IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

REVIEW ON TRAFFIC SIGN DETECTION AND RECOGNITION USING DEEP LEARNING UNDER CHALLENGING WEATHER CONDITIONS

[1]Stephen Mascarenhas, [2]Vaibhav Kolpe, [3]Pratik Shinde, [4]Ankita Tapase, [5] Dr. Aparna Pande

[1]Student, Department of Computer Science and Engineering, Nutan College of Engineering and Research, Pune, Maharashtra, India
[2]Student, Department of Computer Science and Engineering, Nutan College of Engineering and Research, Pune, Maharashtra, India
[3]Student, Department of Computer Science and Engineering, Nutan College of Engineering and Research, Pune, Maharashtra, India
[4]Student, Department of Computer Science and Engineering, Nutan College of Engineering and Research, Pune, Maharashtra, India
[5]Associate Professor, Department of Computer Science and Engineering, Nutan College of Engineering and Research, Pune, Maharashtra, India

Abstract: Traffic sign detection and recognition is an emerging area of research that is gaining popularity with the release of Deep Learning and Computer Vision in modern automobiles. With changing weather conditions and environmental hazards, it is difficult for modern automobiles to identify and recognize road signs. TSDR problem in different CCs (Haze, Snow, Dirty Lens, Lens Blur and Rain) is studied in this paper, with an emphasis on the resulting performance degradation, & suggest a prior enhancement focused TSDR architecture built on Convolutional Neural Networks (CNN). The 4 modules which give most accurate results in detecting and recognizing traffic signs (Challenge Classifier, Enhancement Block, Sign Localizer and Sign Classifier) are flexible, also modifiable on the basis of actual weather conditions. The CURE-TSD dataset, which consists of traffic recordings shot using various CCs, is used to assess the effectiveness of mentioned methods. On the CURE-TSD Dataset, experimental results show that VGG-16 achieves a total accuracy of 99.98%.

IndexTerms **- Deep learning, Convolutional Neural Network, Challenging Conditions, Traffic Sign Detection, Traffic Sign Recognition, Enhancement Block, Challenge Classifier.**

INTRODUCTION

In this generation, Driver Assistance System and autonomous vehicle technology have been vastly demanded as auto driving mode is used for helping the driver for better performance, accuracy and experience. In these systems, traffic sign detections and recognition is one vastly developing area where deep learning based approach is mostly researched. However, most researchers have done research using [2]-[10] which neither includes challenging conditions occurred while capturing images in real world scenarios so, these datasets are not that effective under different weather conditions. In this study, we discovered a video dataset, CURE-TSD that considers both real and unreal challenging environments for traffic sign detection. This dataset includes different CCs that are often possible in real-life situations with more than 1.7 million images, making it the biggest dataset available, and the author has assessed the impact of different CCs based on the total performance of the algorithm. The top two winners of the 2017 IEEE Video and Image Processing (VIP) Cup offered two benchmarks that demonstrated how the CCs may lower the F2 score by 65%. A substantial database or dataset called ImageNet was developed by academics for computer vision study. According to the WordNet hierarchy, there are 14 million annotated pictures in the ImageNet collection. The dataset has been used in the ImageNet Large Scale Visual Recognition Competition (ILSVRC), a benchmark for the categorization of objects and images, since 2010. Researchers used the ImageNet dataset to pre-train a VGG16 network, and from its intermediate layers they extracted characteristics of the target image traffic sign regions and the reconstructed image areas.

In [1] Sabbir Ahmed et al proposed method consists of following methods:

• Challenge Classifier detects challenges present in the captured traffic image and passes the image to enhancement block.
• Enhancement Block performs the enhancement for respective challenges and if the detected challenge type is “No Challenge” then the image bypasses the enhancement blocks and is passed directly to Sign Detection and Recognition modules.
• In Sign Localizer, the traffic sign will be located and only that part will be sent to sign recognizer.

• In Sign Classifier, the region of interest sent by sign localizer will be taken as an input and the classifier will give the appropriate output.
There are lots of studies [11]-[19] and [22] present for all of these modules, and in these articles, we examine those various approaches for each module in an effort to suggest a more effective strategy.

METHODOLOGY

A. Challenge Classifier

In real world scenario, we frequently face different weather challenges which makes difficult to tackle with such variety of possibilities. And we can simplify such complex problem with just adding this module, by doing further survey we have come to known about the methods (VGG-16, Resnet-18, Alexnet and GoogleNet) which are more precise and effective for classifying images.

I. VGG-16

• VGG-16 [11] is shorthand for the VGG model, also known as VGG-Net. The model is a 16-layer convolution neural network (CNN).
• Karen Simonyan and Andrew Zisserman from Oxford University's Visual Geometry Group Lab proposed VGG 16 in their article "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION [32]" from 2014.
• On the ImageNet dataset, which contains 14 million images divided into 1000 classes, this model gets top-5 test accuracy of 92.7%.
• VGG Architecture: A dimensioned picture serves as the network's input. (224, 224, 3). VGG networks are made up of tiny convolution filters. VGG-16 has 13 convolutional layers and three completely connected layers.
• This model was different from earlier, successful models in a number of respects. First of all, AlexNet used an 11 x 11 receptive field with a 4-pixel stride while it only used a small 33 receptive field with a 1-pixel stride. The purpose of a larger receptive field is achieved by the combination of the 33 filters.
• When using numerous smaller layers as opposed to a single large layer, the decision functions are improved and the network can converge more quickly. This is because there are more non-linear activation layers present.
• Second, the smaller convolutional filter used by VGG lessens the likelihood that the network will over-fit during training exercises.
• The best size for a filter is 3 x 3, as smaller sizes can't catch information from the left, right, and up and down. Therefore, VGG is the smallest model that can be used to comprehend the spatial characteristics of an image. The network is simple to control thanks to consistent 3 x 3 convolutions.

II. ResNet

• A Computer vision application use ResNet [13] (residual networks). It is a CNN framework that can accommodate thousands of layers.
• An architecture known as ResNet was unveiled in 2015 by researchers at Microsoft Research.
• Previous CNN architectures supported only a small number of layers, which affected speed, so in order to solve this, experts created a "vanishing gradient" to add more layers.
• Gradient descent is used in the backpropagation technique used to train neural networks which lowers the loss function and identifies the weights that reduce it. If there are too many layers, performance saturates or degrades with each additional layer, and repeated multiplications will ultimately cause the gradient to disappear.
• By using "skip connections," ResNet offers a novel remedy for the disappearing gradient issue.

III. AlexNet

• AlexNet [12] is an eight-layer CNN model which is pre-trained on more than a million pictures of the ImageNet database.
• With a deviation of over 10.8 percentage points from the second-place contender, the network's top 5 error rate stood at 15.3%.
• AlexNet architecture comprises of 5 convolutional layers, 3 max pooling layers, 2 normalization layers, 2 connected layers and 1 softmax layer.
• An apportionment of over 1000 class labels are generated using 1000-directional softmax at the output of the final fully linked layer.

IV. GoogLeNet

• A convolutional neural network with 22 layers, GoogLeNet [12], was developed. A network that has been trained using ImageNet can be loaded in a pre-trained state.
• In 2014, Google researchers proposed GoogleNet in a research paper "Going Deeper with Convolutions" with the support of many universities.
• In the 2014 ILSVRC competition, this model won compared to AlexNet with substantially compressed error rate.
• In the classification task, it has a top 5 error rate of 6.67%. On the ImageNet test dataset, an aggregation of 6 GoogLeNets achieves a mAP of 43.9%.

Table I: Performance comparison among different challenge classifiers on cure-tsd and imagenet

Year	CNN	Developed by	Parameters	Salient Feature	Top – 5 error rate	Accuracy	Dataset
2014	VGG-16		138M	Fixed-size kernels	7.3%	99.98%	Imagenet and CURE-TSD
2015	ResNet	0.02	60M	Shortcut connections	3.6%	0.510	Imagenet and CURE-TSD
2014	GoogLeNet	0.04	4M	Wider-Parallel	6.67%	0.467	Imagenet
2012	AlexNet	0.11	60M	Deeper	15.3%	0.290	Imagenet

B. Enhancement Block

Enhancement block plays significant role in enhancing the quality of images in each cases (Rain, Snow, Haze, Lens Blur and Dirty Lens) with separate and independent block of each case. -Due to the variety of these obstacles, using a single training module for all of them will reduce performance. With each challenge, five distinct enhancement blocks are trained to prevent this issue. By doing further survey, we have come to known EnhanceNet, VDSR, ESPCN, SRCNN are the best practices.

I. EnhanceNet

• EnhanceNet [14] is a Generative Adversarial Network which focuses on getting realistic textures along with higher perceptual quality rather than just improving on the PSNR (Peak Signal to Noise Ratio) values. It is a Single Image Super Resolution model used convert low resolution image (LR) to a high-resolution image (HR).
• The feed-forward fully convolutional network in EnhanceNet has a generic CNN design and consists of 10 residual blocks, which aid in a model's quicker convergence.
• In order to prevent extraneous artifacts, nearest neighbor up-sampling is utilized in the network's up-sampling section in place of the convolution transpose layers, followed by a convolution layer.
• To eliminate any color changes and to guarantee training stability, a bicubic interpolation of the LR image is applied to output.

II. VDSR

• VDSR [15] (Very Deep Super Resolution) is used to enlarge an image. It contains 20 weight layers, which is far deeper than SRCNN, which only has three levels.
• VDSR is a state-of-the-art approach published in 2016 CVPR.
• VDSR architecture consists of:
    o The network receives the LR picture as input.
    o LR image passes through Conv and ReLU layers and then finally in ConvD layer.
    o The output is finally combined with the LR image to create a HR image.

III. ESPCN

• ESPCN [16] (Efficient Sub-Pixel Convolutional Neural Network) is a network of many convolutional layers that up scales LR image through a sub-pixel convolutional layer.
• ESCPN adds an efficient sub-pixel convolutional layer which reduces the image size hence, a small filter can be pre-owned to extract features.
• ESPCN architecture consist of:
    o The LR image is taken as input with form (B, C, N, N).
    o Consist of 3 convolutional layers where first layer has 64 filters with tanh activation layer and size of kernel is 5 x 5.
    o Second layer contains kernel size as 3 x 3 with tanh activation layer and 32 filters.
    o Third layer has a channel (C x r x r) and size of kernel is 3 x 3.
    o Finally, sub-pixel shuffle function for the output with shape and sigmoid layer to obtain HR picture.

IV. SRCNN

• SRCNN [17] (Super Resolution Convolutional Neural Network) is used for single image super resolution. i.e., LR is enhanced to HR Image.
• SRCNN is a state-of-the art which was introduced in ECCV, 2014. SRCNN network is not so deep consisting of only 3 parts/layers:
    o Patch Extraction and Representation: LR image up-scaled using bicubic interpolation.
    o Non-Linear Mapping: Maps n1-dimensional vector to n2-dimensional vector.
    o Reconstruction: After mapping, image is reconstructed using conv to obtain a high-resolution image.

Table II: PSNR for the module trained with multiple image enhancers at 4x resolution

A = 4 Dataset	SRCNN	ESPCN	VDSR	ENet-E
Set5	30.48	30.90	31.35	31.74
Set14	27.49	27.73	28.01	28.42
BSD100	26.90	-	27.29	27.50
Urban100	24.52	-	25.18	25.66

C. Sign Localizer

In a captured image the traffic sign can be present anywhere, so sign localizer locates the traffic sign by breaking the image into smaller segments. There are different methods to be fit for sign localizer and as per survey we have come to known as follows:

I. SegNet

• A semantic segmentation algorithm is SegNet [18]. This basic segmentation design consists of a pixel wise classification which is equivalent to encoder and decoder network.
• The architecture of the encoder network is similar to the layers in the VGG16 network.
• The decoder network is used to convert the low-resolution feature map to full input resolution feature map for pixel-by-pixel classification.
• The decoder performs non-linear up-sampling specifically using pooling indices that were obtained in the max pooling stage of associated encoder.

II. DeConvNet

• The deconvolution network [19] is made up of deconvolution and unpooling layers, which recognize pixel-by-pixel class labels and forecast segmentation masks.
• The network receives a sub-image that may contain an object as input and we refer to this as an instance(s) of an object.
• The key features of the DeConvNet are listed below
    o Envisage input trigger that stimulate individual characteristics.
    o Return the activation of the projected function to the input pixel.
    o Sensitivity analysis on classifier's output.
    o Observation of evolution of features during training.

III. FCN

• FCN [19] (Fully Convolutional Network) is a model pre-owned for semantic segmentation.
• Only use locally connected layers, such as pooling layer, convolution layer and up-sampling layer.
• Avoiding dense layers requires fewer parameters. This also means that FCN can handle different image sizes as long as all connections are local.
• The down sampling path is used by the network to distillate and transcribe context, and the up sampling path enables localization.
• Skip connections are another technique used by FCNs to retrieve the fine-grained contiguous details lost during the down sampling process.

IV. U-Net

• U-Net [23] is a segmentation method primarily designed for medical image segmentation, which evolved from convolutional neural networks and was first developed and used in 2015 to process images in biomedicine.
• This architecture has been successfully used scrupulously identify small tumors from such pictures.
• The U-Net architecture contains a convolutional layer and two networks, encoder and decoder.

Table III: A comparison of time and hardware resources required for computing in sign localizer

Model	Size (MB)	Backward pass (ms)	Forward pass (ms)	Training GPU memory (MB)	Inference GPU memory (MB)
SegNet	117	488.71	422.50	6803	1052
FCN	539	484.11	317.09	9735	1806
DeconvNet	877	602.15	474.65	9731	1872

D. Sign Classifier

There are large varieties of traffic signs, sign classifier classifies these different signs using various attributes of the sign and as per our survey, we have found out different approaches: color based, shape based and edge based.

I. Color-based approach

• The main component of the color-based technique is the threshold-based segmentation of the traffic sign zone in a specific color such as HSI [4] (Hue-Saturation-Intensity), HCL [5] (Hue-Chroma-Luminance) and others [6].
• However, a significant disadvantage of mentioned color-based techniques is that they are extremely vulnerable to changes in lighting, which can happen often in real-world circumstances [8].

II. Shape-based approach
• Object shape is an important and fundamental visual feature for revealing image content.
• Shape feature extraction is also widely used to obtain the user's region of interest. Shape features can be considered as high level features than colors and textures and should be repeatable. That is, if taken from multiple photos of the same scene, they should be the same.
• In the existing literature, shape-based methods such as Canny-Edge detection [8], Histogram-Oriented Gradients (HOG) [6], Haar-Wavelet features [24], and Fast Fourier Transform (FFT) [10] have been widely used.
• Practical use of these techniques is restricted by traffic sign area occlusion caused by size and scale variations, dis-orientation, and motion artefacts while real time video transmission.

III. Proposed method
• For sign classifier, in [18] Uday kamal et al proposed a VGG-16-like architecture consisting of two convolutional blocks for feature extraction.
• These blocks contain two 3 x 3 convolution layers respectively, a ReLU activation layer and max-pooling layer. Also, a dropout layer is appended after every block with a dropout probability of 0.25 to prevent overfitting the model.
• Finally, two fully successive connected layers are used to propagate the features for sign categorization. [18] Contains the specifics of these networks architecture.

CONCLUSION

In this paper, a deep CNN-based modular and robust architecture is proposed for TSDR under different CCs (Rain, Snow, Haze, Dirty Lens and Lens Blur). For Challenge classifier, VGG-16 architecture-based network successfully detects and classifies the challenges, sends the image to the Enhancement Block where Enhance-Net recovers the important features that are useful for traffic sign detection. Proposed Enhance-Net architecture covers loss functions and has pre-trained individual blocks for each CC. This effectively ensures the enhancement of the sign regions subject to their accurate detection. Seg-UNet is the advised hybrid method which combines SegNet and UNet which flourishingly localizes the traffic sign from the enhanced image. Finally, VSSA-NET used for classifying the detected sign accurately.

ACKNOWLEDGEMENT

The authors would like to thank the publishers and researchers for making their valuable resources available, as well as our instructors for their guidance. We would like to express our heartfelt appreciation to our guide, Dr. Aparna Pande, for her encouraging guidance in carrying out this work. We sincerely thank our respective college, Nutan College of Engineering and Research, Pune, for allowing us to take our work this further.

FUTURE SCOPE

In the future, we hope to investigate other architecture and compression strategies to create a virtually optimal network for each module, addressing all of the CCs in the CURE-TSD dataset in terms of performance and inference speed.

REFERENCES

[1] S. Ahmed, U. Kamal, M. K. Hasan, "DFR-TSD: A Deep Learning Based Framework for Robust Traffic Sign Detection Under Challenging Weather Conditions", IEEE Transactions on Intelligent Transportation Systems, 2021. DOI: 10.1109/TITS.2020.3048878.
[2] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey", IEEE Transactions on Intelligent Transportation Systems, 2012. DOI: 10.1109/TITS.2012.2209421.
[3] A. Gudigar, S. Chokkadi, and U. Raghavendra, "A review on automatic detection and recognition of traffic sign", Multimedia Tools Appl., 2016. DOI: 10.1007/s11042-014-2293-7.
[4] S. Xu, "Robust traffic sign shape recognition using geometric matching", IET Intelligent Transport Systems, 2009. DOI: 10.1049/iet-its:20070058.
[5] J. F. Khan, S. M. A. Bhuiyan, and R. R. Adhami, "Image segmentation and shape analysis for road-sign detection", IEEE Transactions on Intelligent Transportation Systems, 2010. DOI: 10.1109/TITS.2010.2073466.
[6] I. M. Creusen, R. G. J. Wijnhoven, E. Herbschleb, and P. H. N. de With, "Color exploitation in hog-based traffic sign detection", IEEE Transactions on Intelligent Transportation Systems, 2010. DOI: 10.1109/ICIP.2010.5651637.
[7] D. Temel, M.-H. Chen, and G. AlRegib, "Traffic sign detection under challenging conditions: A deeper look into performance variations and spectral characteristics", IEEE Transactions on Intelligent Transportation Systems, 2019. DOI: 10.1109/TITS.2019.2931429.
[8] J. Canny, "A computational approach to edge detection", IEEE Transactions on Intelligent Transportation Systems, 1986. DOI: 10.1109/TPAMI.1986.4767851.
[9] H. Luo, Y. Yang, B. Tong, F. Wu, and B. Fan, "Traffic sign recognition using a multi-task convolutional neural network", IEEE Transactions on Intelligent Transportation Systems, 2017. DOI: 10.1109/TITS.2017.2714691.
[10] Y. Yang, H. Luo, H. Xu, and F. Wu, "Towards real-time traffic sign detection and classification", IEEE Transactions on Intelligent Transportation Systems, 2015. DOI: 10.1109/TITS.2017.2714691.
[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv e-prints, 2014. DOI: 10.48550/arXiv.1409.1556.
[12] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016. DOI: 10.1109/TPAMI.2016.2572683.
[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. DOI: 10.1109/CVPR.2016.90.

[14] Mehdi S. M. Sajjadi, Bernhard Scholkopf, Michael Hirsch, "EnhanceNet: Single Image Super-Resolution through Automated Texture Synthesis", IEEE International Conference on Computer Vision (ICCV), 2017. DOI: 10.1109/ICCV.2017.481.
[15] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super resolution using very deep convolutional networks", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. DOI: 10.1109/CVPR.2016.182.
[16] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. DOI: 10.1109/CVPR.2016.207.
[17] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution", European Conference on Computer Vision (ECCV), 2014. DOI: 10.1007/978-3-319-10593-2_13.
[18] U. Kamal, T. I. Tonmoy, S. Das, and M. K. Hasan, "Automatic traffic sign detection and recognition using SegU-Net and a modified Tversky loss function with L1-constraint", IEEE Transactions on Intelligent Transportation Systems, 2019. DOI: 10.1109/TITS.2019.2911727.
[19] E. Shelhamer, J. Long,T. Darrell, "Fully convolutional networks for semantic segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016. DOI: 10.1109/TPAMI.2016.2572683.
[20] D. Temel, T. Alshawi, M.-H. Chen, and G. AlRegib, "CURE-TSD: Challenging unreal and real environments for traffic sign detection", IEEE Dataport, 2019, DOI: 10.21227/en9z-mq69.
[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", Communications of the ACM, 2017. DOI: 10.1145/3065386
[22] Y. Yuan, Z. Xiong, and Q. Wang, "VSSA-NET: Vertical spatial sequence attention network for traffic sign detection", IEEE Transactions on Image Processing, 2019. DOI: 10.1109/TIP.2019.2896952.
[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", arXiv e-prints, 2015. DOI: 10.48550/arXiv.1505.04597.
[24] P. G. Jiménez, S. M. Bascón, H. G. Moreno, S. L. Arroyo, and F. L. Ferreras, "Traffic sign shape classification and localization based on the normalized FFT of the signature of blobs and 2D homographies", Elsevier B.V.,2008. DOI: 10.1016/j.sigpro.2008.06.019.