Traffic Sign Recognition and Classification with Modified Residual Networks

Dhanush Siva Rohit Kollipara
dhanush.20bci7210@vitap.ac.in

Naga Bhavya Sri Panchakarla
nagabhavyasri.20bci7213@vitap. ac.in

Shankar Jagatha
shankar.20bci7218@vitap.ac.in

Abstract—In this paper, we propose a novel methodology to classify traffic signs with the purpose of boosting the classification accuracy. Our model consists of a modified Residual Networks (ResNets). We tried with other model to compare the accuracy. The other models are CNN and VGG16.The image data preprocessing includes color space conversion, and data normalization. The modified Residual Networks yields a competitive performance. Experimental result shows the robustness of our model and its superiority. We have achieved the performance of 97.24% on the German traffic sign recognition Benchmark (TRAFFIC SIGN DATASET) dataset.

I. INTRODUCTION

Traffic sign recognition plays an important role in the visual system of the autonomous vehicle. However, the recognition and classification of traffic sign is just based on 2D images. Based on 2D image, there are many influencing factors, such as poor lighting, motion blur, color deterioration and partial occlusion, etc. How to overcome the above difficulties and make the recognition highly precise are the key works in this paper.

Generally, the traffic sign recognition consists of two stages: one is sign detection and the another one is classification. There are some existing approaches for traffic sign detection: 1) the Viola-Jones approach based on the AdaBoost algorithm [1] [2]; 2) the histogram of gradients (HOG)-based approach[3] [4]; 3) the model-based approach based on the color and shape [5]. At present, for sign recognition, there are two more popular methods: support vector machine (SVM) [6] and convolutional neural networks (CNNs) [7]. Some researchers even combine SVM and CNNs to recognize traffic signs [4]. The traditional methods based on SVM need rich features extracted manually, they are more time-consuming. This paper utilizes colorbased model and a modified Residual Networks (ResNets) [8] to receive highly accurate traffic sign recognition and classification.

This paper focuses on traffic sign recognition and classification based on the 2D image with convolutional neural networks. We first convert color space from RGB (red, green, blue) to YUV (luminance, chrominance), then the CNNs extracts features and makes a classification. We use the dataset from TRAFFIC SIGN DATASET to train and test the CNNs.

The remainder of the paper is structured as follows: Section II describes the way to convert color space from RGB color space to YUV color space and data augmentation. Section III shows the architecture of our convolutional neural network in detail. Section IV mainly analyzes the experimental results.

II. IMAGE PREPROCESSING AND DATA AUGMENTATION

In this paper, we only use the TRAFFIC SIGN DATASET to train our CNNs. Although, there are totally 58 classes with corresponding labels of traffic signs shown in Fig. 1, the number in each class is not enough to train a good CNNs model.

Fig. 1. The total 58 classes in TRAFFIC SIGN DATASET. From top to bottom, there are four categories: prohibitory, danger, mandatory, and others.

A. Data augmentation

Given that the CNNs was trained using only the TRAFFIC SIGN DATASET dataset, the number in each class shown in Fig.2, it was necessary to do some simple data augmentation to avoid overfitting and to improve generalization. For this purpose, each training example was rotated about its geometrical center for angles in the range $[-30^\circ, 30^\circ]$ using steps of three degrees. In this way, the size of training dataset was increased by a factor of 21.

B. Image preprocessing

For image preprocessing, we here first convert color space due to CNNs can extract rich features by itself. The purpose of converting color space is to help CNNs extract

TABLE I

THE DETAILED ARGUMENTS FOR HIDDEN LAYERS. THE OUTPUT SIZE IS AN OUTPUT SIZE OF EACH LAYER. THE NUMBER IN THE OUTPUT SIZE REPRESENTS A*A. IN THE 'OTHER' ROW, THE NUMBERS ARE THE PARAMETERS OF MAX POOLING AND DROPOUT RESPECTIVELY. "-" MEANS THAT HERE DOES NOT NEED THIS KIND OF ARGUMENT.

Layer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Output size	32	32	32	32	32	32	32	32	16	16	16	16	8	8	8	8	8	4	4	4	4	4	4	4	4	1	-
Feature map	-	8	-	8	-	32	-	-	-	64	-	-	-	96	-	-	-	-	128	-	-	43	-	-	-	-	-
Other	-	-	-	-	-	-	-	-	2	-	-	-	2	-	-	-	0.9	2	-	-	-	-	-	-	0.9	4	-

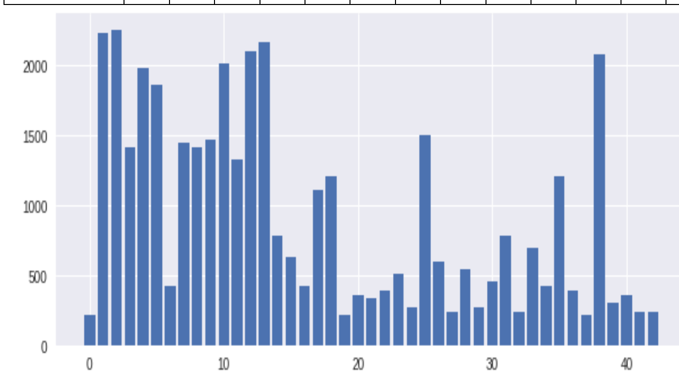

Fig. 2. The number in each class of TRAFFIC SIGN DATASET. The horizontal axis shows class number, the vertical axis shows the number in each class. Each class has different example number and it is necessary to do data augmentation.

features more efficiently. RGB color space is sensitive to light changing, so some researches convert color space from RGB to HSI [10] or normalized RGB [11]. However, the HSI color is partially invariant to illumination changing. Even both the HSI and the normalized RGB space still fail on some severe situations. Some researches [12], [13], and [14] show YUV color space has more advantages in illumination changing and shape detection shown in Fig.3. The process of converting RGB to YUV is shown in the following equations:

(4)

III. CONVOLUTIONAL NETWORKS ARCHITECTURE

A. Architecture

In recent years, machine learning has become more popular for several object classification. Research [14] adopted YUV space and a two-stage convolutional network, which received a good performance, 95.97%, on traffic sign recognition. The above neural architecture shows the excellent performance with *global* and *local* features in the case of two stages. Compared with general ConvNet, this two-stage structure shows its advantage to extract salient features. [15] adopted an inception with dimension reductions and received the 6.67% error of top-5 in ILSVRC 2014. Recently, [8] proposed a Residual Network shown in Fig. 4 and received significant improvements over the state-of-the-art on CIFAR10, CIFAR-100, SVHN, and ImageNet respectively. Taking the advantages of this three methods, our neural network is born. Our ResNets architecture is shown in Fig. 5 and consists of the following components:

- We define the section before the block 1 as preprocessing layers. In these layers, we add two special layers: Batch normalization [16] and Parametric ReLU [17].
- After the preprocessing layers, there are four Blocks. Different from the Block in [8], the layers' order is adjusted. Block 1 and Block 2 have the same architecture. Comparing with Block 1 and 2, Block 3 adds one more Dropout layer.
- Comparing with the others Blocks, Block 4 is more complex. Two sizes of convolutional networks extract features again and again.
- The last layer is softmax layer, which returns several probabilities, the predicted class has the maximum probability.

B. Key ideas and details

As show in Fig. 5, there is a Batch Normalization layer in all blocks. Batch normalization can make the distribution of each layers inputs changes during training more stable and reduces the internal covariate shift. In addition, we use much higher learning rates and be less careful about initialization based on Batch Normalization.

Layer two and layer four followed by a parametric ReLU respectively. Parametric Rectified Linear Unit (PReLU) generalizes the traditional rectified unit and improves model fitting with nearly zero extra computational cost and little overfitting risk.

Comparing with ConvNet, our ResNets is designed to extract deeper features and prevent them overfitting. The purpose of making 1x1 convolution follow 3x3 convolution is that it can reduce the dimension of the 2D image in the convolutional process and it can go wide without simply stacking more layers [18]. There are four Blocks in total. The four blocks are concatenated by order. Different from the ResNets, we adjust the layers' order and do not feed the previous output into all subsequent blocks. The detailed connected method.In order to prevent overfitting effectively, each dropout layer with 90% ratio of dropped outputs. At the end of the network, the layer is a linear layer with softmax loss as the classifier. The detailed arguments of each layer shown in Table I.

C. Training

The networks was trained using the Adam optimization algorithm [19] with an initial learning rate of 0.1, the learning rate is stepped from 0.1, 0.01, 0.001, 0.0001 at rounds 0, 3, 6, 8 respectively. Batch size should be preferably greater than the number of class (43) and hence we choose 128 as the batch size. We use the cross-entropy loss as the objective function.

The proposed network was trained and tested using the machine learning library Tensorflow .

IV. EXPERIMENTAL DATASET AND RESULT

A. Dataset

In this paper, the input image has the size of 32*32. After the data augmentation, the traffic sign classification, 58 classes, which totally possesses 4170 training samples. For iterative validation of training accuracy, the 30% training dataset was divided as validation dataset and the rest 70% is the new training dataset. The detailed number in each dataset shown in Table
II.

TABLE II

```
Train Shape: (3336, 32, 32, 3)
Test Shape : (834, 32, 32, 3)
```

B. Result

In the training process, there are more than 50 epochs. Based on the Python with Tensorflow, the training accuracy reached 100% and the validation accuracy reached 98.66%. The training and validation accuracies over epochs as shown in Fig. 6. There are some predicted results shown in Fig. 7. As we can see in Fig. 7, our model has high recognition degree of reliability under top-5 error.

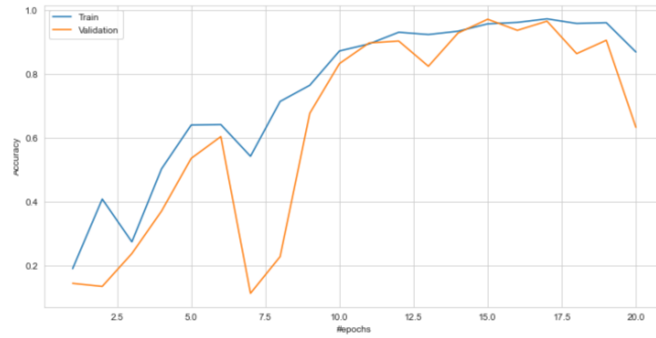

Fig. 6. The training and validation accuracies over epochs. The green color represents the training accuracy and the red color represents the validation accuracy.

Comparing with the previous researches, the proposed method has good performance on the TRAFFIC SIGN DATASET. The details are shown in table III. Comparing with testing accuracy, our method achieves a competitive result with significantly improved computational efficiency.

V. QUALITATIVE ANALYSIS

As our experiments show, the validation accuracy reached 97.56%. To interpret the detailed reasons, we here study the influencing degree to our system of each part: color space, data, and our networks model.

TABLE III

THE COMPARISON BETWEEN OUR PROPOSED METHOD AND THE STATE-OF-ART METHODS ON THE TESTING IMAGES OF TRAFFIC SIGN DATASET

Methods	[13]	[20]	[21]	[22]	[4]	ours
Accuracy(%)	99.46	97.2	99.52	99.65	97.75	99.66
Time(ms)	11.4	-	40	-	3	2

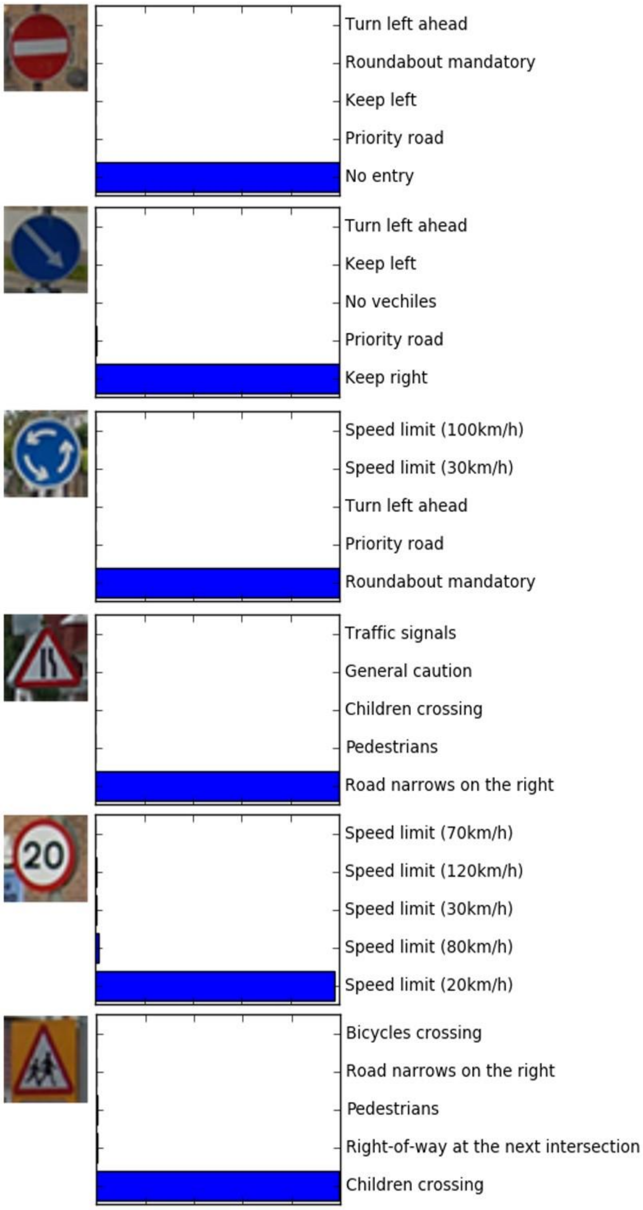

Fig. 7. Some examples and corresponding predicted results. The left represents the example, the right represents the corresponding result.

- For networks model, we keep other parameters same, only to change block number of networks. The corresponding accuracy of different blocks is showed in table IV. From table IV, the networks with four blocks get the highest accuracy, 98.66%. the blocks number from two to five did not impact much accuracy. However, the networks with only one block get a obvious lower accuracy comparing with the other four experiments. The reason is that the networks lack of the structure of our ResNets.
- For color space, the YUV color space improves the validation accuracy up to 1% comparing with RGB color space.

By the above qualitative analysis, we can know our networks' structure is the key to utilize high recognition accuracy. Besides networks' structure, the data augmentation is the other key factor for the system to achieve high accuracy.

Although we use the structure of residual networks, our networks has more advantages than original residual networks. We utilize less than 20 convolutional layers and improve more than 10% validation accuracy comparing with original residual networks.

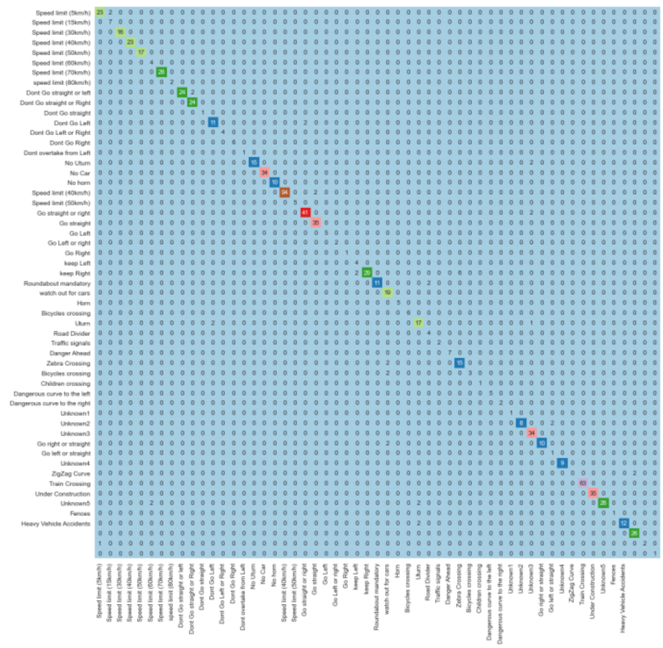

Fig. 8. The Confusion matrix of the trained model

VI. Conclusion

In this paper, we aim at the recognition and classification of TRAFFIC SIGN DATASET with the purpose of boosting the recognition and classification accuracy. To this end, we proposed a fast and high accurate model for traffic sign recognition and classification. Then the 2D images with YUV color space input into our proposed network architecture. Finally, the algorithm we proposed reached 98.66% validation accuracy. Our network extracted salient features based on "global" and "local" features. Comparing with the GoogLeNet network, our network consumes little computing resource and save more training time.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2017-20160-00318) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

[1] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, May 2004.

[2] T. Chen and S. Lu, "Accurate and efficient traffic sign detection using discriminative adaboost and support vector regression," *IEEE*

Transactions on Vehicular Technology, vol. 65, pp. 4006–4015, June 2016.

[3] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05) Volume 1 - Volume 01*, CVPR ’05, (Washington, DC, USA), pp. 886–893, IEEE Computer Society, 2005.

[4] Y. Yang, H. Luo, H. Xu, and F. Wu, “Towards real-time traffic sign detection and classification,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, pp. 2022–2031, July 2016.

[5] S. Houben, “A single target voting scheme for traffic sign detection,” in *2011 IEEE Intelligent Vehicles Symposium (IV)*, pp. 124–129, June 2011.

[6] X. Yuan, X. Hao, H. Chen, and X. Wei, “Robust traffic sign recognition based on color global and local oriented edge magnitude patterns,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, pp. 1466–1477, Aug 2014.

[7] J. Jin, K. Fu, and C. Zhang, “Traffic sign recognition with hinge loss trained convolutional neural networks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, pp. 1991–2000, Oct 2014.

[8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CoRR*, vol. abs/1512.03385, 2015.