Traffic Sign Recognition Based on Convolutional Neural Network Model

Zhilong, He
Qilu University of Technology
(Shandong Academy of Sciences)
Jinan, Shandong, China
hezhilong86@163.com

Zhongjun, Xiao*
Qilu University of Technology
(Shandong Academy of Sciences)
Jinan, Shandong, China
jinanxiaoxiao2012@163.com

Zhiguo, Yan
Qilu University of Technology
(Shandong Academy of Sciences)
Jinan, Shandong, China
yanzg500@sina.com

2020 Chinese Automation Congress (CAC) | 978-1-7281-7687-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/CAC51589.2020.9327830

***Abstract*—Traffic sign recognition (TSR) is a significance research branch in the field of unmanned driving, which is very important for driverless driving and is often used to read permanent or temporary road signs on the roadside. Traffic sign detection (TSD) and traffic sign classification (TSC) constitute a complete recognition system. The paper mainly studies the traffic sign recognition. Traffic sign recognition is mostly applied to portable devices, so the size and detection speed of the model are important factors to be considered. Under the condition of ensuring the speed, the detection accuracy of the model is guaranteed. The accuracy of the model designed in this paper on the German traffic sign recognition benchmark (GTSRB) is 99.30%, the parameter size is only 1.3M, and the trained network model is 4.0M. The results of final experiment show that the network is valid for the classification of traffic signs.**

Keywords—traffic sign; unmanned; portable devices; model;

I. INTRODUCTION

Traffic sign recognition in intelligent driving systems such as automatic driving and assisted driving plays an important role. Road sign recognition methods are split up two categories: manual feature methods and deep learning methods. In the past, traditional recognition methods required manual labeling and feature extraction, such as specific color recognition [10] and other feature recognition methods [11], which greatly reduced the speed of system operation. Manual labeling not only increased the workload, but also the accuracy rate was difficult to guarantee. Artificial feature learning methods generally use SVM and random forest, but this method is not easy to recognize for images with blurred feature boundaries [1].

For recent years, the rapid development of deep learning has changed detection method. The research of neural networks has gradually become a popular research field for researchers. The birth and rapid application of neural network can get rid of the laborious manual annotation, the constructed network can automatically extract the features of the input image. Especially for complex images, the network can get different features, and finally these features are used for target [2] classification. Thanks to the development of deep learning, the recognition of traffic signs is also developing rapidly. For example, Bouti [3] and others modified the LeNet network to achieve a good classification effect on the GSTRB dataset. Article [5] uses hinge loss function and loss of function with a network neural network to achieve the recognition of traffic signs. Wu *et al* [12] used the two branch networks to enhance the learning target classification capabilities, and the classification effect is better.

The main work of this article is as follows: Section II briefly describes the components of the convolutional neural network (CNN) and their role in the network. Section III gives the model structure proposed in this paper and relevant details. Section Ⅳ is the experimental part. The proposed model in this paper was tested on GTSRB, and analyzed and compared with other researchers' models. The final section summarizes the full text.

II. BRIEF REVIEW OF CNN

In recent years, CNN has become one of the research hotspots, which many scholars are devoted to this field [6], [7], [8]. Therefore, CNN has gradually become the most common image classification model in computer vision. Usually a complete CNN includes three basic parts: convolutional layer, pooling-layer and fully-connected layer. Convolutional layer is an important part of CNN. The convolution kernel convolves the corresponding region of the image with a specified step size and outputs a two-dimensional feature map [9]. The image develops from low-dimensional to high-dimensional, and then obtains high-dimensional features of the image. Compared with traditional machine learning methods, adding convolutional layers can automatically extract features at different levels in the image, and has translation invariance to the input image. In addition, the convolution kernel in the convolution layer is parameters of the shared, which greatly reduces the size of the parameters. The pooling layer in the convolution process can reduce the image dimension, retain the ability of key information, and speed up the network training process. Common pool methods include maximum pool, average pool and random pool. No matter which pooling method is used, its main purpose is to reduce the spatial feature dimension, reduce the system load, and speed up the network training speed. As the end of the neural network, there is usually one, or more fully-connected layer. Its function is to be extended to one-dimensional feature map, use the extracted high-dimensional feature information to classify the image, and use the last fully-connected layer as output layer, and then network outputs classification result. In addition, the classification activation function can convert the feature information of the image into the (0, 1) interval, which reduces the computer performance consumed during the training process.

III. METHODOLOGY

The target neural network constructed in this paper is trained on the training set to verify the recognition accuracy of the network on the validation set. According to the results on the validation set, the training is continued on the training set. Finally, the accuracy of the network on the test set is tested.

Zhongjun, Xiao is the corresponding author
(e-mail: jinanxiaoxiao2012@163.com).

This research was funded by the National Natural Science Foundation of China (No. 61877062) (in Chinese).

A. Data Enhancement and Processing

Fig. 1 shows the distribution of 43 categories GTSRB. The horizontal coordinate is 43 categories, and the vertical coordinate is the number of each category. We can clearly see that the distribution of the image dataset is uneven, which is easy to cause the network to classify certain categories (more data) accurately, while for other categories (less data) the classification effect is inaccurate, So this paper uses data enhancement methods to expand the dataset. The generalization ability of the network and the classification ability of different shooting angles are improved.

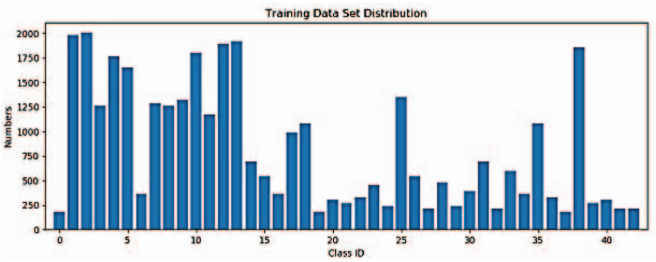

Fig. 1. Dataset distribution.

Algorithm enhancement uses imgaug [19], imgaug is a machine learning library for processing images. There are various enhancement methods, such as rotation, blur, grayscale, etc. Therefore, this paper uses imgaug to expand the GTSRB data and divide it into small batches for network training, which not only improves the generalization ability of the network, but also reduces computing load of computer.

Data augmentation is a commonly used image expansion method to improve network generalization ability. Therefore, this paper uses data enhancement [5], [8] to perform 50% image shading on the training set, random cropping and filling of specified pixels, and 50% image color conversion to increase the size of datasets and improve effectiveness.

The following is the specific algorithm flow:

Algorithm 1: Enhance image.

Input: Randomly select the input images to obtain small batches, *data*, *labels* for training;

Output: Enhanced data, *X_batch* and label, *y_batch*;

1: Use the imgaug library to draw out the methods that need to be enhanced, the image light and dark changes, random cropping and filling, and color transformation;

2: Using the method of generating random numbers, randomly select any one of the three image transformations to enhance the *X_batch* data;

3: Get the generated image *X_batch* and label *y_batch*;

4: **return** (*X_batch*, *y_batch*);

In addition to data enhancement, we also use data normalization processing, the specific method is to normalize the image pixel value from (0, 255) to (0, 1), image normalization processing helps network training to quickly converge and accelerate Network training speed. The following formula:

$$X_{inorm} = \frac{x_i - min(x)}{max(x) - min(x)} \tag{1}$$

Where min (x) is image pixel minimum, max (x) is image pixel maximum, x_i is pixel value of point i^{th}, and X_{inorm} is normalized value of pixels.

B. Network Structure

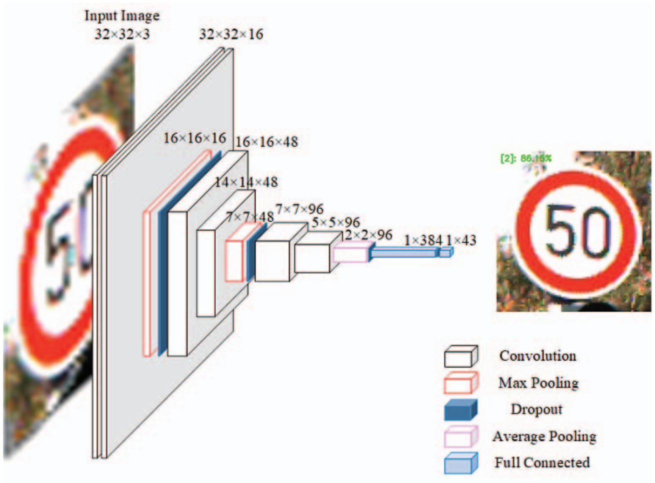

Fig. 2. Network model diagram

A total of 10 layers of convolutional neural networks are constructed in this paper, we call it TS-CNN, as shown in Fig. 2. The network is mainly composed of convolutional and pooling layer.

By gradually increasing the convolutional layer, extracting the feature map from the input image, using max-pooling to reduce the dimensionality of feature map, obtaining features at different feature scales by fusing more layers. Finally, the fully connected layer performs dimensional transformation on the input features, and uses soft extremum functions to classify the traffic signs.

C. Network Details

1) Dropout

We apply Dropout technology to the construction of the network [13]. In the process of forward propagation, this method randomly inactivates neurons with a certain probability P to reduce the scale of parameters, improve the generalization ability of the model. Fig. 3 is a before-and-after diagram.

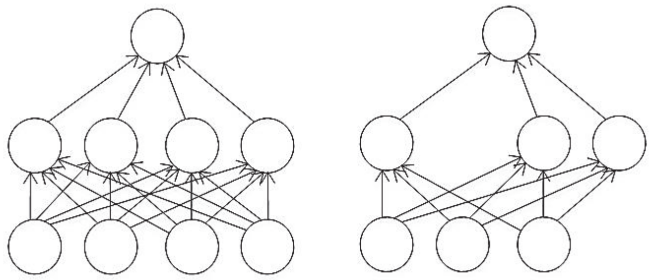

Fig. 3. Before and after using Dropout..

The left picture in the above picture is not using Dropout, the right picture is using Dropout, it can be clearly seen that the complexity of the network structure after using Dropout is reduced, which is helpful to improve the network training efficiency and generalization ability.

2) Activation Function

This article does not use the common ReLU function, but the ELU function [14]. This function combines the advantages of the ReLU and Soft-Max functions. The expression and figure of the function are as follows,

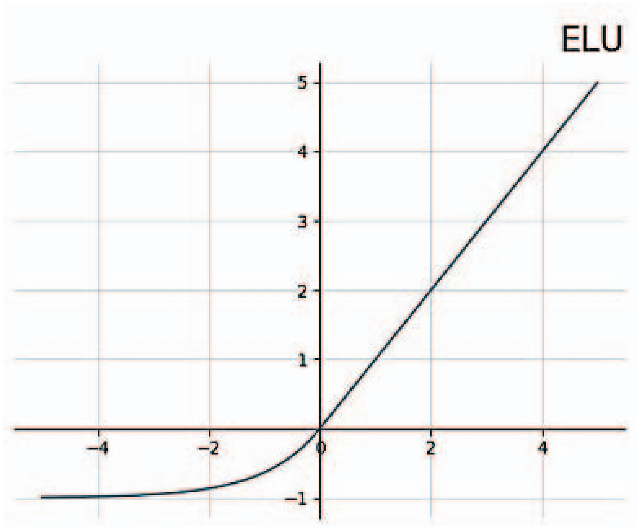

Fig. 4. ELU function.

$$f(\mathrm{t})=\begin{cases} t, & if\ \ t>0 \\ \lambda\left(e^{t}-1\right), & if\ \ \mathrm{t}\le 0 \end{cases} \tag{2}$$

Where, λ is a non-zero constant. When t > 0, the output equals the input, which guarantees linear growth of the function. Therefore, the gradient disappearance problem can be alleviated like the ReLU function. The left side has soft saturation characteristics and is more robust to changes in the input image, which is not available in the ReLU function. As shown in Fig. 4, here λ is -1.

IV. Experiments

A. Datasets and Setups

In this paper, the dataset uses the German traffic sign recognition benchmark (GTSRB) [4]. It was randomly obtained from the camera in a real-time scene, and was first provided by the International Joint Conference on Neural Networks (IJCNN) in 2011. The dataset includes 51839 images in 43 categories, of which training set images have 39,209 samples testing images have 12,630 samples, and each image resolution is dynamically changed from 15 × 15 to 250 × 250.

Fig. 5. Examples of datasets.

TABLE I. A List of 43 Categories

Name	Content
Mandatory	
Prohibitory	
Danger	
Other	

Fig. 5 is an example image of the dataset in GTSRB. The image with constantly changing scales increases the richness of the data and can improve the fitting effect of the network. TABLE Ⅰ is the type of all datasets, which contains 43 road sign categories.

The evaluation of the neural network is carried out on the following configuration, and the experiment is carried out according to the division of the GTSRB dataset.

1) CPU: Intel (R) I5-7300HQ, 2.50GHz
2) Memory: 20G DDR4
3) Disk capacity: 1T
4) Operating system: UBUNTU 18.04

B. Evaluation and Analysis

We trained the network and stopped the training when the loss of the training fluctuated stably, so iterated 100 rounds in total to get the trained model. Figure 6 below is the loss and accuracy images during the training process.

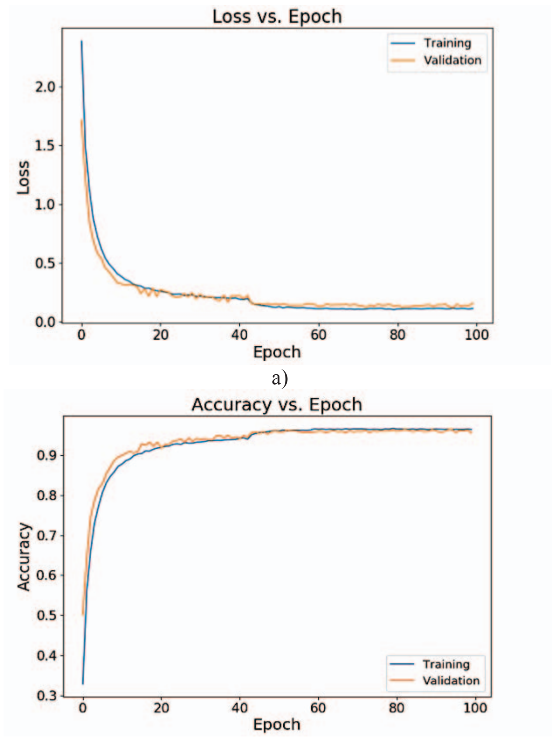

Fig. 6. Network iteration graph of loss and accuracy.

In Fig. 6, a) shows that the loss value gradually approaches zero, but the accuracy rate stabilizes at around 95.6 in b). We think this is a normal phenomenon, because the use of data enhancement, the image after each batch of data enhancement is different. In other words, the image data is not fixed every time, so the accuracy may not achieve a high effect, which also prevents the network from overfitting, thereby showing better effect on test dataset.

To verify the validity of the proposed, a variety of advanced (Compare with the state-of-the-arts) traffic sign classification methods were selected for comparison, as shown in TABLE Ⅱ and Ⅲ.

TABLE II. COMPARISON OF DIFFERENT METHODS IN SPEED.

Name	Method	Accuracy	Speed (ms/image)
Ciregan et al [20]	MCDNN	99.46%	11.4
Tang et al [16]	SVM + HOG + Gabor + LBP	98.65%	39.8
Li et al[17]	ELM+HOG+LBP	96.82%	2.97
The proposed	CNN	**99.30%**	**0.22**

TABLE III. COMPARISON OF DIFFERENT METHODS IN MODEL SIZE.

Name	Method	Accuracy	Model size
Zheng et al [15]	CNN+PCA+LVQ	94.62%	-
AlexNet [6]	CNN	95.90%	233.8MB
GoogLeNet [21]	CNN+ Inception	96.50%	41.8MB
Zhou et al [18]	ReduceV2Net	98.20%	6.3MB
The proposed	CNN	**99.30%**	**4.0MB**

In [16], a variety of features are fused, specifically HOG, Gabor filter features, and LBP features. Multi-feature complementarity is used to enhance the recognition effect, and finally SVM is used for classification. Literature [17] combines local features with ELM neural network to classify traffic signs, and the classification detection speed is faster. Literature [15] was combined CNN, principal component analysis (PCA), Learning Vector Quantization (LVQ), and used dimensionality reduction and classification processing to eliminate image interference information to classify traffic signs, resulting in 94.62% detection accuracy. But we believe that unnecessary processing links have been added. In [18], they divided detection and classification into two modules. The classification module inputs the detected traffic signs into the convolutional neural network for classification. The images used are grayscale images to speed up the recognition. The final classification accuracy is 97.75%. Similarly, the paper has been verified on AlexNet [6] and GoogLeNet [21], but the performance is not satisfactory, and both models are very large, unable to meet the mobile, small equipment requirements. The results show that this method has the potential to be applied to real - time systems.Although there is a certain gap with the recognition accuracy of MCDNN [20], it greatly reduces the detection time and complexity.

V. CONCLUSION

We propose a lightweight convolutional network suitable for traffic sign recognition classification in this paper. The network completes the recognition of traffic signs through simple convolution and pooling operations, theoretically guarantees the calculation efficiency of the algorithm, and is verified on the GTSRB data. Another advantage of this network is its processing time, which is faster than the detection speed of current algorithms, and it has a simple architecture and strong scalability. In future research, we consider recognizing traffic signs under severe weather and conducting experiments on more benchmark datasets. Of course, we also hope to apply this model to the detection of traffic signs.

REFERENCES

[1] Chaiyakhan, K., Hirunyawanakul, A., Chanklan, R., Kerdprasop, K. and Kerdprasop, N., 2015. Traffic Sign Classification using Support Vector Machine and Image Segmentation.

[2] Aghdam, H.H., Heravi, E.J. and Puig, D., 2016. A practical approach for detection and classification of traffic signs using convolutional neural networks. Robotics and autonomous systems, 84, pp.97-112.

[3] Bouti, A., Mahraz, M.A., Riffi, J. and Tairi, H., 2019. A robust system for road sign detection and classification using LeNet architecture based on convolutional neural network. Soft Computing, pp.1-13.

[4] Stallkamp, J., Schlipsing, M., Salmen, J. and Igel, C., 2011, July. The German traffic sign recognition benchmark: a multi-class classification competition. In The 2011 international joint conference on neural networks (pp. 1453-1460). IEEE.

[5] Jin, J., Fu, K. and Zhang, C., 2014. Traffic sign recognition with hinge loss trained convolutional neural networks. IEEE Transactions on Intelligent Transportation Systems, 15(5), pp.1991-2000.

[6] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[7] Zhang, J., Jin, X., Sun, J., Wang, J. and Li, K., 2019. Dual model learning combined with multiple feature selection for accurate visual tracking. IEEE Access, 7, pp.43956-43969.

[8] Zhou, Y., Shi, J., Yang, X., Wang, C., Wei, S. and Zhang, X., 2019. Rotational objects recognition and angle estimation via kernel-mapping CNN. IEEE Access, 7, pp.116505-116518.

[9] Luo, W., Li, Y., Urtasun, R. and Zemel, R., 2016. Understanding the effective receptive field in deep convolutional neural networks. In Advances in neural information processing systems (pp. 4898-4906).

[10] Berkaya, S.K., Gunduz, H., Ozsen, O., Akinlar, C. and Gunal, S., 2016. On circular traffic sign detection and recognition. Expert Systems with Applications, 48, pp.67-75.

[11] Yang, Y., Luo, H., Xu, H. and Wu, F., 2015. Towards real-time traffic sign detection and classification. IEEE Transactions on Intelligent transportation systems, 17(7), pp.2022-2031.

[12] Wu, Y., Liu, Y., Li, J., Liu, H. and Hu, X., 2013, August. Traffic sign detection based on convolutional neural networks. In The 2013 international joint conference on neural networks (IJCNN) (pp. 1-7). IEEE.

[13] Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.

[14] Clevert, D.A., Unterthiner, T. and Hochreiter, S., 2015. Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289.

[15] Zheng, Q. and Xie, X., 2018, May. Traffic Sign Recognition Based on Learning Vector Quantization and Convolution Neural Network. In Proceedings of the 3rd International Conference on Intelligent Information Processing (pp. 178-183).

[16] Tang, S. and Huang, L.L., 2013, November. Traffic sign recognition using complementary features. In 2013 2nd IAPR Asian Conference on Pattern Recognition (pp. 210-214). IEEE.

[17] Li, W., Chen, Q., Dong, T., Wei, L. and Zhang, Q., 2017, December. Traffic Signs Classification Based on Local Characteristics and ELM. In 2017 10th International Symposium on Computational Intelligence and Design (ISCID) (Vol. 1, pp. 127-130). IEEE.

[18] Zhou, Y., Feng, Y., Zeng, S. and Pan, B., 2019, May. Design of Lightweight Convolutional Neural Network Based on Dimensionality Reduction Module. In IOP Conference Series: Materials Science and Engineering (Vol. 533, No. 1, p. 012045). IOP Publishing.

[19] https://imgaug.readthedocs.io/en/latest/index.html#

[20] Ciregan, D., Meier, U. and Schmidhuber, J., 2012, June. Multi-column deep neural networks for image classification. In 2012 IEEE conference on computer vision and pattern recognition (pp. 3642-3649). IEEE.

[21] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9)