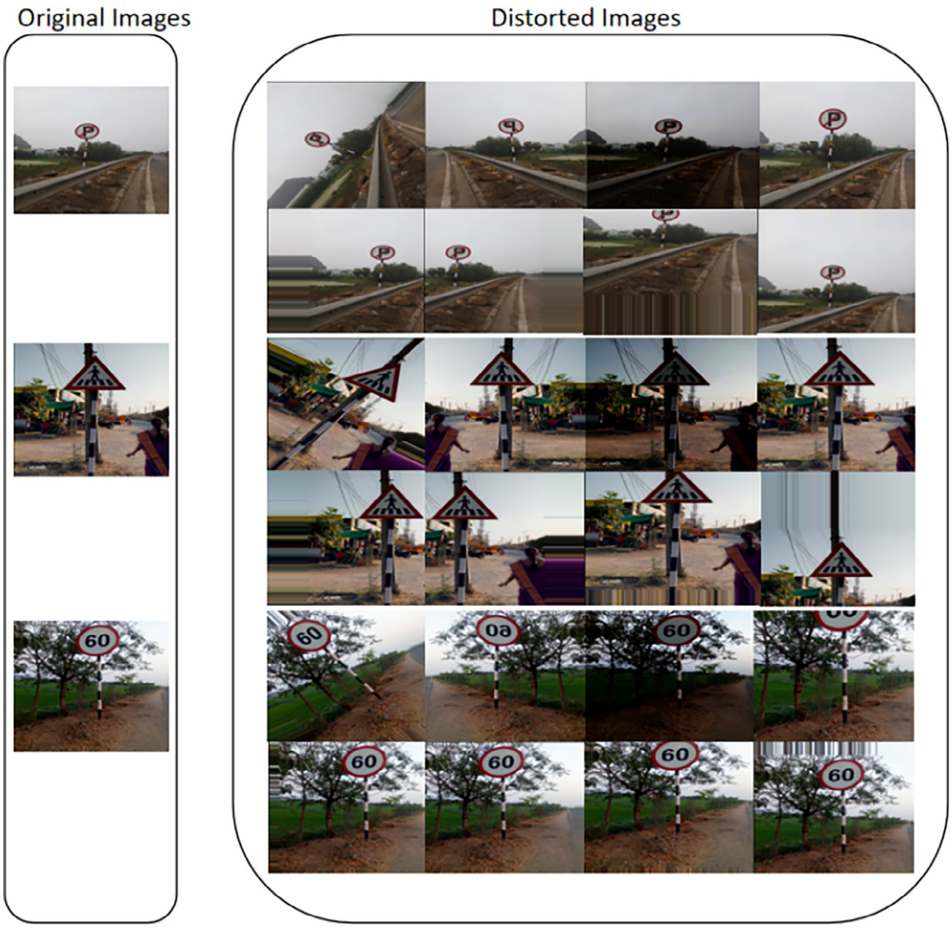

Fig. 10. Samples of data augmentation techniques.

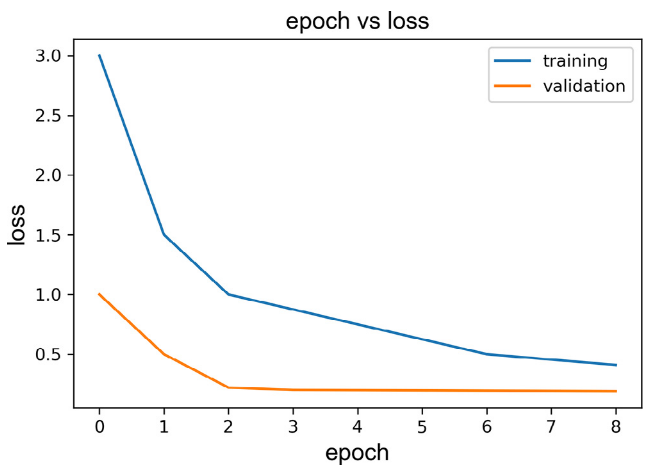

Fig. 11. Epoch vs. loss. In the graph *X*-axis shows the epoch value and *Y*-axis shows the loss value.

Traffic sign detection using Mask R-CNN

Detailed account of Mask R-CNN, as an improved version of Fast R-CNN, was presented in He et al. (2017). Each of the detector models has 2 parts that are connected convolutionally. RPN, the first part of the detector model, (Abbas and

Singh, 2018) takes the input image and proposes or outputs rectangular boxes called bounding boxes. Bounding boxes are likely to find an object. The second part of the detector model is the Fast R-CNN, which is a region-based CNN. Fast R-CNN classifies the objects enclosed within the bounding boxes. Mask R-CNN uses Feature Pyramid Network (Lin et al., 2017) that helps retain the prominent features of low-resolution. Visual Geometry Group (VGG) 16, an underlying network present in Fast R-CNN, is replaced with a residual network in Mask R-CNN (Simonyan and Zisserman, 2015). Mask R-CNN model, trained, and validated on the custom dataset, turned in an accuracy of 0.94. Fig. 12 shows a graph of the epoch vs. loss values, evincing the reduced loss of the validation data compared with the loss of the training dataset.

Traffic sign detection using Refined Mask R-CNN

This section reviews the model proposed in this study, for detection of traffic signs using the Mask R-CNN detector, with several improvements in the distribution of ROI, and assignment of weights to foreground and background in an image, data augmentation and parametric changes. RMR-CNN, proposed in Section IV, turned in an accuracy of 0.974. Fig. 13 presents a graph of epoch vs. loss values for the proposed RMR-CNN, with data augmentation. As the number of iterations increases, loss value decreases, attended by concurrent increases in accuracy of identifying the traffic signs. The enhanced performance is justifiable, for as the iterations increases, the parameters are truthfully learned, resulting in augmented recognition.

Fig. 14 shows the proposed RMR-CNN model's performance in the detection of traffic signs. All the three categories of road signs in India, mandatory or cautionary and informatory signs were covered. Fig. 15 portrays detection of several Indian Road Traffic Signs, sporting intra-category variations.

Traffic sign recognition

This section presents Optical Character Recognition (OCR)-centric techniques for character recognition and Histogram Matching apropos sign recognition.

Optical character recognition

Classic Mask R-CNN can be applied for detection of traffic signs where there are no textual characters involved. Resume OCR was employed for the masked image to derive the circumscribed numbers or textual characters. This subtlety is prescribed, as all the circular or triangular traffic signs call for infallible recognition of the traffic sign (Mainkar et al., 2020). The basic model proposed by Mainkar's team was enhanced [via 'easyocr'], to conform to this investigation's requirements of short latency and unerring accuracy.

Histogram matching

OCR's downside is its ability to only recognize road traffic signs that enclose textual characters. Thus, OCR is unable to detect traffic signs devoid of any encompassed characters [i.e., containing only visual signs]. The Histogram matching method can be marshalled to handle such traffic signs. In this model the threshold value used is 0.82. The threshold value is the percentage of matching, when the derived image is compared to the existing traffic sign instances. The threshold value is used to find the correct category of traffic sign. Whenever the matching rate is above the threshold, the image is given the

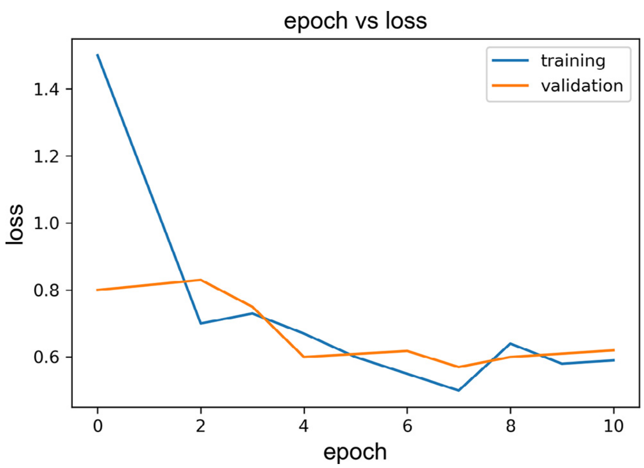

Fig. 12. Epoch vs. loss values for Mask R-CNN. In the graph *X*-axis shows the epoch value and *Y*-axis shows the loss value.

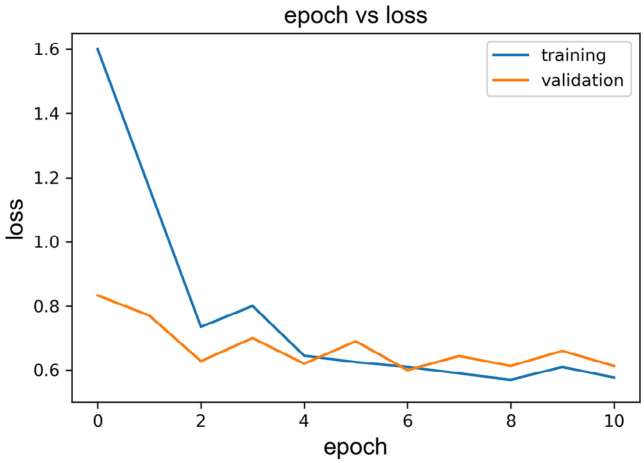

Fig. 13. Epoch vs loss value for Mask R-CNN with model changes and data augmentation.

Fig. 14. Examples of Indian Traffic Signs with various signs.

name of the traffic signal to which it is matched. In the RMR-CNN model, we first checked whether the masked image has any characters. If yes, then the characters and the traffic sign are recognized accordingly. If the characters are unavailable, then we used histogram matching and recognized the corresponding traffic sign. Fig. 16 shows the cropped images, and their corresponding masked images, which are used in OCR and Histogram matching.

Evaluation

The hyperparameters that are set in our model are the number epochs, batch size, minimum and maximum dimension of the image in the dataset. We have not used any optimisation techniques but used physical method where we try different values and compare the accuracies of predictions on the test dataset. The number of epochs is set to 20 and the batch size is set to 100 images per batch. The minimum dimension of the image required is 32 $\times$ 32 pixels and the maximum dimension

Fig. 15. Indian Road Traffic Sign categories, sorted by the proposed RMR-CNN method.

of the image is 2048 × 1080 pixels. Large batch size will speed up the computation using the parallelism of GPUs but very large batch size will result in poor generalization. Using an epoch value which is larger than required value will result in overfitting.

Table 1 shows the precision and recall results for detection of Indian Traffic Sign, using the proffered RMR-CNN concept. The RMR-CNN model achieved an impressive precision of 99.6% and recall at 96.75%. The precision score for recognition is inevitably expected to be lesser than the detection accuracy as the former needs additional steps for recognition that work up to a decrease in the precision value. As part of the assessment of the proposed model, the Precision and Recall values of the Fast R-CNN and Masked R-CNN methods were employed for comparison with the RMR-CNN model are presented in Table 2.

Fig. 16. Cropped image and masked image using OCR and histogram matching.

Table 1
Results of Indian Traffic Sign detection.

Traffic sign	RMR-CNN	
	Precision (%)	Recall (%)
Speed 30	100	100
Speed 40	100	98.4
Speed 60	99.7	97.5
Speed 70	100	100
Speed 80	99	91.2
No horn	100	90.5
Warning	98.9	98.1
Average	99.6	96.75

Table 2
Precision and Recall % comparison for Indian Traffic Sign detection.

Traffic sign	Fast R-CNN		Mask RCNN	RMR-CNN		
	Prec (%)	Rec (%)	Prec (%)	Rec (%)	Prec (%)	Rec (%)
Sign 40	92.1	92.5	98	96.5	98.7	97.6
Sign 60	91.3	97.1	94.5	97.3	99.5	98.1
Sign 70	92.4	92.4	81.4	99.1	88.7	98.5
Go left	99.1	94.2	99.5	92.4	98.6	93.6
Go right	98.9	95.1	94.6	95.5	96.3	95.1
warning	81.2	91.5	96.5	96.5	98.7	98.5
Speed 80	96.4	95.2	96.8	96.8	99.1	95.9
Average	93.05	94	94.4	96.7	97.08	96.75

Note: Prec – Precision, Rec – Recall.

Table 2 summarizes results for the 3 types of models. Precision and Recall of Fast R-CNN were 93.05% and 94%, with a miss rate of 6% and false positive rate of 6.95%. When trained with Mask R-CNN, Precision improved to 94.4% and Recall increased to 96.7%, while the false positive rate decreased from 6.95% to 5.6%, and miss rate decreased from 6% in Fast R-CNN to 3.3% in Mask R-CNN. The precision of the bounding boxes was improved by using data augmentation in RMR-CNN. An accuracy of 97.08% Precision and Recall of 96.75% were achieved in conjunction with the RMR-CNN model. The false positive rate of 2.92% is even less than that of both FCN and Mask R-CNN. Table 3 presents a summary of the precision, recall and F-measure of all the 3 models.

Fig. 17 shows that a precision of 97.08% was achieved for the proffered model with a miss rate of only 3.25% and false positive rate of only 2.92%. This was an expected outcome from the proposed improvements. We can see the enhanced performance of the submitted model after adaptations. When compared with fast RCNN and Mask R-CNN without adaptations, which have error rates of 6% and 3.3% respectively, the model put forth has an error rate of 3.25%. The error rate is nearly half of FCNN's error rate, and the accuracy value is better for RMR-CNN compared to Fast RCNN and Mask R-CNN state of art models. The false positive rates of our model are no more than half, when compared to the standard FCNN and Mask R-CNN model.

Table 3
Precision, recall, and F measure for all three models.

Average	Fast RCNN (%)	Mask RCNN (%)	RMR-CNN (%)
Precision	93	94.4	97.08
Recall	94	96.7	96.75
F-measure	93.4	95.53	96.87

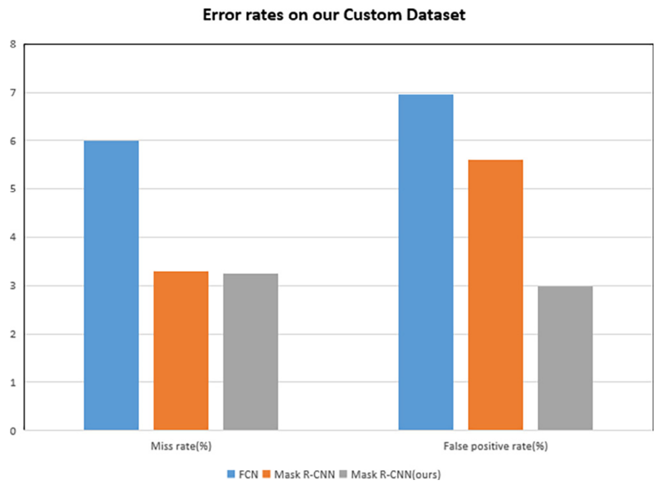

Fig. 17. Miss rates and false positive rates on Indian Traffic Sign dataset.

Results and conclusion

This paper proposed a viable deep learning strategy for the detection and recognition of Indian road traffic signs which revealed good performance under various conditions like light variation, orientation variation and scale variation. The paper introduces RMR-CNN which is a refined version of Mask R-CNN and includes improvements in architecture, data augmentation and refinements in parametrical values. An innovative customized dataset was obtained in real time for the training and validation of the submitted RMR-CNN model. Besides data augmentation, several adaptations to the conventional CNN model served as testimonials for the RMR-CNN tactic's accurate, efficient, fast detection, and learning capability of a decent amount of Indian traffic signs. The observed performance improvements were validated by notable reductions in miss rate and false positive rate. The proffered RMR-CNN model outperformed both the Fast R-CNN and Mask R-CNN models in terms of precision, recall and F-measure.

The developed CNN variant approach was successfully validated on a diverse set of traffic sign categories. The error rate of nearly 3% is mainly due to the similarity with other traffic signs, occlusion, and wide viewing angle. Error rate can be reduced by using multiple instances of the same traffic sign with a stereo camera. The authors submit that the RMR-CNN strategy should be pursued for further refinements in detection and recognition of dirty unclean traffic signs as mentioned in Majid and Heaslip (2016), as well as enhanced performance in terms of decreasing the miss rate and false positive rate.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements and Declarations

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank HuT Labs and, Electronics and Communication Engineering Department at Amrita Vishwa Vidyapeetham, Amritapuri campus for their encouragement and guidance. The authors declare that the contents of this article has not been published previously. All the authors have contributed to the work described, read and approved the contents for publication in this journal.

References

Abbas, S.M., Singh, D.S.N., 2018. Region-based object detection and classification using faster R-CNN. In: 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT), pp. 1–6. https://doi.org/10.1109/CIACT.2018.8480413.
Avramović, A., Sluga, D., Tabernik, D., Skočaj, D., Stojnić, V., Ilc, N., 2020. Neural-network-based traffic sign detection and recognition in high-definition images using region focusing and parallelization. IEEE Access 8, 189855–189868. https://doi.org/10.1109/ACCESS.2020.3031191.
Aziz, L., Haji Salam, M.S.B., Sheikh, U.U., Ayub, S., 2020. Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review. IEEE Access 8, 170461–170495. https://doi.org/10.1109/ACCESS.2020.3021508.
Chen, T., Lu, S., 2016. Accurate and efficient traffic sign detection using discriminative AdaBoost and support vector regression. IEEE Trans. Veh. Technol. 65 (6), 4006–4015. https://doi.org/10.1109/TVT.2015.2500275.
Girshick, R., (2015). Fast R-CNN. IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.
Greenhalgh, J., Mirmehdi, M., 2012. Real-time detection and recognition of road traffic signs. IEEE Trans. Intell. Transp. Syst. 13 (4), 1498–1506. https://doi.org/10.1109/TITS.2012.2208909.
Greenhalgh, J., Mirmehdi, M., 2015. Recognizing text-based traffic signs. IEEE Trans. Intell. Transp. Syst. 16 (3), 1360–1369. https://doi.org/10.1109/TITS.2014.2363167.
Guo, J., You, R., Huang, L., 2020. Mixed vertical-and-horizontal-text traffic sign detection and recognition for street-level scene. IEEE Access 8, 69413–69425. https://doi.org/10.1109/ACCESS.2020.2986500.
He, K., Gkioxari, G., Dollár, P., Girshick, R., (2017). Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.
Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M., Igel, C., 2013. Detection of traffic signs in real-world images: The German traffic sign detection benchmark. Proc. IJCNN, 1–8.
Hu, Q., Paisitkriangkrai, S., Shen, C., van den Hengel, A., Porikli, F., 2016. Fast detection of multiple objects in traffic scenes with a common detection framework. IEEE Trans. Intell. Transp. Syst. 17 (4), 1002–1014. https://doi.org/10.1109/TITS.2015.2496795.
Huang, Z., Yu, Y., Gu, J., Liu, H., 2017. An efficient method for traffic sign recognition based on extreme learning machine. IEEE Trans. Cybern. 47 (4), 920–933.
Kamada, H., Naoi, S., Gotoh, T. (1990). A compact navigation system using image processing and fuzzy control. In: IEEE Proceedings on Southeastcon, 1990, pp. 337-342 vol.1, doi: 10.1109/SECON.1990.117829.
Kamal, U., Tonmoy, T.I., Das, S., Hasan, M.K., 2019. Automatic traffic sign detection and recognition using SegU-Net and a modified Tversky loss function with L1-constraint. IEEE Trans. Intell. Transp. Syst. 1–13. https://doi.org/10.1109/tits.2019.2911727.
Lee, H.S., Kim, K., 2018. Simultaneous traffic sign detection and boundary estimation using convolutional neural network. IEEE Trans. Intell. Transp. Syst. 19 (5), 1652–1663. https://doi.org/10.1109/TITS.2018.2801560.
Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. Proc. Comput. Vis. Pattern Recognit., 936–944
Liu, C., Chang, F., Chen, Z., Liu, D., 2016. Fast traffic sign recognition via high-contrast region extraction and extended sparse representation. IEEE Trans. Intell. Transp. Syst. 17 (1), 79–92. https://doi.org/10.1109/TITS.2015.2459594.
Lopez-Montiel, M., Orozco-Rosas, U., Sanchez-Adame, M., Picos, K., Ross, O.H.M., 2021. Evaluation method of deep learning-based embedded systems for traffic sign detection. IEEE Access 9, 101217–101238.
Luo, H., Yang, Y., Tong, B., Wu, F., Fan, B., 2018. Traffic sign recognition using a multi-task convolutional neural network. IEEE Trans. Intell. Transp. Syst. 19 (4), 1100–1111. https://doi.org/10.1109/TITS.2017.2714691.
Mainkar, V.V., Katkar, J.A., Upade, A.B., Pednekar, P.R., 2020. Handwritten character recognition to obtain editable text. In: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 599–602. https://doi.org/10.1109/ICESC48915.2020.9155786.
Majid, K., Heaslip, K., 2016. Analysis of factors temporarily impacting traffic sign readability. Int. J. Transp. Sci. Technol. 5 (2), 60–67. https://doi.org/10.1016/j.ijtst.2016.09.003. ISSN2046-0430.
Mammeri, A., Boukerche, A., Almulla, M., 2013. Design of traffic sign detection, recognition, and transmission systems for smart vehicles. IEEE Wireless Commun. 20 (6), 36–43.
Mogelmose, A., Trivedi, M.M., Moeslund, T.B., 2012. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. IEEE Trans. Intell. Transp. Syst. 13 (4), 1484–1497.
Segvic, S., Brkic, K., Kalafatic, Z., Stanisavljevic, V., Sevrovic, M., Budimir, D., Dadic, I. (2010). A computer vision assisted geoinformation inventory for traffic infrastructure. In: Proc. 13th Int. IEEE Conf. Intell. Transp. Syst. (ITSC), pp. 66–73.
Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. Proc. Int. Conf. Learn. Represent., 1–14
Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C., 2012. Man vs. Computer: Benchmarking machine learning algorithms for traffic sign recognition. Neural Netw. 32, 323–332.
Temel, D., Chen, M.H., AlRegib, G., 2020. Traffic sign detection under challenging conditions: a deeper look into performance variations and spectral characteristics. IEEE Trans. Intell. Transp. Syst. 21 (9), 3663–3673. https://doi.org/10.1109/TITS.2019.2931429.
Timofte, R., Prisacariu, V., Van Gool, L., Reid, I., 2011. Combining traffic sign detection with 3D tracking towards better driver assistance. In: Emerging Topics in Computer Vision and Its Applications. World Scientific, Singapore, pp. 425–446. https://doi.org/10.1142/8103.
Wong, A., Shafiee, M.J., St. Jules, M., 2018. MicronNet: a highly compact deep convolutional neural network architecture for real-time embedded traffic sign classification. IEEE Access 6, 59803–59810.
Yang, Y., Luo, H., Xu, H., Wu, F., 2016. Towards real-time traffic sign detection and classification. IEEE Trans. Intell. Transp. Syst. 17 (7), 2022–2031. https://doi.org/10.1109/TITS.2015.2482461.
Zaklouta, F., Stanciulescu, B., 2012. Real-time traffic-sign recognition using tree classifiers. IEEE Trans. Intell. Transp. Syst. 13 (4), 1507–1514.
Zhang, D., Chen, X., Li, F., Sangaiah, A.K., Ding, X., Gao, H., 2020. Seam-carved image tampering detection based on the cooccurrence of adjacent LBPs. Secur. Commun. Netw. 2020, 12. https://doi.org/10.1155/2020/8830310. Article ID 8830310.
Zhang, D., Li, Q., Yang, G., Li, L., Sun, X., 2017. Detection of image seam carving by using weber local descriptor and local binary patterns. J. Inf. Secur. Appl. 36, 135–144.
Zhang, J., Sun, J., Wang, J., Yue, X.-G., 2021. Visual object tracking based on residual network and cascaded correlation filters. J. Ambient Intell. Human Comput. 12 (8), 8427–8440.
Zhou, S., Qiu, J., 2021. Enhanced SSD with interactive multi-scale attention features for object detection. Multimed. Tools Appl. 80, 11539–11556. https://doi.org/10.1007/s11042-020-10191-2.
Zhu, Y., Zhang, C., Zhou, D., Wang, X., Bai, X., Liu, W., 2016a. Traffic sign detection and recognition using fully convolutional network guided proposals ISSN 0925-2312 Neurocomputing 214, 758–766. https://doi.org/10.1016/j.neucom.2016.07.009.
Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., Hu, S., 2016b. Traffic sign detection and classification in the wild. Proc. CVPR, 2110–2118.

Further Reading

He, S., Chen, L., Zhang, S., Guo, Z., Sun, P., Liu, H., Liu, H., 2021. Automatic recognition of traffic signs based on visual inspection. IEEE Access 9, 43253–43261. https://doi.org/10.1109/ACCESS.2021.3059052.