Recognition of Traffic Signs using Deep Learning

Article · October 2020

CITATIONS
0

READS
251

1 author:

Tharmi Vijayakumaran
University of Moratuwa
1 PUBLICATION **0** CITATIONS

SEE PROFILE

Recognition of Traffic Signs using Deep Learning

Tharmi.V

Department of Interdisciplinary studies,

Faculty of Information technology

Moratuwa, Sri Lanka

tharmi14t@gmail.com

Abstract- Traffic sign recognition is an important consideration in advanced driver assistance systems, intelligent autonomous vehicles and real-world computer vision and pattern detection problem. This independent study aims to deal with the state-of-the-art machine learning algorithms in the recognition of traffic signs compare to the human traffic sign recognition performance and baseline results. This study mainly reveals how Deep learning influence the traffic sign recognition in a traffic recognition system. Convolutional Neural network, a kind of Deep Neural Network showed particularly high classification accuracies in recognition process.

Keywords—Traffic sign recognition, Traffic sign detection, Traffic sign classification, Image processing, Segmentation, Deep Convolutional Neural Network.

I. INTRODUCTION

The goal of computer vision is to detect and classify different sizes of objects in the scene. In image classification deep learning contributes a major role. One of the deep neural networks, Convolutional Neural Network (CNN) involves in the process of image classification, localization and detection. In a typical traffic sign recognition, there are two tasks, traffic sign detection and traffic sign classification. Detection of traffic sign is detecting the size and location of traffic sign in images and classification is classifying the detected traffic signs into their related sub classes. These two tasks depending on each other in a traffic sign recognition system.

If a system has a classifier without the detection step then it does not make a sense. Traffic sign recognition can commonly classifies as traditional classification methods and comprehensive learning methods. Traffic signs are located along the roads to convey drivers about road rules and limitation on the directions and speed on the road they drive. They are designed with covering a small area with large surrounding. Laws standardized the design of traffic signs in a country and it varies across the countries. Traffic signs are categorized into various shapes like circular, triangular and rectangular, octagonal and diamond shapes.

A human being can identify the traffic signs easily as it has designed in regular shapes and noticeable colors attract human beings. However, it is difficult for computer algorithms to identify these traffic signs due to motion blur, changing of illumination, decaying of colors, weather condition, littered background, partial occlusion and rotations [1]. Fig.2 shows the complicated traffic signs which contain above stated problems. There were many studies have been published on traffic sign detection but there are no systematic methodological comparisons of the approach and reference datasets are also freely not available at the initial stage of researches.

Appearance of traffic signs differs between the classes in terms of colors, presence of text and shapes. However, subclasses are very similar to each other. As the above stated problems existing it is the need to be with significant differences in visual appearances. Researchers has used one of these datasets to evaluate their approaches.

- [21] The German Traffic Sign Detection Benchmark(GTSDB). It contains 900 full images with 1206 traffic signs. Mainly three categories of traffic sings are included. The dataset mainly focused on detection purpose. Fig. 1 shows the examples for GTSDB.

- [11] The German Traffic Sign Recognition Benchmark(GTSRB) which comprises of 51840 images with 43 categories of traffic signs. It was intended for classification purpose only.

- [22] Belgium Traffic Sign Classification dataset(BTSC) comprises of 62 categories of traffic signs which are cropped images.

- [2] The Laboratory for Intelligent and Safe Automobiles dataset (LISA) consists of 6610 images with 49 classes. This is with acquired traffic signs on United States of America.

- [3] The Tsinghua-Tencent 100K dataset which is a large dataset contains the statistics of 100,000 cropped images. Of these cropped images, 10000 consists of 30000 traffic signs. This is considered as large dataset and it comprises of 45 classes.

- [23] DFG traffic sign dataset consists of 6957 images. 200 categories with 13,239 annotated traffic signs.

GTSRB dataset has been used most widely which created for the competition "The German Traffic Sign Recognition Benchmark" held at International Joint Conference on Neural Networks (IJCNN)2011[2]. In [4], the authors proposed some effective localization algorithms which conducted in real time. These algorithms were implemented using the extensive datasets from GTSRB and later on introduced database German Traffic Sign Recognition Benchmark (GTSRB) which allows unbiased comparison of different approaches to

categorizing and detecting traffic signs, respectively.

Fig.1.Examples of three categories of GTSDB.

These databases consists of training set and testing set which are used for developing the recognition and localization algorithms.

II. Deep Learning

Deep learning is a collection of techniques from artificial neural network (ANN), which is a branch of machine learning. ANNs are modeled on the human brain; there are nodes linked to each other that pass information to each other. It has the feature of complex structure connecting layers in more complex ways. "Cambrian explosion" of computing power to train, automatic feature extraction[15]

III. Convolutional Neural Network (CNN)

CNN is a multilayered biologically-inspired structure that automatically learns hierarchies of fixed features. It is a type of ANN. ANN is a network of nodes(neurons) where every node is only connected to certain other nodes and all node connections are assigned weights based on how important those connections are[8]. Many vision approaches use HOG and SIFT hand crafted features while Convolutional nets learn features at every level from the input data and shows a high performance in accuracy[12]. CNN has performed well in classification of entire images, localizing objects by ROIs. CNN provides a high accuracy in object detection. Fig. 3 shows how CNN extract features of an input image.

IV. Deep Neural Network (DNN)

Deep Neural Network is a sequence of convolutional and max pooling layers. In the DNN every layer get connections from its previous layer. It consists of fully connected layers. When the image is passed into this DNN the image will be diminished in its original form as passes every layers but the features of the image will be extracted.

Fig. 2. Examples of traffic signs which which are affected by the environmental factors.

V. Convolutional Neural Network Architecture

Deep hierarchical neural network composed of convolutional layer, max pooling layer, spatial transformer, rectified linear unit and local contrast normalization. These layers functions as feature extractors. Features in the input image is extracted using these layers and feature maps will be created. When the input image is passed through the succession of convolutional layer and max pooling layers, feature maps will be created and go through all and reach the fully connected layers.

ReLu layers acts as activation function in DNN which are made up of neurons. ReLu layers increase the non linearity property of the network without interfering the learnable parameters of the convolutional layer. The contrast of an input map is normalizes by local contrast normalization layers. Subtractive local normalization and divisive local normalization use a Gaussian kernel. Max pooling layers are used to educe the spatial size of the feature maps gradually by reducing the amount of parameters. Overfitting also controlled by the selection of superior invariant features and improvement of generalization.

Fully connected layer neurons have complete connections to all the passed layers and it will produce 1- dimensional feature vector. In CNN architecture, the last fully connected layer performs the task of classification using softmax activation function. Fig. 6 shows the Convolutional Neural Network Architecture.

VI. Traffic Sign Detection

The approaches in detection based on colours and shapes. The methods depend on colours having the advantage of easily

distinguishes from their surroundings and the task of searching for specific colours as that is straight forward in an image. Although there are some problems such as colours vary greatly with illumination, age and also with status of sign.

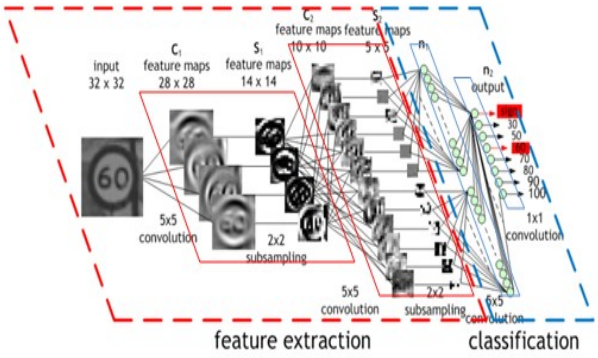

Fig. 3. CNN feature extraction and classification of an image.

Shapes of an image invariant to lighting and age. There are many detection algorithms which are rely mostly on edge detection. Edges are strong to changes in lightning, but occlusion of parts in a traffic sign makes the detection harder and when the background colour resembles the colour of the traffic sign, this spoils the edge discovery. Researchers have developed various methods to reduce these problems when implementing the systems. Eigen-gradients based oriented gradient maps, Karhunen-Loeve transform, Hough transformation, Adaptive thresholding and Adaptive shape analysis, Self-Organizing Map are some methods [5].

Traditionally features are extracted by hand crafted methods. Features extracted from the image using algorithms are defined as hand crafted features. HOG, SIFT, BOW, Gabor, Fish vectors are used. These methods are used in the case of object classes are separated properly. However above stated feature extraction methods are not preferable in huge scale object detection. Because of the non linear separation of objects. Non linear models such as SVM, Random forest have the ability to learn the non linear decision boundaries. Viola-Jones detector depend on Haar features and linear classifier which based on HOG descriptors are some of baseline algorithms commonly used in detection of traffic sign detection. In some AdaBoost and colour sensitive Haar wavelet features, and a temporal information propagation are used[9].

Initially, most researchers focused on the classification of traffic signs and the complex phase, the detection of traffic signs was ignored by researchers. Benchmarks that address the traffic sign detection, primarily cover the traffic sign subsets which are considered as important for autonomous advanced driver assistance systems and autonomous vehicles. Those benchmarks have low inter-category variance. Hand crafted detectors and classifiers are able to detect those signs. But those sign categories are not suited for real world instances. Therefore feature learning based on real world examples become a solution.

Now the world gets into deep learning. This is an interesting and complex field in machine learning. CNN has been holding a significant place in object detection. CNN has performed well in classification of entire images, localizing objects by ROIs. CNN provides a high accuracy in object detection. CNN is a non linear function and it is an end-to-end learning method which have the ability to display the image into feature space where layers are non-overlapping and separate linearly. In CNN approach to detect traffic signs from background, in the initial stage the RGB images are transformed to Gray scale images using support vector machines. Then the results are feed to CNN. Fixed layers detect the Region of Interests and learnable layers extract distinctive features for the classifier to find out traffic signs of the target group. Fig. 4. shows the process.

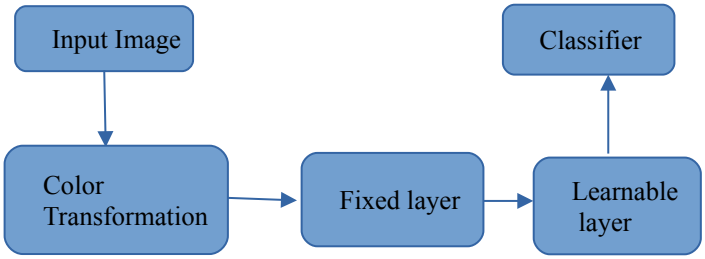

Fig. 4. Flow diagram of CNN approach in image detection

On the account of being looked at colours, most of the colour based approaches consider shapes into account.[3] Detection process can be split as the segmentation, feature extraction and detection. Most shape only based approaches don't have a segmentation step. A system that can find various kinds of signs at lower detection rate is better than a system which can detect only one particular sign type. Segmentation step is performed as it gives a rough idea about where signs may be and narrow the search area.

[6] There are many segmentation approaches available but, the best approach is selected which gives best recognition results. High recognition rate, high speed, low number of false positives and the low number of lost signs decide the best results of recognition process. In traditional approaches, segmentation is carried out based on colours. Colour based segmentation depends on a threshold of input image in some colour space. In the detection edge is another important feature. Edges can be directly obtained from raw pictures and sometimes from pre segmented images. In some Hough transform is used to process edges but, Hough transforms are cost effective and inappropriate to real-world applications.

Cascading classification another famous algorithm that is primarily based on positive and negative image samples. Positive images are considered as images in an object of interest, here traffic signs. Negative image is arbitrary image does not contain traffic signs. Cascade classifiers work on this algorithm. By combining Haar cascade classifier and deep CNN, enhances the capability of detection in real time performance.

[7] employed a concept to better the accuracy in the detection process. Haar cascade classifier and deep CNN were used. A quick search mechanism depend on simple features to detect ROI is implemented, then confirmation done on each

ROI to detect whether it contains traffic sign or not and to categorized into traffic signs and or not in the purpose of removing some false positives. This research has used GTSRB dataset and the evaluation depend on recall and precision values. This algorithm has an average precision of 98.81% and average recall rate of 98.22%.

In [13] approach, in the detection module a colour image is transformed to traffic sign probability map in which high intensity is identified in pixels situated in traffic signs and low intensity in the pixels outside the traffic sign. Then Maximally Stable Extremal Regions(MSER) are extracted and the false positives are filtered out. Here a multi-class Support Vector Machine(SVM) based on a novel color HOG feature classify the traffic sign proposals into super classes.

In [10] implemented multi scale CNN on GTSRB dataset. Here each layer's output is fed into the classifier. Combining representations of multiple stages contribute to various scales of receptive fields to the classifier. 98.97% accuracy was reached in phase 1 and at later 99.17% accuracy reached by using 2 layer classifier and by avoiding colour information. Fig. 5. shows that how CNN architecture implemented by Sermanet and LeCun learn task-specific fixed features in a hierarchical manner.

[14] Proposed a CNN based approach for traffic sign detection. Author used the SVM to transform the original image into Gray scale image and use CNN with fixed and learnable layers for detection and recognition. GTSDB dataset was used and results with an area under the precision-recall(AUC) of 99.73% in the category "Danger", and an AUC of 97.62% in the category "Mandatory".

VII. Traffic Sign Classification

The other major process in traffic sign recognition is traffic sign classification. When the detection process is completed those detected signs are for classification. Before the involvement of CNN researchers have adapted various object

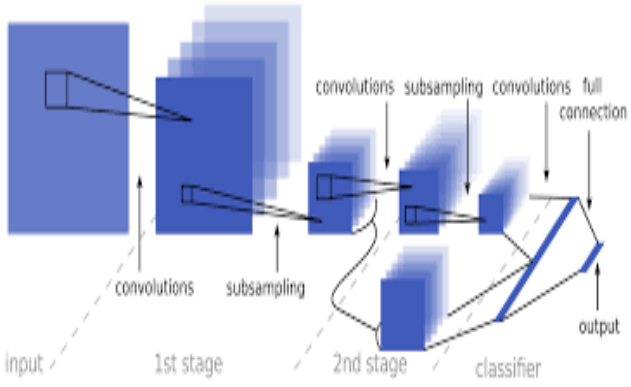

Fig. 5. CNN architecture implemented by Sermanet and LeCun

detection methodologies for traffic sign classification.

[16]Segmentation was carried out using HSI color space and candidate blobs are extracted. Then the blobs are classified into shapes using linear SVMs and those classified blobs are recognized using SVMs and Gaussian kernels. [19]K-d trees and Random Forests were used in GTSB dataset to classify the traffic signs. Two tree classifiers' performance are evaluated using various sizes HOG descriptors and distance transforms. K-d trees with HOG descriptors achieve 92.9% classification result and K-d trees with Distance Transforms achieve classification result of 67%. Random Forests improve the performance as Random Forest with HOG Descriptors achieve 97.2% and Random Forests with Distance Transforms achieve 81.8%.

Although the above methods have been widely used in traffic sign recognition Convolutional Neural Network(also called ConvNets) show high accuracy in classification in traffic sign recognition competitions. CNN became a trend in traffic sign classification during the GTSRB competition. [17] used a committee of a CNN and MLP trained on HOG and HAAR features and provide the recognition with best accuracy of 99.15%.

[10] Sermanet team has used multi scale neural network on GTSRB dataset with EBLearn open source library. 98.97% accuracy was gained in the 1st phase of competition using 32×32 coloured data. Here the convolutional network architecture is differed from traditional architecture by using multi scale features, type of non linearities used. Here first stage output is used after pooling and passed to the classifier. Then output of first stage is more subsampled to gain high accuracy. Rectified sigmoid function tanh is used with subtractive and divisive local normalization. 99.17% accuracy was gained later by ignoring the colored information and using gray scale image information and with the increase of network capacity and depth.

[3] carried out detection and classification process simultaneously. A benchmark was created which contains more images with more variableness, smaller and higher resolution when compared to others benchmarks. They have trained two networks, one as traffic sign detector and other as classifier which simultaneously detect and classify traffic signs. This approach outperforms the performance of fast R-CNN in smaller objects. Fast R-CNN has 0.56 and 0.50 recall and accuracy respectively. But this approach has 0.91 and 0.88 recall and accuracy respectively.

[25] used two neural networks on Swedish Traffic Sign Dataset. Fully convolutional neural network which is trained to predict coarse sign regions. Another CNN is used for final classification of traffic signs. The proposed method presents 97.69% and 92.9% precision and recall respectively when compared to R-CNN approach which presents 91.25% and 87.24% precision and recall respectively.

[20] has used ConvNet on the GTSRB dataset. Data augmentation was applied to reduce the effect of overfitting. In this steps are taken to enlarge the training dataset which lead CNN to have better generalization. Dataset is augmented by application of 12 transformations. Classification ConvNet is used with PReLu activation function on grayscale images. 95.55% of the test image is classified successfully with single

classification ConvNet and ensemble of three ConvNet is accomplished to classify 97.70% of samples correctly.

[23] used Faster R-CNN and Mask R-CNN in traffic sign recognition. They have compared how fully convolutional method, Faster R-CNN and Mask R-CNN have showed their results on Swedish Traffic Sign dataset and DFG datasets. Here data augmentation was applied to avoid overfitting of dataset. [24] implemented a hinge loss stochastic gradient descent approach to train deep neural network consists of 20 CNNs. GTSRB dataset was used and it presents 99.65% accuracy.

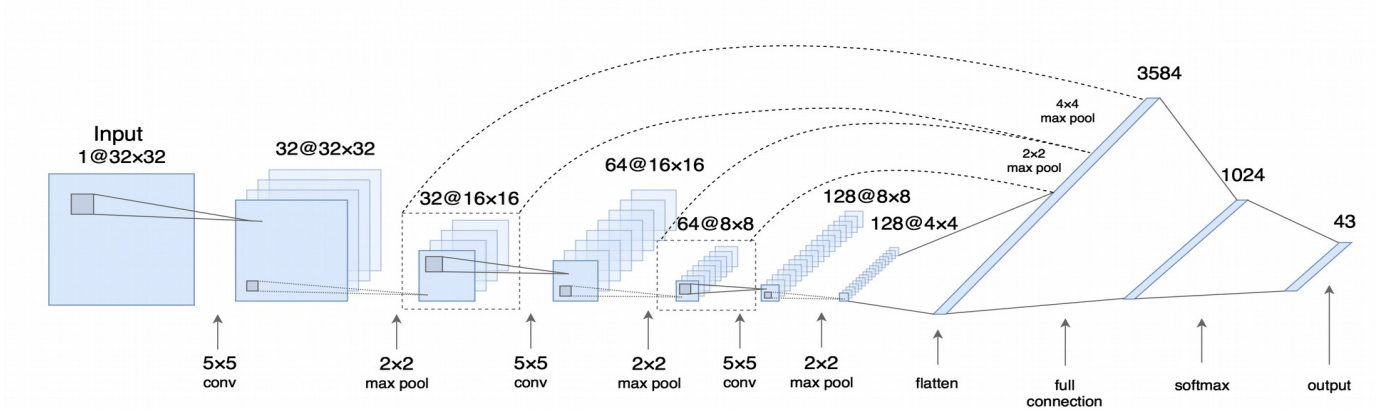

Fig. 6. Convolutional Neural Network Architecture.

VIII. Conclusion

This independent study has centered on the state of sign detection and classification in traffic sign recognition. In autonomous driving vehicles traffic sign recognition has been carried out in traditional and end-to-end learning methods.

The detection process has been done on sequence of segmentation, feature extraction and detection. Detection process can be categorized as color based detection and edge based detection. Challenging factor on color based detection are colors can be occluded, motion blur, illumination change. Edge detection has the challenge of background color being similar to color of traffic sign.

Previously detection was carried out using hand crafted methods like HOG, SIFT, Fish vectors and classification was carried out using SVM. CNN was introduced to traffic sign recognition when the need for recognition process which exists in real world scenario implemented. Approached based on CNN produced good results related to accuracy. Deep learning become popular as available of large datasets. Deep Neural Networks are used in traffic sign recognition with various kinds of convolutional neural network architectures. Deep learning approach leads to high accuracy than the traditional methods of traffic sign recognition.

Acknowledgements

The author would like to thank Dr. Sagara Sumathipala for his many useful suggestions, which have greatly improved the presentation of this paper.

References

[1] [1]Štefan Toth,"Difficulties of Traffic Sign Recognition",MICT 2012, 7-TH WINTER SCHOOL OF MATHEMATICS APPLIED TO ICT, SACHTICKY, FEBRUARY 2012, pp. 1-4

[2] [2] Møgelmose, A., Trivedi, M. M., & Moeslund, T. B. (2012)", Vision based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems",Perspectives and Survey, IEEE Transactions on Intelligent Transportation Systems, 13(4), pp.1484-1497

[3] Zhe Zhu, Dun Liang, Songhai Zhang, Xiaolei Huang, Baoli Li, Shimin Hu,"Traffic-Sign Detection and Classification in the Wild",The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), [2016], pp. 2110-2118

[4] Samuele Salti, Alioscia Petrelli, Federico Tombari, Nicola Fioraio and Luigi Di Stefano," A Traffic Sign Detection pipeline based on interest region extraction", Proceedings of International Joint Conference on Neural Networks, Dallas, Texas, USA, August 4-9, 2013, pp. 723-729

[5] Safat B. Wali , Mohammad A. Hannan , Aini Hussain , Salina A. Samad," Comparative Survey on Traffic Sign Detection and Recognition: a Review",Dept. of Electrical, Electronic & Systems Engineering, Universiti Kebangsaan Malaysia, Malaysia, 2015, pp. 38-42

[6] Hilario Gómez-Moreno, Member, IEEE, Saturnino Maldonado-Bascón, Senior Member, IEEE, Pedro Gil-Jiménez, and Sergio Lafuente-Arroyo," Goal Evaluation of Segmentation Algorithms for Traffic Sign Recognition", IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 11, NO. 4, DECEMBER 2010, pp. 917-930

[7] Lotfi Abdi, Aref Meddeb, "Deep learning traffic sign detection, recognition and augmentation", Conference Paper · April 2017.

[8] Alexander Shustanova, Pavel Yakimova*,"CNN Design for Real-Time Traffic Sign Recognition", 3rd International Conference "Information Technology and Nanotechnology", ITNT-2017, 25-27 April 2017, Samara, Russia, pp. 718-725

[9] Claus Bahlmann∗, Ying Zhu∗, Visvanathan Ramesh∗,Martin Pellkofer†, Thorsten Koehler†, "A System for Traffic Sign Detection, Tracking, and Recognition Using Color, Shape, and Motion Information", pp. 1-6

[10]Pierre Sermanet and Yann LeCun,"Traffic Sign Recognition with Multi-Scale Convolutional Networks", Courant Institute of Mathematical Sciences, New York University

[11] J. Stallkampa,∗, M. Schlipsinga, J. Salmena, C. Igelb,"Man vs. Computer: Benchmarking Machine Learning Algorithms for Traffic Sign Recognition", Institut fu ̈r Neuroinformatik, Ruhr-Universita ̈t Bochum, Universita ̈tsstraße 150, 44780 Bochum, Germany , Department of Computer Science, University of Copenhagen, Universitetsparken 1, 2100 Copenhagen, Denmark, pp.1-11

[12]Dan Cire ̧san, Ueli Meier, Jonathan Masci and Ju ̈rgen Schmidhuber,"Multi-Column Deep Neural Network for Traffic Sign Classification", 23 January 2012, pp.1- 15

[13] Yi Yang, Hengliang Luo, Huarong Xu, and Fuchao Wu," Towards Real-Time Traffic Sign Detection and Classification", IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 17, NO. 7, JULY 2016, pp. 2022-2031

[14] Yihui Wu, Yulong Liu, Jianmin Li, Huaping Liu, Xiaolin Hu,"Traffic Sign Detection based on Convolutional Neural Networks", Proceedings of International Joint Conference on Neural Networks, Dallas, Texas, USA, August 4-9, 2013, pp. 747-753

[15] Adam Gibson; Josh Patterson, Getting started with deep learning. [E-book] Available: O'Reilly Media, Inc.

[16] Saturnino Maldonado-Bascón, Sergio Lafuente-Arroyo, Pedro Gil-Jiménez, Hilario Gómez-Moreno, and Francisco López-Ferreras,"Road-Sign Detection and Recognition Based on Support Vector Machines", IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 8, NO. 2, JUNE 2007, pp. 264-278

[17] Dan Cires ̧an, Ueli Meier, Jonathan Masci and Ju ̈rgen Schmidhuber,"A Committee of Neural Networks for Traffic Sign Classification",pp. 1-4

[18] Yingying Zhu, Chengquan Zhang, Duoyou Zhou, Xinggang Wang, Xiang Bai, Wenyu Liun,"Traffic sign detection and recognition using fully convolutional network guided proposals", Neurocomputing 214 (2016) , pp.758–766

[19] Fatin Zaklouta and Bogdan Stanciulescu and Omar Hamdoun, "Traffic Sign Classification using K-d trees and Random Forests", pp. 1-5

[20] Hamed Habibi Aghdam ∗, Elnaz Jahani Heravi, Domenec Puig, "A practical approach for detection and classification of traffic signs using Convolutional Neural Networks", Robotics and Autonomous Systems 84 (2016), pp.97–112

[21] Sebastian Houben, Johannes Stallkamp, Jan Salmen, Marc Schlipsing, and Christian Igel, "Detection of Traffic Signs in Real-World Images: The German Traffic Sign Detection Benchmark",Conference Paper, August 2013, pp. 1-9

[22] SaouliAzizaEl, AroussiMohamedb,FakhriYoussef, "Traffic Sign Recognition Based On Multi-feature Fusion and ELM Classifier", Procedia Computer Science, Volume 127, 2018, Pages 146-153

[23] Domen Tabernik and Danijel Skocaj, "Deep Learning for Large-Scale Traffic-Sign Detection and Recognition", pp. 1- 14

[24] Junqi Jin, Kun Fu, and Changshui Zhang, Member, IEEE, "Traffic Sign Recognition With Hinge Loss Trained Convolutional Neural Networks", IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 15, NO. 5, OCTOBER 2014, pp. 1991- 2000

[25] Yingying Zhu, Chengquan Zhang, Duoyou Zhou, Xinggang Wang, Xiang Bai, Wenyu Liu, "Traffic sign detection and recognition using fully convolutional network guided proposals", Neurocomputing 214 (2016), pp. 758–766