

PAPER • OPEN ACCESS

Traffic Sign Recognition with a small convolutional neural network

To cite this article: Wenhui Li *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **688** 044034

View the [article online](#) for updates and enhancements.

You may also like

- [Application of machine learning methods for traffic signs recognition](#)
D V Filatov, K V Ignatev, A V Deviatkin et al.
- [Traffic Sign Recognition with Neural Networks in the Frequency Domain](#)
Florian Franzen, Chunrong Yuan and Zhong Li
- [Traffic sign detection method based on Faster R-CNN](#)
Linxiu Wu, Houjie Li, Jianjun He et al.

Recent citations

- [Impact of COVID-19 pandemic on low-carbon shared traffic scheduling under machine learning model](#)
Xin Liu and Shunlong Li
- [A complex junction recognition method based on GoogLeNet model](#)
Chengming Li *et al*

Traffic Sign Recognition with a small convolutional neural network

Wenhui Li¹, Daihui Li² and Shangyou Zeng^{1*}

¹ School of Electronic Engineering, Guangxi Normal University, Guilin, Guangxi Zhuang Autonomous Region, 541004, China

²School of Electronic Engineering, Guangxi Normal University, Guilin, Guangxi Zhuang Autonomous Region, 541004, China

*Corresponding author's e-mail: zsy@mailbox.gxnu.edu.cn

Abstract. Traffic sign recognition is an important part of the intelligent transportation system and has important application prospects in driverless vehicles and driver assistance systems[1]. In the image recognition of traffic signs, according to the image features of traffic signs, the common methods include traditional template matching method[2], SVM method[3], random forest[4] and the best Convolutional Neural Networks (CNN) method. In this paper, a new CNN is proposed. Feature extraction, compared with the traditional CNN method, has higher accuracy, fewer parameters, smaller models and easier training, which is evaluated on the German Traffic Sign Recognition Benchmark (GTSRB) and the Belgium Traffic Sign Dataset (BTSD). The results show that this method is superior to the traditional CNN method in traffic identification.

1. Introduction

With the improvement of material life, the car penetration rate has also increased substantially, and the expressway has also become complicated[5]. The frequent occurrence of traffic accidents has caused a lot of personal injury or property loss[6], and traffic signs play an important role in road safety. The role of the traffic signs contains a wealth of road information, can timely convey important traffic information to the driver (for example, prohibit overtaking, speed limit, reduce traffic accidents), guide the driver to make a reasonable response, ease traffic pressure, reduce traffic accidents, It is conducive to road traffic safety.

In the actual situation, the traffic signs image collected by the camera on the car, real-time traffic signs recognition[5], in the actual scene, there is various factors affect the identification of traffic signs, such as lighting factors, light intensity will lead to image exposure, light weakness will result in dim image, pictures collected during car driving, the image will be blurred due to vibration, traffic signs are outdoors all the year, there are uncontrollable factors Traffic signs are corroded, etc. [7]. The CNN extracts the features of the input images by itself, and continuously learns the feature information of the pictures through training. It is more abundant than the artificial filtering features and more anti-interference. This paper designed two networks a traditional CNN, an improved CNN, uses convolution pooling to extract low-dimensional features and high-dimensional features of images to achieve higher accuracy and lightweight models.



2. Related work

Using the collected traffic signs image information to identify the traffic signs image has its important theoretical research significance and practical value[8]. At present, the identification methods of traffic signs include template matching method, SVM method and random forest method. Template matching method compares image features with feature database data, When the characteristic parameter of the acquired image is within a certain range, it is determined that the traffic signs information is. The SVM method is according to the correspondence between the color and shape attributes of the traffic signs, thereby completing the shape classification and type discriminating function of the traffic signs[9]. The HSV-HOG-LBP descriptor was constructed by feature fusion, and the random forest was used to identify the traffic signs[10]. However, since the image is inevitably subject to errors such as shape distortion and color distortion during the acquisition, the above-mentioned method is used to identify the traffic signs in the actual use scenario, and the success rate and accuracy are not particularly high, even if the image processing algorithm is optimized. There are also many limitations. The CNN network is a network that extracts image features by itself. In traffic signs recognition, the traditional CNN network has a good effect on recognition rate and training speed. For example, VGGNet is a set of convolutional neural network algorithms developed by the Visual Geometry Group (VGG) at Oxford University, including VGG-11, VGG-11-LRN, VGG-13, VGG-16, and VGG-19 [11]. GoogLeNet is the winner of the 2014 ILSVRC image classification algorithm, and is the first large-scale convolutional neural network formed by stacking Inception modules. ResNet from Microsoft's artificial intelligence team, Microsoft Research, is the winner of the 2015 ILSVRC image classification and object recognition algorithm, which outperforms GoogLeNet's third-generation version, Inception v3[12]. ResNet[13] is a large-scale convolutional neural network built using residual blocks. ResNet is the focus of its residual blocks built by hopping connections in the hidden layer. The stacking of residual blocks alleviates the gradient vanishing problem that is common in deep neural networks and is used by many subsequent algorithms, including Inception v4 in GoogLeNet[14]. This paper is a small convolutional neural network for traffic signs recognition. It has a more accurate feature extraction than traditional convolution. It has higher accuracy and is more targeted than traditional networks. It is much smaller in parameters than traditional CNN. The model is easy to transplant, suitable for small platforms and more practical.

3. Network architecture

3.1. Convolutional neural network works

The picture input convolutional neural network, then forwards the network through the forward propagation of the network to obtain the actual output of the network. By calculating the error of the actual output and the label, the weight and offset of the network are updated, and the principle is as follows.

- [1] Randomly seek N samples from the sample set as a training set;
- [2] Set each weight and offset initial value to initialize the learning rate;
- [3] The picture is input into the network to calculate the actual output vector of the network;
- [4] Compare the elements in the output vector with the elements in the target vector to calculate the output error;
- [5] Calculate the adjustment amount of each weight and the adjustment amount of the offset, adjust the weight and threshold;
- [6] After experiencing M, judge whether the indicator meets the accuracy requirement, if not, then return [3], continue the iteration; if it is satisfied, go to the next step;
- [7] At the end of the training, save the weights and offsets in the file. At this time, it can be considered that the weights have been stabilized and the classifier has been formed.

3.2. Traditional Convolutional Network Model

The traditional network uses the fully connected layer to integrate features into one-dimensional vectors[7], but the parameters are more. This paper uses global average pooling[15] instead of the fully connected layer, adding Normal Normalization (BN)[16] after convolution makes the network easier to train and reduces training time.

Table 1. TraditionNet structure.

Type	Kernel size/stride/pad	Output size
Input	---	227*227*3
Conv1/BN	7*7/3/0	74*74*64
MaxPool1	3*3/2/0	37*37*64
Conv2/BN	3*3/1/1	37*37*64
Conv3/BN	3*3/1/1	37*37*64
MaxPool2	3*3/2/0	18*18*64
Conv4/BN	3*3/1/1	18*18*64
Conv5/BN	3*3/1/1	18*18*64
Max/Pool3	3*3/2/0	9*9*64
Conv6/BN	3*3/1/1	9*9*64
MaxPool4	3*3/2/0	4*4*64
Conv7	3*3/1/1	4*4*43
MaxPool5	3*3/2/0	2*2*43
AVEPool6	3*3/2/0	1*1*43

3.3. Improved Convolutional Network Model

This paper designs a feature extraction module (TS-Module) for traffic signs instead of standard convolution. The traditional convolution adopts a single 3*3 convolution check for image convolution, and the extracted features are not sufficient[17]. This design uses concatenated three sets of convolution kernels to perform convolution extraction on the input image. In the module, 1*3 and 3*1 can extract the horizontal and vertical features of the image separately, and then integrate the feature information of 1*3 and 3*1 by 1*1, which can not only achieve the same receptive field of 3*3, widening The network structure is deepened, and 3*3 is to extract the main feature information of the image. By connecting 3*3 and 1*1, more features can be extracted, and the parameters can be reduced accordingly, so that the model is smaller, and the module is as figure 1.

Explain the TS-Module module of this article with an example of inputting 64 feature maps. Input 64 feature maps into three branches a, b, and c respectively (for convenience of explanation, 3*3 is called a branch, 1*3 is called b branch, and 3*1 is called c branch road), In order to ensure the image feature extraction is sufficient, this paper sets the number of a branch output channels to 24, the number of b branch output channels to 20, and the number of c branch output channels to 20, a, b, c three branches for independent convolution, The feature can be extracted more diversely, and then the feature information of the b and c branches is integrated by 1*1 to obtain a new feature map, and the new feature map and the feature map extracted by the channel are connected on the channel to obtain The output of the TS-Module module. The TS-Module module parameter settings are shown in Table 2.

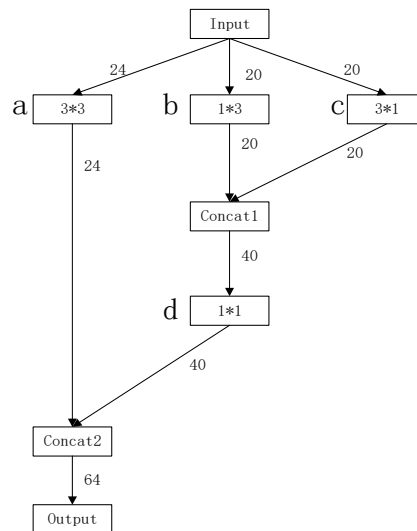


Figure 1. TS-Module module.

Using the MyNet network structure of the TS-Module module of this paper, as shown in Table 3, the overall framework remains basically unchanged, replacing the original five convolutional layers with the TS-Module module. As the network deepens, in order to avoid the gradient disappearing, a residual is added between the output of MaxPool2 and the input of MaxPool3.

Table 2. TS-Module module parameters.

Layers	Kernel size/stride/number	Pad/Pad_h/Pad_w
a	3*3/1/24	Pad=1
b	1*3/1/20	Pad w=1
c	3*1/1/20	Pad h=1
Concat1	b&c	---
d	1*1/1/40	Pad=0
Concat2	a&d	---

Table 3. MyNet structure.

Type	Kernel size/stride/pad	Output size
Input	---	227*227*3
Conv1/BN	7*7/3/0	74*74*64
MaxPool1	3*3/2/0	37*37*64
TS-Module1	3*3/1/1	37*37*64
TS-Module2	3*3/1/1	37*37*64
MaxPool2	3*3/2/0	18*18*64
TS-Module3	3*3/1/1	18*18*64
TS-Module4	3*3/1/1	18*18*64
Max/Pool3	3*3/2/0	9*9*64
TS-Module5	3*3/1/1	9*9*64
MaxPool4	3*3/2/0	4*4*64
Conv7	3*3/1/1	4*4*43
MaxPool5	3*3/2/0	2*2*43
AVEPool6	3*3/2/0	1*1*43

4. Experiment

4.1. Image Preprocessing

Data enhancement is performed on the dataset image before training, normalization and de-averaging operations are performed, all images are converted into 256*256 size, and the average value of the training image is subtracted for each pixel of the image, and all the images are respectively The five orientations (upper left, lower left and lower right and middle) are randomly tailored to 227*227 size to enhance the data set.

4.2. Experimental setup

Experimental environment: This experiment uses Caffe as the underlying framework for deep learning. The computer uses 32GB RAM, i7-6700K quad-core and eight-thread CPU and NVIDIA-GTX1070Ti GPU. The operating system is Ubuntu16.04.

Experimental data set: This experiment used the German Traffic Sign Recognition Benchmark (GTSRB) and the Belgium Traffic Sign Dataset (BTSD). Both data sets included traffic signs under adverse conditions, as shown in Figure 2 and Figure 3. The GTSRB dataset includes 43 types of German traffic signs, with a total of 51831 pictures, of which 39209 are in the training set and 12630 in the test set. The Belgian traffic signs data set includes 62 Belgian traffic signs, for a total of 7093 pictures. There are 4575 training sets and 2518 test sets.

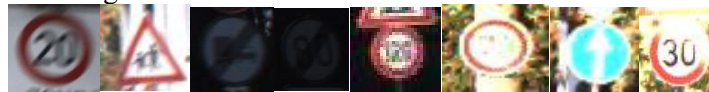


Figure 2. Traffic signs under adverse conditions in the GTSRB.



Figure 3. Traffic signs under adverse conditions in the BTSD.

Solver file configuration: When the experiment is trained on the GTSRB data set and the BTSD data set, The experiment uses the learning rate change mode as multistep, gamma is 0.1, and the network performance is tested once every 500 times, initial learning rate is 0.001, step value is 24000 and 48000, and the maximum number of iterations is set to 60000 times.

4.3. Experimental results and analysis

This experiment compares the recognition performance of the TraditionNet network and the MyNet network in the GTSRB and the BTSD data sets.

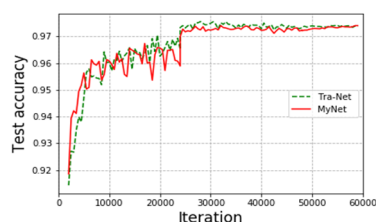


Figure 4. Accuracy curve of different CNN models on GTSRB.

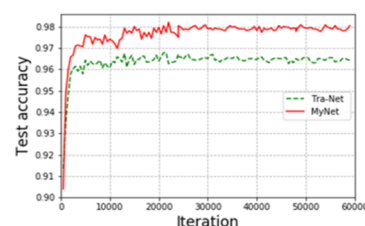


Figure 5. Accuracy curve of different CNN models on BTSD.

Figure 4 shows the accuracy curves of the two models on the GTSRB data set. The accuracy of GTSRB on Tra-Net can reach 97.35%, which is already very high accuracy. MyNet has a slightly higher accuracy than Tra-Net.

Table 4. Performance of each model on GTSRB and BTSD

Data sets	Network model	Accuracy (%)	Caffemode size(KB)	Forward Operation time(ms)
GTSRB	Tra-Net	97.35	884.6	692.34
GTSRB	MyNet	97.4	620.3	700.67
BTSD	Tra-Net	96.4	928.5	694.42
BTSD	MyNet	98.1	664.2	705.10

From Table 4, the network model is reduced by 264.3 KB. MyNet can achieve a smaller model while having higher accuracy, making the model more portable. Figure 5 shows the accuracy curves of the two models on the Belgian dataset. When the preset maximum number of iterations is 60000, MyNet's accuracy rate is 1.5 percentage points higher than the traditional convolution network, the accuracy rate is 98.1%, and the model is reduced by 264.3KB.

Select the GTSRB data set for comparison experiments, compare the CNN of the TS-module module with the recent classic CNN, as shown in Table 5, the CNN of the TS-module module can not only obtain a lightweight model, but also maintain high accuracy.

Table 5. Performance of each model on the GTSRB.

Network model	Accuracy (%)	Caffemode size (MB)
LeNet-5[18]	87	---
SqueeNet	94.2	3.0
AlexNet	95.9	233.8
CNN[18]	95	---
LBP[19]	96.3	---
GoogLeNet	96.5	41.8
Tra-net	97.35	0.864
MyNet	97.4	0.606

5. Conclusion

This paper designs a TS-Module module CNN for traffic signs recognition. Through several experiments in the traffic signs dataset, it is proved that the CNN of the module is more suitable for traffic signs recognition than the traditional CNN, because the traditional convolution is through a single The 3*3 convolution kernel performs feature extraction, which has the disadvantage of insufficient feature extraction, and generates a large number of parameters. As the network deepens, the gradient disappears, resulting in poor network performance. For the TS-Module module CNN, the TS-Module module uses multiple sets of convolutions. The 1*3 convolution kernel and the 3*1 convolution kernel are connected through a 1*1 convolution kernel to achieve a 3*3 convolution kernel. The experience field, effectively deepen and widen the network, while multi-channel convolution can not only extract a variety of information, but also greatly reduce the parameters, the practicality of the network is enhanced, the CNN also quotes the residual, If you want to design a deeper network, you can add residuals, which can effectively avoid gradient disappearance, improve accuracy, and lightweight models. And the network design base the TS-module can be more widely applied to mobile terminals. Combined with the above, this makes the convolutional neural network using the module better applicable to traffic signs recognition.

References

- [1] Suisui Tang. (2014) Research on Traffic Sign Recognition Algorithm. Beijing Jiaotong University, PP. 1-7.
- [2] Varan S, Singh S, Kunte R S, et al. (2007) A Road Traffic Signal Recognition System Based on Template Matching Employing Tree Classifier. In: Proc of the International

- Conference on Computational Intelligence & Multimedia. Sivakasi, Tamil Nadu, India. pp. 360-365.
- [3] Soendoro D, Supriana I. (2011) Traffic Sign Recognition with Color-based Method, Shape-arc Estimation and SVM. In: International Conference on Electrical Engineering & Informatics. IEEE. Bandung, Indonesia.
 - [4] Wahyono W, Jo K H. (2014) A comparative study of classification methods for traffic signs recognition. In: IEEE International Conference on Industrial Technology. Bhubaneswar, India, Busan. PP. 614-619.
 - [5] Yang Wang. (2013) A Traffic Sign Recognition Method based on Template Matching. Jilin University, pp. 1-8.
 - [6] Zhu Shuang-dong, Lu Xiao-feng. (2006) A Survey of the Research on Traffic Sign Recognition. Computer Engineering & Science, 28: 50-52+102.
 - [7] HUANG Lin, ZHANG Yousai. (2015) Traffic signs recognition applying with deep-layer convolution neural network. Modern Electronics Technique, 38: 101-106.
 - [8] QIAO Kun, GU Han-zhou, LIU Jia-ming. (2017) Traffic Sign Recognition Based On Deep Learning. Information Technology and Informatization, 12: 25-31.
 - [9] Li Zheng. (2011) Research on method of traffic signs recognition based on support vector machine, Yanshan University.
 - [10] Wenbin Fu. (2018) Real-time detection and recognition of road traffic signs using MSER and random forests. Jiangxi University of Science and Technology.
 - [11] K. Simonyan and A. Zisserman. (2015) Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations. San Diego.
 - [12] Szegedy C, Vanhoucke V, Ioffe S, et al. (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas. PP. 2818-2826.
 - [13] He K, Zhang X, Ren S, et al. (2016) Deep Residual Learning for Image Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas. PP. 770-778.
 - [14] Szegedy C, Ioffe S, Vanhoucke V, et al. (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: the Association for the Advance of Artificial Intelligence. San Francisco. PP. 4278-4284.
 - [15] Lin M, Chen Q, Yan S. (2014) Network In Network. Computer Science.
 - [16] Ioffe S, Szegedy C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: International Conference on Machine Learning. Lille. PP. 448-456.
 - [17] YANG Yuanfei, ZENG Shangyou, GAN Xiaonan, et al. (2017) Application of improved depth convolution network in traffic signs recognition. Television technology, 41: 214-219.
 - [18] XU Bin-sen, WEI Yuan-zhou, et al. (2017) Research and Implementation of Traffic Sign Recognition Algorithm. Computer engineering & Software, 38: 74-81.
 - [19] HAN Wei, LIU Lei. (2017) Traffic Sign Recognition Method Based on Polar Coordinate Partition LBP. Natural Science Journal of Xiangtan University, 39: 92-94.