Real-Time Embedded Traffic Sign Classification

MD.Nazirul Islam

Computer Science and Engineering
Bangladesh University Of Business & Techonology(BUBT)

Dhaka, Bangaladesh
16173103041@cse.bubt.edu.bd

Mohammod Masrukul Haque Mridul

Computer Science and Engineering
Bangladesh University Of Business & Techonology(BUBT)

Dhaka, Bangladesh
16172103238@cse.bubt.edu.bd

***Abstract*—Traffic sign recognition plays an important role in driver assistant systems and intelligent autonomous vehicles. Its real-time performance is highly desirable in addition to its recognition performance. This paper aims to deal with real-time traffic sign recognition, i.e., localizing what type of traffic sign appears in which area of an input image at a fast processing time. To achieve this goal, we first propose an extremely fast detection module, which is 20 times faster than the existing best detection module. Our detection module is based on traffic sign proposal extraction and classification built upon a color probability model. Then, we harvest from a convolutional neural network to further classify the detected signs into their subclasses within each superclass. Experimental results on both German roads show that both our detection and classification methods achieve comparable performance with the state-of-the-art methods, with significantly improved computational efficiency.**

Keywords—Traffic sign detection, traffic sign recognition, real-time, color probability model.

I. INTRODUCTION

Traffic sign recognition is an important function for driver assistance systems. Although it has been studied for many years, existing methods are still far from mature. The major difficulties include bad lighting condition, similar background color, partial occlusion, low quality, etc. Some difficult examples are shown in

Fig.1. The goal of the traffic signal recognition problem is to identify which type of traffic sign is in the scene.

There are usually two steps for most traffic sign recognition methods: detection and classification. The detection step finds out the region of interests (ROI), each of which contains a traffic sign; and the classification step determines the classes of the traffic signs. As methods for classification have shown great success in recent researches, we only focus on traffic sign detection method in this paper. Color-based detection methods are very popular because traffic signs are usually in fixed colors. These methods are usually fast and robust to projective deformation, but sensitive to the lighting condition. To solve this problem, some methods use the normalized RGB color space, while some other methods convert the color space from RGB to HSI (hue-saturation-intensity), which is based on human color perception. Both the two ways are effective in a certain extent, but fail on some extreme cases such as serious backlight or color cast. Another popular class of detection methods is shape-based since traffic signs are usually with regular shapes. Hough transform is a commonly used method in detecting regular shapes such as line and circle. Some methods utilize Hough transform to detect circles or triangles in the extracted edge bitmap to locate potential traffic signs, but the memory and computational requirement is quite high for large images. Some other methods adopt genetic algorithm to detect circles and achieve high robustness to projective deformation. However, these methods are too slow for practical use. In recent years, the sliding window scheme is popular in the object detection field. Two representatives are histogram of oriented gradient and Viola-Jones. Both the two approaches have been adopted in traffic sign detection and reported good results. However, as pointed out in, different methods are hard to compare because they all use different dataset. The GTSDB (German traffic sign detection benchmark) announced recently provides an open platform for testing various traffic sign detection methods. We propose a coarse to-fine method based on histogram of oriented gradient, and have won the first place for the prohibitory and mandatory category, and the third place for the danger category. In this paper, we first give a description of the proposed method, then describe the training strategy and present experimental results. Finally, we make a conclusion of the proposed method..

II. RELATED WORK

Chronologically, approaches of published works on trac sign recognition systems evolved from color and shape based methods to machine learning based methods. In recent times, CNNs have attracted attention in pattern recognition and computer vision research, and have been widely adopted for both object detection and recognition. Color based approaches are very common. These methods use different color spaces for segmentation of the road image such as RGB (Escalera et al., 1997), HIS (Maldonado Bascon et al., 2007) or HSV (Shadeed et al., 2003), among others. The shape-based method is another popular approach for trac sign recognition and detection. Symmetry information of circular, triangular, squared and octagonal shapes are used in (Loy & Barnes, 2004), a radial symmetry detector is proposed in

(Barnes et al., 2008), Hough transforms are investigated in (Barnes et al., 2010) and a circular trac sign recognition system is studied in (Kaplan Berkaya et al., 2016). One of the main problems before the year 2011 was the lack of a publicly available trac sign dataset. The Belgian Trac Sign Dataset (BTSD) (Timofte et al., 2011), the German Trac Sign Recognition and Detection Benchmark (GTSRB and GTSDB) (Stallkamp et al., 2011), and more recently, the Croatian trac sign dataset (MASTIF) (Jurasic et al., 2015), the Dataset of Italian Trac Signs (DITS) (Youssef et al., 2016) and the Tsinghua-Tencent 100K benchmark (Zhu et al., 2016) solved this issue and boosted the research in TSRS because some of them are commonly used to evaluate the performance of computer vision algorithms for trac sign detection and recognition. Mathias et al. (2013) propose fine grained classification applying die rent methods through a pipeline of three stages: feature extraction, dimensionality reduction and classification. On GTSRB, they reach 98.53% of accuracy merging grayscale values of trac sign images and Histogram of Oriented Gradients (HOG) based features, reducing the dimensionality through Iterative Nearest Neighbour based Linear Projections (INNLP) and classifying with Iterative Nearest Neighbours (Timofte & Van Gool, 2015) (INNC). Although Support Vector Machines (SVM) (Salti et al., 2015), Random Forests (Zaklouta et al., 2011) and Nearest Neighbours (Gudigar et al., 2017) classifiers have been used to recognise trac sign images, Convolutional Neural Networks (Lecun et al., 1998), (2012) won the GTSRB contest (Stall Kamp et al., 2012) with a 99.46% accuracy thanks to a committee of 25 CNN by using data augmentation and jittering. Sermanet & LeCun (2011) use multi-scale CNN achieving an accuracy of 98.31%, second place in the GTSRB challenge. Later, Jin et al. (2014) propose a hinge loss stochastic gradient descent method to train an ensemble of 20 CNNs that brought o↵ 99.65% accuracy and eared a faster and more stable convergence than previous works. However, these approaches can still be improved avoiding the use of hand-crafted data augmentation and keeping away from applying multiple CNNs in an ensemble or in a committee way for the reason that it normally leads to higher memory and computation costs. Abbreviations and Acronyms

III. MODEL ARCHITECTURE

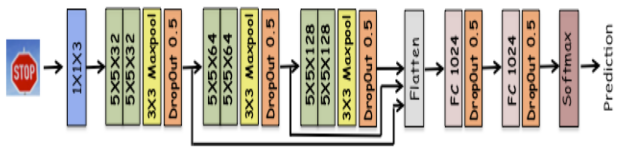

Fig.2 presents the model architecture we will use. This architecture was converged upon after trying several different architectures.

he first module in the model above is comprised of 3 1X1 filters. These filters have the effect of changing color maps. In most applications, changing color map can result in significant improvements in performance. However, it is not clear what the best color map is for different applications, therefore using 3 1X1 filters results in a case were the network itself ends up choosing the best color map.

The next 3 modules are composed of 32, 64 and 128 (respectively) 3X3 filters followed by max pooling and dropouts. The output from each of the convolution module is fed into a feedforward layer. Rationale being that the fully connected layer has access to outputs from low level and higher level filters and has the ability to choose the features that works the best. The feed forward layers are composed of 2 hidden layers with 1024 neurons in each layer. Additional dropout layers are applied after each of the fully connected layers. The last soft max layer is used to compute the log-loss of model prediction. In addition a l2-regularization cost is included to penalize large model weights.

IV. DATASET

The traffic sign recognition benchmark (TSRB) [20] used for evaluation purposes in this paper consists of color images of traffic signs (one traffic sign per image, with a total of 43 types of traffic signs) with image sizes varying from 15 15 to 250 250 pixels. There are a total of 39,209 color images in the training set and a total of 12,630 images in the test set. To balance the number of samples in different classes as well as improve the generality of the resulting network, a number of different data augmentation techniques were leveraged including: i) rotation, ii)shifting, iii)sharpening, iv) Gaussian blur, v) motion blur, vi) HSV augmentation, and vii) mirroring. As standard for evaluating. Performance using GTSRB, all images are cropped and all images are resized to 48 *48 pixels .To evaluate the accuracy of the network, the top-1 accuracy was computed on the GTSRB test set

Fig.3. Example of images from training set

V . EXPERIMENTS

In this section, we conduct our experiments on German roads to show the performance of our methods. We first give a brief introduction of data sets and experimental setups. Then, the detection results as well as the classification results are reported.

For the tests we use a CPU Intel Core2 Duo T6500 at 2.10GHz with 2GB DDR3 RAM (used for 16x16 data) and a DETECTOR Classifier B Classifier A Classifier C CPU Intel Core(TM)2 DuoE8400 at 3.0GHz with 4GB DDR3 RAM (32x32 data). It is no secret that this type of hardware does not guarantee high processing speed. We must stress the fact that the purpose of this thesis is in fact not to reach top classification performances and neither to run the training for a very high number of epochs. We focus on comparing performances of different study cases with the same network model and training parameters. We are satisfied with the results once the training is converged. Then we compare the study cases and draw conclusions.

A. Dataset organization

As mentioned in Section4.1the GTSRB is composed of highly unbalanced data. Some of the classes only have 200 images, which is quite a limitation for deep learning. We use the same subset of images for validation and testing (there is no update of hyper parameters during the training to justify validation-only data). We still keep track of both training and validation accuracy to avoid over fitting. The data is sorted out randomly as shown in Figure4.4: roughly one third of images are used for training and two thirds for testing.

Class num.	Class name	Traffic sign	IJCNN 2011 class group	Total available samples
00	20 Speed limit		red-round, speed	210
01	30 Speed limit		red-round, speed	2220
02	50 Speed limit		red-round, speed	2250
03	60 Speed limit		red-round, speed	1410
04	70 Speed limit		red-round, speed	1980
05	80 Speed limit		red-round, speed	1860
06	End of 80 speed limit		end-of	420
07	100 Speed limit		red-round, speed	1440
08	120 Speed limit		red-round, speed	1410
09	No overtaking		red-round, red-other	1470
10	No overtaking by heavy vehicles		red-round, red-other	2010
11	Crossroads		danger	1320
12	Priority road		spezial	2500
13	Give way		spezial	2510
14	Stop and give way		spezial	780
15	Restricted vehicular access		red-round, red-other	630
16	No large heavy vehicles		red-round, red-other	420
17	No entry for vehicular traffic		spezial	1110
18	Other danger		danger	1200
19	Bend, to left		danger	210
20	Bend, to right		danger	360
21	Double bend, first to left		danger	330
22	Uneven road		danger	390
23	Slippery road		danger	510
24	Road narrows on right		danger	270
25	Road works		danger	1500
26	Traffic lights		danger	600
27	Pedestrians		danger	240
28	Children crossing		danger	540
29	Bicycles crossing		danger	270
30	Snow		danger	450
31	Animal crossing		danger	780
32	End of all speed limits		end-of	240
33	Turn right ahead		blue	659
34	Turn left ahead		blue	423
35	Ahead only		blue	1200
36	Go straight or right		blue	390
37	Go straight or left		blue	210
38	Keep right		blue	2070
39	Keep left		blue	300
40	Roundabout mandatory		blue	360
41	End of no overtaking		end-of	240
42	End of no overtaking by trucks		end-of	240

Fig.4. All 43 benchmark classes in GTSRB.

B. Experimental results

We using the raw 32x32 rescaled images reaches 98.98% accuracy on the test data in about 30 iterations. Below in the table the classification rate and the false alarm rate obtained by averaging class rates. The confusion matrix is shown as well. The empty cells in the matrix are zeros.

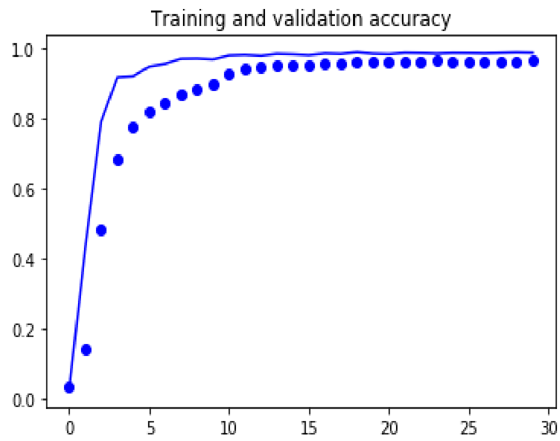

Fig.5. Training ad Validation Accuracy

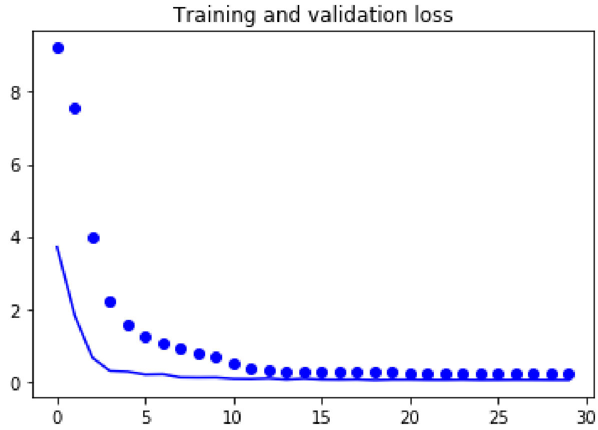

Fig.6. Training and validation Loss

N. of signs	Avg. classification rate [%]	Avg. false alarm rate [%]
11763	99.40	0.63

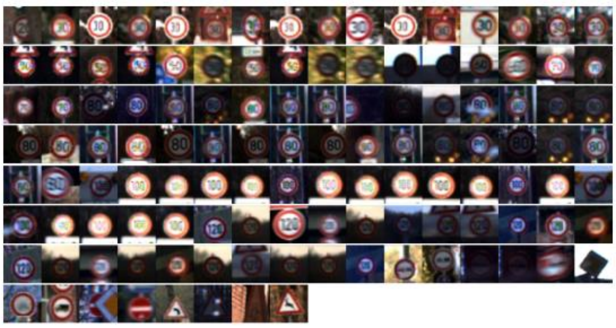

Fig.7 The classification errors in CS1.

VI. CONCLUSIONS

In this paper, a highly compact deep convolutional neural network called Micron Net is introduced for real time embedded traffic sign recognition. By designing a highly optimized network architecture where each layer's microarchitecture is optimized to have as few parameters as possible, along with macro architecture augmentation and parameter precision optimization, the resulting Micron Net network achieves a good balance between accuracy and model size as well as inference speed. The resulting Micron Net possess a model size of just ~1MB and ~510,000 parameters (~27x fewer parameters than state-of-the-art), requires just ~10 million multiply-accumulate operations to perform inference (with a time-to-compute of 32.19 ms on a Cortex-A53 high efficiency processor), while still achieving a top-1 accuracy of 98.9% on the German traffic sign recognition benchmark, thus achieving human-level performance. These experimental results show that very small yet accurate deep neural network architectures can be designed for real-time traffic sign recognition that are well-suited for embedded scenarios. Future work involves exploring extensions upon Micron Net across a larger range of traffic datasets to improve generalizability in different scenarios. Furthermore, it is also worth exploring and investigating this integrated micro architecture level and macro architecture-level design principles and optimization strategies on deep neural network architectures for different tasks outside of traffic sign recognition, and the fundamental tradeoffs between microarchitecture-level and macro architecture-level design principles and optimization strategies on such deep neural network architectures and mechanisms to optimize for such tradeoffs to improve generalizability of such an approach. Furthermore, model stability studies that also involve assessing the performance of this approach in the case of less training data given smaller model sizes would be quite interesting to explore as future work. Finally, model performance studies with a wider variety of embedded processors at different floating-point and fixed-point precision levels would be interesting to explore as future work.

REFERENCES

[1] H. H. Aghdam, E. J. Heravi, and D. Puig, ''A practical approach for detection and classification of traffic signs using convolutional neural networks,'' Robot. Auton. Syst., vol. 84, pp. 97–112, Oct. 2016.

[2] H. H. Aghdam, E. J. Heravi, and D. Puig, ''A practical and highly optimized convolutional neural network for classifying traffic signs in real-time,'' Int. J. Comput. Vis., vol. 122, no. 2, pp. 246–269, 2017.

[3] A. Arcos-García, J. Álvarez-García, and L. M. Soria-Morillo, ''Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods,'' Neural Netw., vol. 99, pp. 158–165, Mar. 2018.

[4] C. Fang, S. Chen, and C. Fuh, "Road-sign detection and tracking," IEEE Transactions on Vehicular Technology, vol. 52, no. 5, pp. 1329– 1341, 2003.

[5] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez Moreno, and F. Lopez-Ferreras, "Road-sign detection and recognition based on support vector machines," IEEE Transactions on Intelligent Transportation Systems, vol. 8, no. 2, pp. 264–278, 2007.

[6] Y. Y. Nguwi and S. Y. Cho, "Emergent self-organizing feature map for recognizing road sign images," Neural Computing & Applications, vol. 19, no. 4, pp. 601–615, 2010.

[7] M. A. Garcia-Garrido, M. A. Sotelo, and E. Martin-Gorostiza, Fast road sign detection using hough transform for assisted driving of road vehicles, ser. Lecture Notes in Computer Science. Springer Berlin, Heidelberg, 2005, vol. 3643, pp. 543–548.

[8] N. Barnes, G. Loy, D. Shaw, and A. Robles-Kelly, "Regular polygon detection," in Proceedings of the Tenth IEEE International Conference on Computer Vision, 2005, pp. 778–785.

[9] Y. Aoyagi and T. Asakura, "A study on traffic sign recognition in scene image using genetic algorithms and neural networks," in Proceedings of the 1996 IEEE IECON 22nd International Conference on Industrial Electronics, Control, and Instrumentation, vol. 3, 1996, pp. 1838–1843.

[10] A. d. l. Escalera, J. Armingol, and M. Mata, "Traffic sign recognition and analysis for intelligent vehicles," Image and Vision Computing, vol. 21, no. 3, pp. 247–258, 2003.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 886–893.

[12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2001, pp. 511–518.

[13] Y. Xie, L.-f. Liu, C.-h. Li, and Y.-y. Qu, "Unifying visual saliency with hog feature learning for traffic sign detection," in Proceedings of the IEEE Intelligent Vehicles Symposium, 2009, pp. 24–29.