ISSN: 2249-7196

IJMRR/April. 2025/ Volume 15/Issue 2s/264-269

Sanapati Syam/ International Journal of Management Research & Review

Semantic segmentation-based traffic sign detection and recognition using deep learning techniques

Sanapati Syam
PG scholar, Department of MCA, DNR College, Bhimavaram, Andhra Pradesh.
Ch.Jeevan Babu
(Assistant Professor), Master of Computer Applications, DNR college, Bhimavaram, Andhra Pradesh.

Abstract We present a method for detecting and classifying traffic signs based on two deep neural network architectures. A Fully Convolutional Network (FCN) – based semantic segmentation model is modified to extract traffic sign regions of interest. These regions are further passed to a Convolutional Neural Network (CNN) for traffic sign classification. We propose a novel CNN architecture for the classification step. In evaluating our approach, we contrast the efficiency and the robustness of the deep learning image segmentation approach with classical image processing filters traditionally applied for traffic sign detection. We also show the effectiveness of our CNN-based recognition method by integrating it in our system.

I. INTRODUCTION

1.1 Introduction

Convolutional neural networks have recently become ubiquitous in large-scale image recognition tasks, owing to the exponential advancement in computing power. In addition to the considerable gain in hardware performance, widely available big datasets have contributed towards state-of-theart improvements. Having pushed the boundaries in several computer vision tasks, such as object classification and detection [1], they have likewise been proven to excel at semantic segmentation.

The latter is perhaps one of the biggest challenges of the modern deep learning era. The detection and classification of traffic signs constitutes one of the key challenges in obtaining a good visual perception of traffic scenes. This stems from the fact that not only is the traffic environment very complex and subjected to continuous changes, such as weather conditions,

but also to the aspect of the traffic signs in themselves.

The 'yield' traffic sign in Germany is very different from other countries, albeit it may not seem at first sight. Such discrepancies preclude both a facile pathway towards solving this task and the generalization of its solutions. In this paper we propose a novel traffic sign detection and recognition method.

Instead of using traditional image processing methods for traffic sign detection, we train a Convolutional Neural Network (CNN) that outputs a semantic segmentation of the traffic scene. For this purpose we modify a Fully Convolutional Network (FCN) [2] architecture and extract traffic sign regions from the resulting semantic map.

The second novelty in our paper is a new CNN architecture for traffic sign recognition, which is trained to classify the traffic sign regions obtained from the aforementioned semantic map.

II. LITERATURE SURVEY

[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, "SSD: single shot multibox detector," in ECCV (1), ser. Lecture Notes in Computer Science, vol. 9905. Springer, 2016, pp. 21–37.

We present a method for detecting objects in images using a single deep neural network. Our approach, named SSD, discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape.

Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes. SSD is simple relative to methods that require object proposals because it completely eliminates proposal generation and subsequent pixel or feature resampling stages and encapsulates all computation in a single network.

[2] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640–651, April 2017.

Convolutional networks are powerful visual models that yield hierarchies of features. We show that convolutional networks by themselves, trained end-to-end, pixelsto-pixels, exceed the state-of-the-art in semantic segmentation. Our key insight is to build "fully convolutional" networks that take input of arbitrary size and produce correspondingly-sized output with efficient inference and learning.

[3] X. Ren and J. Malik, "Learning a classification model for segmentation," in Proceedings Ninth IEEE International Conference on Computer Vision, Oct 2003, pp. 10–17 vol.1.

We propose a two-class classification model for grouping. Human segmented natural images are used as positive examples. Negative examples of grouping are constructed by randomly matching human segmentations and images. In a preprocessing stage an image is over-segmented into super-pixels.

We define a variety of features derived from the classical Gestalt cues, including contour, texture, brightness and good continuation. Information-theoretic analysis is applied to evaluate the power of these grouping cues. We train a linear classifier to combine these features. To demonstrate the power of the classification model, a simple algorithm is used to randomly search for good segmentations. Results are shown on a wide range of images.

[4] J. Mehra and N. Neeru, "A brief review: Super-pixel based image segmentation methods," 2016.

Object-based image analysis is an important task in image processing. For the extraction of an accurate object, segmentation plays a vital role. The major problem raised with image segmentation is over segmentation, which can be corrected using superpixel based image segmentation.

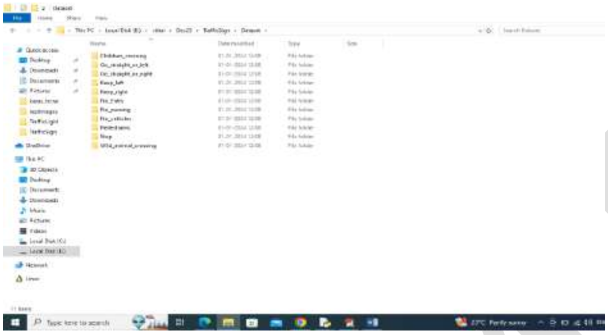
Superpixel generation is affected by the divergent distribution of pixel intensity values while acquiring the image. To overcome this problem, a novel colorization method is proposed as a preprocessing technique to enhance superpixel generation. Colorization transforms image into unique intensity values with particular color. Quantitative evaluation of experimental results yields promising accuracy, for the proposed approach.

PROPOSED METHOD

The proposed system presents a comprehensive deep learning pipeline for traffic sign detection and classification, beginning with the input of 1024x512 RGB images processed through a modified FCN8s semantic segmentation model trained on the Cityscapes dataset. This model efficiently delineates regions of interest (ROIs), particularly traffic signs, which are then cropped and passed to a CNN-based classification network trained on the GTSRB dataset. The segmentation model outperformed traditional OpenCV-based methods significantly, achieving up to 89% accuracy in complex urban scenes, while traditional filters only reached 28%. For classification, two CNN architectures were evaluated using different hidden layer configurations and activation functions (ReLU and eLU), with the best results reaching 91.5% accuracy on a balanced dataset. To handle class imbalance, random oversampling was employed, and various regularization methods were tested, though dropout and batch normalization had limited effect. The entire system, including ROI extraction and classification, processes each image in under 0.035 seconds, demonstrating both accuracy and efficiency. Limitations such as the segmentation of traffic signs not present in the classification dataset and issues with closely placed signs suggest the need for a more diverse dataset and potentially more advanced segmentation models. Nonetheless, the integration of semantic segmentation, ROI extraction using convex hulls, and deep learning classification forms a robust framework for real-time traffic sign recognition.

RESULT

In this project you asked to develop Faster-RCNN algorithm to detect and recognize Traffic signs by using GTSRB dataset. GTSRB dataset contains over 40 different traffic signals and it's difficult to train entire dataset as Faster-RCNN taking longer time for training so we have used below traffic signs from GTSRB dataset.



We took 11 different traffic signs which is showing above screens and then implemented below Faster-RCNN algorithm with images and bounding boxes

In above screen read red colour comments to know about FRCNN algorithm which getting train by using image features, class names and bounding boxes which can be used to detect and recognize signs.

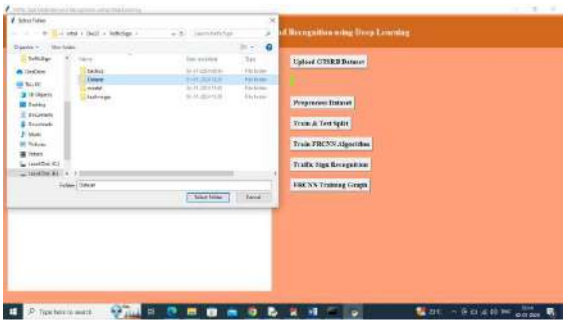
To implement this project we have designed following modules

1) **Upload GTSRB Dataset:** using this module we can upload dataset to application which will read all images, bounding boxes and class labels and then resize all images to equal sizes and then extract features from each image to generate training array
2) **Pre-process Dataset:** this module we will shuffle, normalize all images and bounding boxes features
3) **Train & Test Split:** this module we split all processed image features into train and test where application will be using 80% dataset images for training and 20% for testing
4) **Train FRCNN Algorithm:** 80% processed image features will be input to FRCNN algorithm to train a model and this model will be applied on 20% test data to calculate detection accuracy
5) **Traffic Sign Recognition:** using this module we can upload test image to application and then FRCNN will detect and recognize sign
6) **FRCNN Training Graph:** will plot FRCNN training accuracy and loss graph

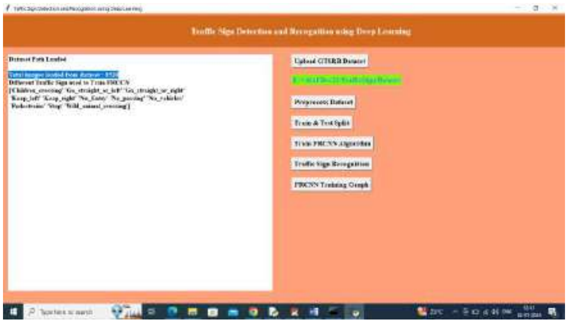
SCREEN SHOTS

To run project double click on 'run.bat' file to get below screen

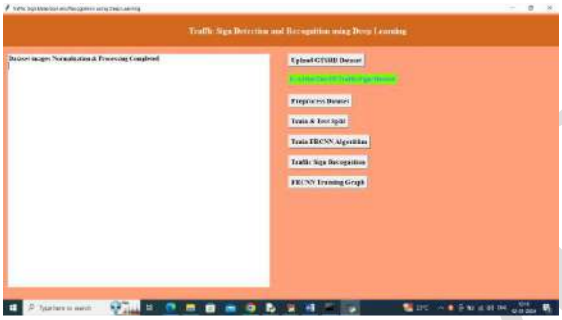
In above screen click on 'Upload GTSRB Dataset' button to upload dataset and get below output



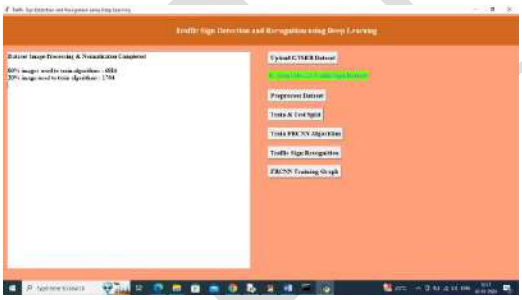
In above screen selecting and uploading 'Dataset' folder and then click on 'Select Folder' button to load dataset and get below page



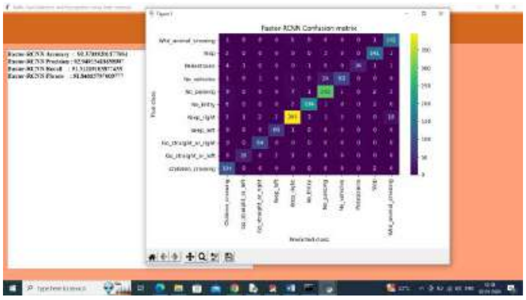
In above screen dataset loaded and in blue colour text can see total number of loaded images and then can see different signs images using to train FRCNN and now click on 'Pre-process Dataset' button to normalize image features and get below output



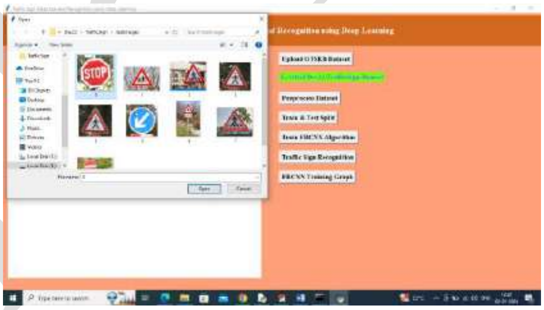
In above screen dataset normalization completed and now click on 'Train & Test Split' button to split dataset into train and test to get below output



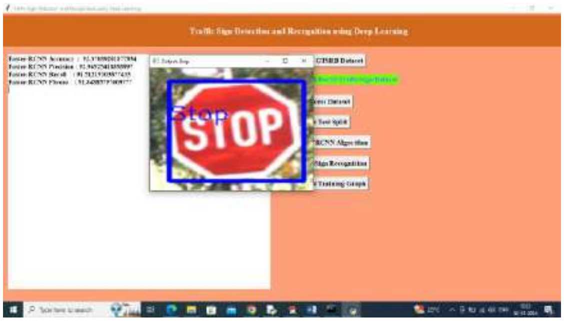
In above screen application using 6816 images (80%) for training and remaining for testing and now click on 'Train FRCNN Algorithm' button to train FRCNN and get below output



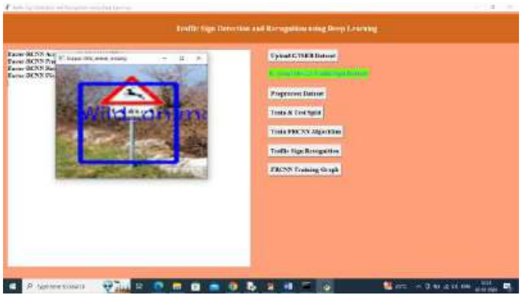
In above screen FRCNN got 92% accuracy and can see other metrics output and in confusion matrix graph x-axis represents 'Predicted Labels' and y-axis represents 'True Labels' and all different colour boxes in diagnol represents correct prediction count and remaining all blue boxes represents incorrect prediction counts which are very few and now close above graph and then click on 'Traffic Sign Recognition' button to upload test image and recognize sign



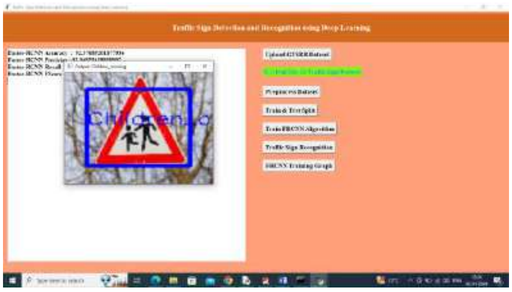
In above screen selecting and uploading '0.jpg' image and then click on ''Open' button to get below output



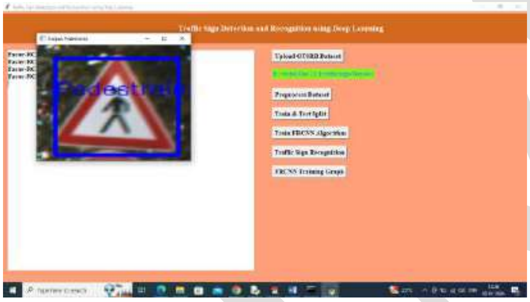
In above screen sign recognised as 'Stop' and similarly you can test other image and below are other output



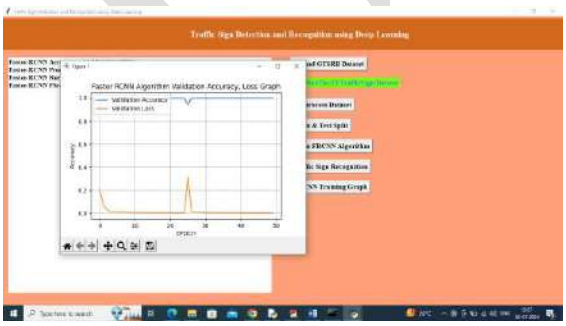
In above screen in image title can see recognised sign as ‘wild animal crossing’



In above screen sign recognized as ‘Children crossing’.



In above screen sign is ‘Pedestrians crossing’ and now click on ‘FRCNN Training Graph’ button to get below screen



In above graph x-axis represents training epoch and y-axis represents accuracy and loss values. In above graph blue line is for accuracy and orange line for loss and can see with each increasing epoch accuracy got increase and reached closer to 1 and loss got decrease and reached closer to 0.

CONCLUSION

In this paper we presented a deep learning-based traffic sign detection and recognition using semantic segmentation. An RGB image is fed into a modified FCN8s which outputs a semantically segmented image. The FCN8s is modified by replacing the last convolutional layer with dilated convolution (of rate 4) instead of the traditional convolution (of rate 1). This modification generated an increase of 0.7% in the final accuracy. Although the accuracy may have not increased by a great margin, it improved the segmentation quality of the less well-represented object categories, such as traffic signs and semaphores. The semantically segmented image is then passed on to a Region Of Interest (ROI) module, which extracts the candidate traffic signs. These extractions are cropped and sent to a classification module which assigns a label to each cropped ROI image received. The classification module consists of two different models (eLU) and (ReLU) that we created in order to contrast the efficiency of the former activation function to the latter. To the best of our knowledge, the proposed approach is novel. The possible improvements to the actual implementation have been discussed in the previous section.

REFERENCES

[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, “SSD: single shot multibox detector,” in ECCV (1), ser. Lecture Notes in Computer Science, vol. 9905. Springer, 2016, pp. 21–37.

[2] E. Shelhamer, J. Long, and T. Darrell, “Fully convolutional networks for semantic segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640–651, April 2017.

[3] X. Ren and J. Malik, “Learning a classification model for segmentation,” in Proceedings Ninth IEEE International Conference on Computer Vision, Oct 2003, pp. 10–17 vol.1.

[4] J. Mehra and N. Neeru, “A brief review: Super-pixel-based image segmentation methods,” 2016.

[5] X. Tian, L. Jiao, L. Yi, K. Guo, and X. Zhang, "The image segmentation based on optimized spatial feature of superpixel," J. Vis. Comun. Image Represent., vol. 26, no. C, pp. 146–160, Jan. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.jvcir.2014.11.005