Accepted Manuscript

International Journal of Pattern Recognition and Artificial Intelligence

Article Title: Study on Traffic Sign Recognition by Optimized Lenet-5 Algorithm

Author(s): Chuanwei Zhang, Xiangyang Yue, Rui Wang, Niuniu Li, Yupeng Ding

DOI: 10.1142/S0218001420550034

Received: 02 November 2018

Accepted: 19 March 2019

To be cited as: Chuanwei Zhang *et al.*, Study on Traffic Sign Recognition by Optimized Lenet-5 Algorithm, *International Journal of Pattern Recognition and Artificial Intelligence*, doi: 10.1142/S0218001420550034

Link to final version: https://doi.org/10.1142/S0218001420550034

Study on Traffic Sign Recognition by Optimized Lenet-5 Algorithm

Chuanwei Zhang[1,2]

1 College of Mechanical Engineering, Xi'an University of Science and Technology

2 Shaanxi Key Laboratory of Integrated and Intelligent Navigation

Xi'an 710054, P.R. China

963923940@qq.com

Xiangyang Yue*

College of Mechanical Engineering, Xi'an University of Science and Technology

Xi'an 710054, P.R. China

yuexyxust@163.com

Rui Wang

College of Mechanical Engineering, Xi'an University of Science and Technology

Xi'an 710054, P.R. China

1214111936@qq.com

Niuniu Li

College of Mechanical Engineering, Xi'an University of Science and Technology

Xi'an 710054, P.R. China

809071374@qq.com

Yupeng Ding

College of Mechanical Engineering, Xi'an University of Science and Technology

Xi'an 710054, P.R. China

130623663@qq.com

Abstract: Traffic sign recognition is a key technology of intelligent vehicle, which is based on visual perception for road information. In view of the fact that the traditional computer vision identification technology cannot meet the requirements of real-time and accuracy, the traffic sign recognition algorithm has been proposed on the basis of improved Lenet-5. Firstly, we performed picture noise elimination and image enhancement on selected traffic sign images. Secondly, we used Gabor filter kernel in the convolution layer for convolution operation. In the convolution process, we added normalization layer BN after each convolution layer and reduced the data dimension. In the down-sampling layer, we replaced Sigmoid with the Relu activator. Finally, we selected the expanded GTSRB traffic sign database for the comparison experiment on the Caff platform. The experimental results showed that the proposed improved Lenet-5 network test set had the recognition accuracy of 96%, which was better than the method combined Gabor with Support Vector Machine (SVM) in terms of recognition accuracy and real-time performance.

Key words: Traffic sign recognition; Gabor filter; Lenet-5; SVM

1 Introduction

According to the statistics, there are about 50,000 traffic accidents every day in China on average. On average, more than 300 people die in traffic accidents every day. The economic toll from China's traffic congestion accounts for 20 percent of the urban population's income, reaching 250 billion RMB yuan a year. Annual automobile exhaust emissions reach 120 million tons, and a large amount of exhaust pollution results in haze weather that seriously affects people's health[1].

Therefore, the Intelligent Transportation System (ITS)[2] emerged in this context to solve practical traffic problems. ITS aim to enable various traffic users to be better informed and make safer use of transport networks. Considerable techniques have been proposed in ITS during the past years [3-5].

Traffic sign recognition is the key technology of intelligent vehicle for road environment information perception[6-7]. It involves static and dynamic traffic sign recognition, and includes the key technologies such as image acquisition, image preprocessing, image feature extraction, classification recognition. With the development of intelligent vehicle technology, the traffic sign recognition technology based on vision sensor is more and more popular by domestic and foreign scholars[8]. Traffic sign detection (TSD) and traffic sign recognition (TSR). The former focuses on the localisation of the target in a frame, while the latter performs a fine-grained classification to identify the type of the detected target[9].

At present, scholars at home and abroad have made some achievements in the research of traffic sign recognition. The main methods are as follows: template-matching traffic sign recognition method[10-11] makes the recognition by finding the similarity from two matching maps. The recognition accuracy is high. However, the two images must be of the same scale, and the sample images need to be stored. The large amount of computation can not guarantee the real-time requirement. BP neural network recognition method[12] has good fault-tolerance and learning ability. However, it needs to artificially determine the initial weights and thresholds of network nodes as well as the number of iterations. Therefore, it is easy to be overstudied. The recognition algorithm of Convolutional Neural Networks can recognize the rotation[13-14]. The accuracy of the images under different illumination conditions is high, but a large number of training samples are needed. Meanwhile, it needs to set network layers number and iterations number. But the training of the network model takes a long time, which can not guarantee the real-time performance. HOG[15] and Support Vector Machine (SVM) recognition algorithm[16] can recognize traffic signs accurately but needs to be improved in term of real-time.

In this paper, an improved Lenet-5 traffic sign recognition algorithm has been proposed. Firstly, the Gabor filter is used in convolution layer to extract the feature vector of traffic sign image and the Gabor wavelet is sensitive to the edge of the image, so it has good adaptability to the image under different illumination conditions. After each convolution layer, the normalized layer Batch Normalization(BN) is added to reduce the matrix dimension. Then the Relu activator is used instead of the Sigmoid activation function, and the support vector machine is used to classify the final classification layer. Through the contrast experiment of different recognition algorithms, it is proved that the algorithm in this paper has high real-time and accuracy.

2 Method

According to the color and shape features of traffic signs and the sign characters contained in each image, the image preprocessing for the sample image is carried out firstly to eliminate the influences of external factors and shooting angle distance, then the pre-processed image is normalized. The size of image is unified so as to facilitate the feature extraction and classification recognition. Gabor feature is adopted to extract effective traffic sign information, and BN normalization is used to deal with the setting of BN network parameters and training of CNN model. Finally, support vector machine classifier is used to output recognition information. The system process is shown in figure 1:

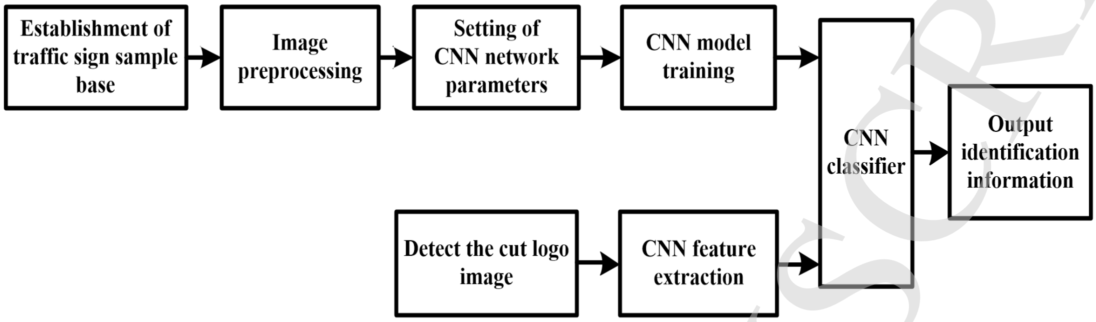


Figure 1. Overall flow chart of the system

3 Image detection

3.1 Detection of RGB and HSV

De La Escalera et al.[17] set a threshold segmentation method for R, G and B, but it is greatly affected by light and obviously cannot be detected for faded traffic signs. Ruta et al[18] proposed the RGB spatial color enhancement method to enhance the color brightness of traffic signs to achieve rapid and accurate identification. Fatin Zaklouta[19] improved color enhancement algorithm on the basis of Ruta et al., which can detect images that are less affected by light. Benallal M and Meunier J[20] found in the study of RGB color space that there are certain differences between the three color channels of RGB, and the differences can form two stable features in the sign detection, which have certain advantages for the detection of traffic signs.

The segmentation of color space based on RGB threshold is easily affected by light, and there is no fixed threshold range between the three components, so the threshold needs to be adjusted according to different lighting conditions. In this paper, the method of three-component difference is selected to effectively improve the influence of illumination. First, the pixel values of the three components of R, G and B after the color image histogram equalization are 0~255, and the gray scale images of the three components of R, G and B are extracted respectively, as shown in figure 2.

(a)Original image **(b)R component** **(c)G component** **(d) B component**

Figure 2. Three-component grayscale

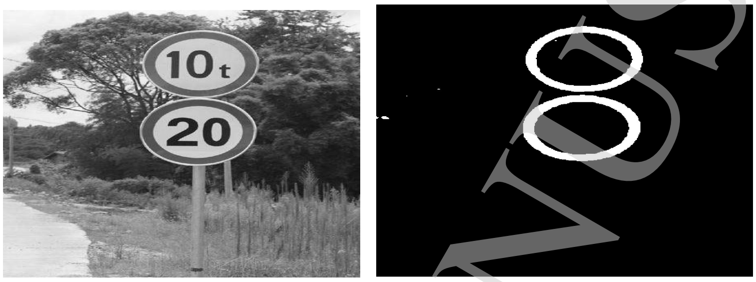
(1) feature operator extraction

The key of three-component color difference method is the selection of feature operator to separate the marker area from the background. For this method, it is similar about values among the gray-level values of R, G, and B. The feature operator is selected that meets the threshold segmentation requirements, and finally the segmentation mark image can be gained.

(2) threshold segmentation based on OTSU algorithm

For the image segmented by three-component color difference method, the gray scale diagram which has similar background to the logo color exists in the image. OTSU threshold segmentation calculation score is proposed to segment the binary image of similar background. OTSU algorithm divides the gray histogram of the image into two parts with the optimal threshold, getting the maximum variance between the two classifications, that is, the maximum separability.

The effect diagram of traffic sign image segmentation by three-component color difference method is shown in figure 3.



(a)The original image **(b)Segmentation image**

Figure 3. RGB three-component segmentation

HSV threshold segmentation algorithm is based on color space. Vitabile et al.[21] transformed RGB image into HSV color space and divided it into non-color region, unstable color region and color region. The dynamic pixel aggregation method was used to segment the color. In reference [22], the candidate region of traffic signs was segmented and located based on HSV in HSV color space. Chen Yixin et al.[23] used HSV space color extraction combined with shape discrimination and SVM to complete the logo detection work.

HSV color space to image threshold segmentation is less affected by light. Through a small amount of calculation, the RGB color space are turned into HSV space. Because brightness (V) value in HSV color space is set as a fixed value regardless of the influence of brightness, logo image segmentation depends on values of hue (H) and saturation (S). The steps of traffic sign detection based on HSV color segmentation are as follows:

(1) The sample image is converted from RGB color space to HSV color space. In the HSV color space, $H \in \left[0^0, 360^0\right]$, $S,V \in \left[0,1\right]$.

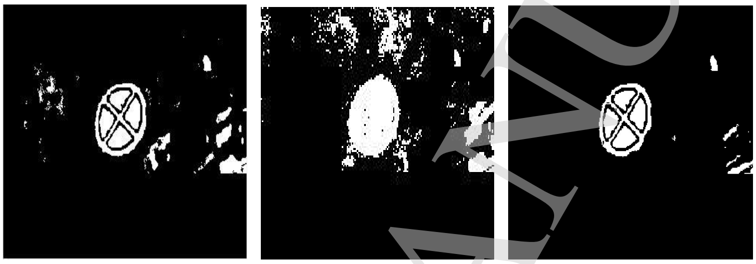
(2) H, S and V images are displayed respectively in the HSV color space and set as H values of yellow, red and blue. The mean and standard deviation of the obtained H and S matrices are directly calculated, and then the threshold range is determined.

(3) Color segmentation is conducted by comparing pixels and thresholds in the image to obtain binary images. The pixel points that meet the threshold interval are set as foreground white, while those that do not meet are set as background black. The segmentation threshold of HSV space is determined, and multiple traffic sign images are selected for verification, then the best H, S and V values are selected.

There is a small amount of noise spots. For color segmentation in HSV space, the segmented image has been influenced by many interference regions. It is disturbed seriously by the background during the blue sign segmentation. Hence, the combination of results from three-component color difference threshold segmentation algorithm and HSV color segmentation can effectively solve the shortages of these two algorithm and has high accuracy for binary image segmentation. Based on the segmentation results of the previous three-component color difference threshold segmentation algorithm, the HSV chromaticity space threshold segmentation results are directly fused to the preprocessed image color segmentation steps as follows.

RGB is converted to HSV chromaticity space to obtain H and S values. After contrast enhancement, the original image is segmented by three-component color difference method in RGB space to obtain the segmentation results. Based on the HSV threshold segmentation results and RGB segmentation results, the segmentation results were judged by superposition. Only the pixels within the threshold range of HSV threshold segmentation results and RGB segmentation results are marked with 1 as the traffic sign area: otherwise, the mark is 0 as the background.

The final mark image segmentation results of the superimposed color threshold segmentation algorithm are shown in figure 4.



(a)RGB segmentation (b) HSV segmentation (c)Fusion segmentation

Figure 4. Superimposed color segmentation

3.2 Traffic signs identification based on color and shape

The information of color image segmentation is strongly influenced by external interference in the binary image after segmentation, especially the billboards on both sides of streets. Most clothes and colors of pedestrians remain in interference fields of the binary image after segmentation, and the contour of the binary image is incomplete after the color segmentation of the logo image which is blocked by branches and damaged. The accurate shape detection can effectively filter out the interference field in the binary image, and the traffic signs that are occluded and damaged can also be effectively detected. At the same time the logo region that meets the color segmentation and shape detection algorithm is extracted and are input in the trained classifier for the classification identification.

4 Traffic sign recognition

4.1 Image preprocessing

(1)Image gray processing

The true color image grayscale processing can effectively reduce the influence of the outside environment, and also make use of the color particularity of the traffic sign. Through the grayscale processing, it is convenient to extract the feature by reducing the dimension of the image. The final gray value is obtained through the weighted mean of three color channel components of R,G and B.[24] Its weighting formula is as follow:

$$\text{Gray} = 0.11\text{B} + 0.59\text{G} + 0.3\text{R} \tag{1}$$

(2)Image enhancement

The histogram equalization is used to enhance the image[25], and the uniform distribution of the pixels of the gray image is changed to make the details of the image more clear so that the contrast of the image can be improved.

(3)Dimension normalization

The collected traffic sign images are easily affected by the shooting angle, shooting distance and other factors, which cause different sizes and seriously affect the feature extraction and classification recognition. The size of traffic sign image is adjusted to 36×36 by bilinear interpolation operation, then the image is used for feature extraction and classification recognition.

4.2 Improved Lenet-5 traffic sign recognition

Lenet-5 consists of 7 layers including convolution layer, downsampling layer, full connection layer and output layer. In the convolution layer, 5×5 templates are used to extract the filter eigenvector. The size of the downsampling layer window is 2×2, and the step size is 1. Input image 32×32 network structure parameters, characteristic number, number of neurons, and Lenet-5 traffic sign recognition is improved. Normalized layer BN is added after each convolution layer; Sigmoid is substitute for ReLU activator ; Support vector machines are used in the final classification layer; So the improved Lenet-5 structure is shown in figure 5.

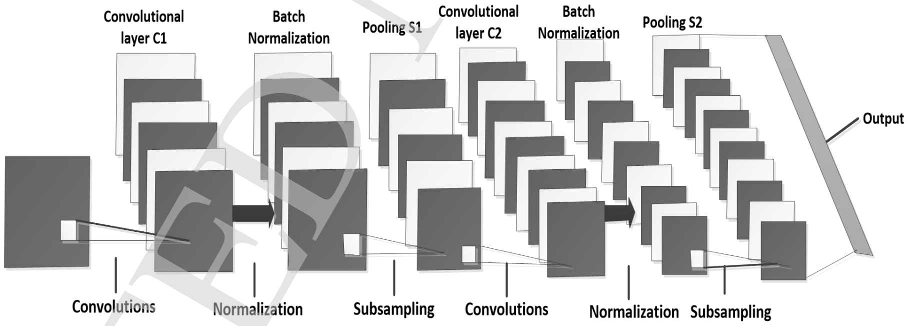


Figure 5. Improved Lenet-5 structure

(1)Convolution layer

In the convolution layer, Gabor convolution kernel is used for convolution operation, because Gabor transform has been widely used in the field of image processing and pattern recognition. On feature extraction, Gabor transform has good characteristics especially in local spatial and frequency domain information process. Gabor wavelet[26] is sensitive to the edge of the image, so it can provide good direction selection and scale selection characteristics, and is not sensitive to illumination changes so that it has good adaptability for the light changes. In airspace, two-dimensional Gabor filter is a band-pass filter, and its impulse response functions are as follows:

$$g(x,y,f,\theta)=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{k_1^2}{2\sigma_x^2}-\frac{k_2^2}{2\sigma_y^2}\right)\exp\left(i\left(f_x x+f_y y\right)\right) \tag{2}$$

$$k_1 = x\cos\theta + y\sin\theta \text{, } k_2 = -x\sin\theta + y\cos\theta \tag{3}$$

Where σ_x and σ_x are the standard deviations in the x and y directions respectively, $f_x = f \cdot \cos\theta$ and $f_y = f \cdot \sin\theta$ represent the frequency in space, f is the central frequency of the bandwidth, and θ is the spatial direction. When $\sigma_x = \sigma_y$, the formula is:

$$g(x,y,f,\theta) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\exp\left(i\left(f_x x + f_y y\right)\right) \tag{4}$$

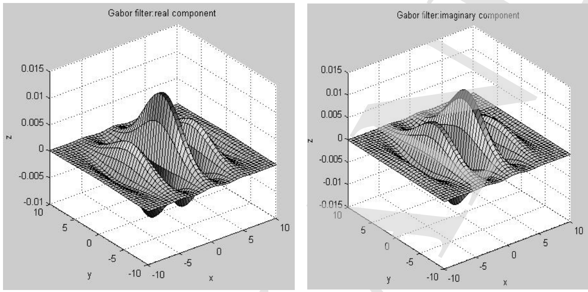
Calculation formula of its real part is written as:

$$g(x,y,\lambda,\theta,\varphi,\sigma,\gamma) = \exp\left(-\frac{(x'^2+y'^2)}{2\delta^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\varphi\right) \tag{5}$$

Calculation formula of its imaginary part is written as:

$$g(x,y,\lambda,\theta,\varphi,\sigma,\gamma) = \exp\left(-\frac{(x'^2+y'^2)}{2\delta^2}\right)\sin\left(2\pi\frac{x'}{\lambda}+\varphi\right) \tag{6}$$

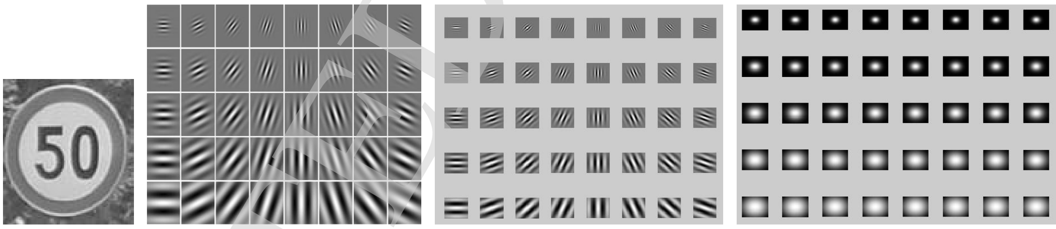
The real and imaginary parts of the Gabor filter are shown in figure 6.



(a)Real component **(b)Imaginary component**

Figure 6. The real and imaginary components of the Gabor filter

$x' = x\cos\theta + y\sin\theta$, where λ represents the wavelength of the positive wave, θ represents the direction, φ is the initial phase, and θ is the interval of $[0,\pi]$. The parameters of Gabor filter are set as follows: a filter bank consisting of 40 different filters with central frequency $v \in \{0,1,2,3,4\}$ and filter direction $U \in \{0,1,2,3,4,5,6,7\}$. Its real part response, imaginary part response and amplitude response are shown in figure 7.



(a)Real part response **(b)Imaginary part response** **(c)Amplitude response**

Figure 7. Real part response, imaginary part response, amplitude response graph

According to the principle of Gabor wavelet transform, the feature extraction is conducted for the traffic sign images of training and testing samples. The feature picture $I(x,y) \in R^{wh}$ that needs to be extracted (where w and h represent the width and height of image pixels in turn) convolves the Gabor filter in the 8-direction of 5 scale of the image to obtain the Gabor feature map of the image. In the experiment, amplitude is used as the feature vector of Gabor. According to the expression:

$$H(x,y) = I(x,y) * g(x,y,f,\theta) \tag{7}$$

$H(x,y)$ means the convolution characteristic graph, and the amplitude graph is divided into 4 x 4 overlapping regional blocks. The amplitude of the sampled regional blocks is used as the fast output feature vector of the region through the weighted average value, and the region block amplitude is concatenated as the output vector of Gabor filter, as shown in figure 8.

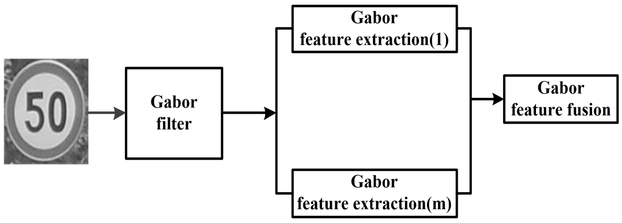


Figure 8. Gabor feature extraction

(2) Adding normalized layer BN

When the original image input convolution layer carries on the convolution training, the gradient descent method is mainly used to carry on the training. The parameter and learning rate of network must be set artificially before the entire program runs, so this process is quite tedious and the accuracy of recognition is greatly influenced by human being which cannot guarantee the real-time requirement. To solve the problem effectively, the normalized (BN) layer can increase the learning rate and speed up the convergence. The calculation formula of BN is as follow.

$$\widehat{x}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{Var[x^{(k)}]}} \tag{8}$$

$E(x^{(k)})$ is the average value of input data in each batch, and the denominator in the above formula is the standard deviation of input data in each batch. If the data is normalized directly by using the above formula, the expression capacity of the layer will be reduced. If the activation function is used, two learnable parameters γ, β must be introduced. The final output is:

$$y^{(k)} = \gamma^{(k)} \widehat{x}^{(k)} + \beta^{(k)} \tag{9}$$

(3)Improved sampling layer

The result of down-sampling is that the features and parameters are reduced, which is not the only one purpose of down-sampling. The purpose of down-sampling is to keep some invariance (rotation, translation, stretching, etc.). The methods of average sampling and maximum sampling are commonly used. Mean-pooling means that the average values of adjacent points are taken；max-pooling means the maximum value of adjacent points. Maximum Pool: the maximum value from adjacent points is selected as output value. Mean pool: the output value is an average of the coordinates of adjacent points. Pool operation is a way to reduce data dimension and accelerate network iteration.[27] Two different convolution pool operations are shown in figure 9.

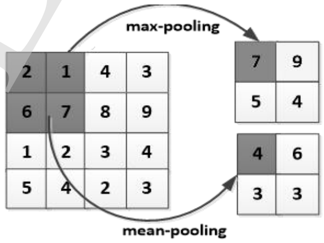
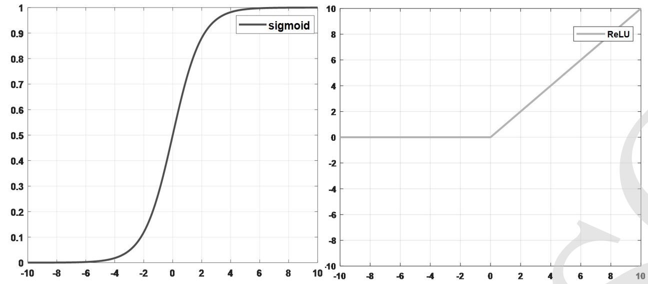


Figure 9. Pool diagram

We take the maximum value from adjacent points to prevent non-maximum value, which not only reduces the computational complexity and shortens the operation time, but also keeps the feature invariance (rotation, translation, translation, etc.) of the object. Scaling can effectively reduce the estimated average deviation caused by parameter error of convolution layer and retain texture information.

(4) The activation function selects ReLU

The most commonly used activation functions in convolution neural networks such as the most original linear function, Sigmoid activation function and the ReLU function, are the core of the research on convolutional neural network and an important operating algorithm of the excitation target feature, as shown in figure 10：



(a)Sigmoid function **(b)ReLU function**

Figure 10. Activation function

By using the Sigmoid function, we can see that its range of values is 0 to 1. In the whole calculation process of forward and backward propagation, due to the high calculation complexity, the ReLU activation function is used. Its minimum value range is 0 and it is an operation process that the value increases progressively. Sigmoid function is easy to saturate when calculating the backpropagation of parameters. As a result, the gradient descent rate is slow and the convergence rate slows down. ReLU can alleviate the problem of overfitting. Compared with the training time of Sigmoid function, ReLU obviously has advantages so it is chosen as the activation function[28]. The recognition rates of the different activation functions GTSRB library are shown in Table 1：

Table 1 Recognition rates of activation function GTSRB library

Activation function	Training time Iterative 150 times	Training accuracy	Test accuracy
Sigmoid	150ms	97.5%	89.3%
ReLU	96ms	95.8%	91.5%

From the experimental results of the two types of activation functions in the above table, we can see that the training time of the Sigmoid function is longer than that of the ReLU activation function at the same iteration number 150 times, although the ReLU function does not reach the expected value in the training accuracy. However, the test accuracy of the ReLU activation function is obviously higher than that of the Sigmoid activation function, and the ReLU function avoids the over-fitting phenomenon. Therefore, ReLU is chosen as the network activation function.

(5) SVM classifier replaces Softmax classifier

SVM is a new machine learning method based on statistical learning. It is based on the principle of structural risk minimization and minimizes the risk of machine learning by selecting certain discriminant functions. A classifier with minimum error is obtained by selecting the training sample and the test sample to search an approximate estimation of the relation between the input variable and the output variable. Therefore, the result of the output variable can be predicted with a higher accuracy rate [31-32].

Classification of traffic sign sample image should set selected sample training set as $T=\{(x_1,y_1),...,(x_l,y_l)\}\in(X\times Y)^L$.

$x_i\in X=R^n$； x_i is feature vector, $y_i\in Y=\{-1,1\}(i=1,2,...,l)$ defines the attribute label of the image. There are only two kinds of sample categories including belonging and not belonging to. In the feature space, the discriminant formula of the sample image through the interval maximization is:

$$w*x_i+b>0\text{, if } y_i=+1 \tag{10}$$

$$w * x_i + b < 0 \text{, if } y_i = -1 \tag{11}$$

where w is the normal vector, and b is the intercept. According to the above formulas and the feature of support vector machine in the feature space, the linear classification function formula is obtained as follow:

$$f(x) = \text{sgn}(w \cdot x + b) \tag{12}$$

The maximum geometric classification gap is found by the best hyperplane. The formula is as follow:

$$m\arg in = \frac{2}{\|w\|} \tag{13}$$

According to the formula of maximum gap calculation, we can know that W represents the characteristic normal vector of input sample. The classification distance of hyperplane is determined by linear classification function and the interval between M1 and margin in space. The hyperplane spacing affects the error classification times of support vector machine, and the error rate formula is as follow:

$$\eta \le \left(\frac{2R}{\delta}\right)^2 \tag{14}$$

W represents the feature normal vector of the input samples, and the optimal classification hyperplane is shown in figure 11 below.

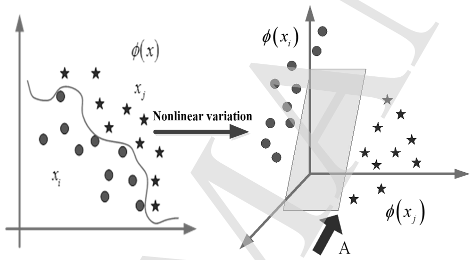


Figure 11. Support vector machine diagram

To facilitate the processing and analysis, Lagrange multiplier is introduced:

$$L(w,b,a) = \frac{\|w\|^2}{2} - \sum_{i=1}^{n} a_i \left[y_i(wx_i + b) - 1 \right] \tag{15}$$

According to the above formulas, the classification formula of X can be obtained:

$$f(x) = \text{sgn}(wx + b) = \text{sgn}\left(\sum_{i=1}^{n} y_i a_i (x, x_i) + b \right) \tag{16}$$

The improved Lenet-5 network layer, convolution nucleus size, number of neurons, and trainable parameters are shown in Table 2.

Table 2 Improvement of Lenet-5 network parameters

Type	convolution kernel	Feature map	Neure quantity	Trainable parameter
C1	5×5	28×28	4704	156
S1	2×2	14×14	1176	12
C2	5×5	10×10	1600	1516
S2	2×2	5×5	400	32
F6		1×1	320	10164
		1×43	43	

5 Experiment and result analysis

5.1 Experimental environment

The image acquisition device is Point Grey gray point camera with 13 million pixels and frame rate 85fps. Ethernet communication CMOS camera, pixel is 1280×1024. To reduce the amount of calculation, the picture has been processed appropriately. The image pixel to be recognized is 640×480. The TCP/IP protocol is used as the communication protocol of the server and the host computer PC client. The video transmission part is mainly connected by the Server_thread through the Socket communication based on the TCP protocol and the upper computer PC client. The image video processing platform is Asus desktop computer. The controller is AMD Ryzen7 1700, the frequency is 3.0 GHz, the graphics card is GTX1080Ti, 120GB solid state hard disk, the memory size is 16G. The experimental platform is shown in figures 12 and 13.

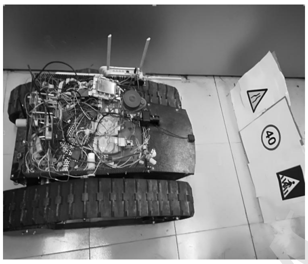


Figure 12. Intelligent car

Figure 13. Identification platform

5.2 Establishment of sample database

To obtain sample data, this paper collected a large number of traffic signs videos in all kinds of weather and state in the natural environment. Those videos are decomposed into images[29], and then, the traffic signs with different sizes, angles and degrees are selected as the positive sample for classifier training. At the same time, the image color and shape similar to the symbol interference objects are extracted as the training negative sample. In addition, in order to enrich the type and quantity of samples, some traffic signs in complex environment are adopted as supplements.

5.3 Image detection

For the routine traffic sign detection model, it has some limitations in dealing with the problems such as complex environment. This paper adopted fused segmentation method based on RGB three-component color difference method, adaptive threshold method and HSV color threshold segmentation method. It has been applied to the color segmentation of traffic signs. The collected traffic sign samples were divided into three categories including red, yellow and blue according to the color features, and the image enhancement processing was executed to the sample pictures of the training set. The pre-processed image directly uses the difference on the RGB three primary color space; Secondly, the image was transformed from RGB space to HSV space for expression. Certain color thresholds were given for components H and S, and HSV space threshold segmentation was realized by combining with OTSU threshold segmentation algorithm. At last, the segmentation results of RGB and HSV color space threshold were combined to obtain the color discrimination model of traffic signs in this paper. Then, shape analysis and morphological processing were performed on the binary images output by model discrimination. Finally, the location information of the binary detection results of the signs was marked in the original image. Hence, the experimental result shows that the proposed model has relatively strong robustness for the disturbance from the varieties of complex environments such as light and weather.

5.4 Localization of traffic signs

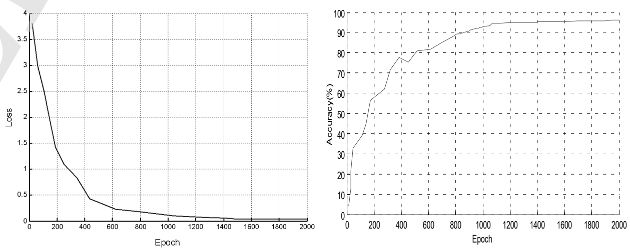
The image is preprocessed to remove the interference factors in the image and judge the shape of the image. After the traffic signs in the image are identified by detecting the shape of the model, the segmented traffic signs should be marked in the original image, that is, the positioning of traffic signs. In this paper, MBR (minimum bounding rectangle) is used to mark the position of the foreground target, that is, the maximum and minimum values of the ordinates of the connected region in the binary image are respectively scanned as the vertex of the minimum external rectangle of the foreground target. When marking in the original image, a certain amount of deviation should be added for correction, and the traffic signs are positioned with different colors and shapes (circular, octagon, triangle, rectangle).

(a)Circular positioning **(b)Octagon positioning**

Figure 14. Location of traffic signs

5.5 Improved Lenet-5 traffic sign recognition experiment

In the process of improving Lenet-5 network training, by drawing Accuracy and Loss curves, the dynamic process of improved Lenet-5 training can be observed intuitively and clearly. Figure 15 shows the corresponding Accuracy and Loss curves of the improved Lenet-5 network during the training process.



(a)Accuracy curve **(b)Loss curve**

Figure 15. Accuracy and Loss curves trained

From figure 15, we can see that when the number of iterations of the network is 1400-2000 Epoch, the training of the network is stable. After the improved Lenet-5 network training is completed, the test samples are input to test the model. The trained network model is matched with the test samples, and three different recognition algorithms are compared and analyzed. It is known that the improved Lenet-5 network has two advantages of higher recognition accuracy and faster recognition. Therefore, the improved Lenet-5 is used to recognize traffic signs in this paper. In this section, the trained and improved Lenet-5 convolutional neural network is used to identify 1,886 images, of which 1,789 images are correctly classified, and the classification accuracy reaches 95.77%. 79 images cannot be classified correctly, and the classification results are shown in figure 16.

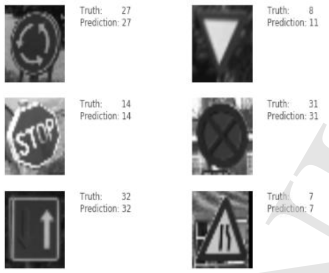


Figure 16. Test sample classification results

5.6 Comparison and Analysis of algorithms

In order to test the effectiveness of the improved Lenet-5 traffic sign recognition algorithm proposed in this paper, Gabor combined with SVM recognition algorithm is selected to compare with Lenet-5 convolution neural network and improved Lenet-5 convolution neural network algorithm . These three kinds of recognition algorithms are conducted classifier training by the extended GTSRB training set. The test data are 12630 images of GTSRB, and the recognition accuracy of 45 kinds of traffic signs is counted. As shown in the following figure.

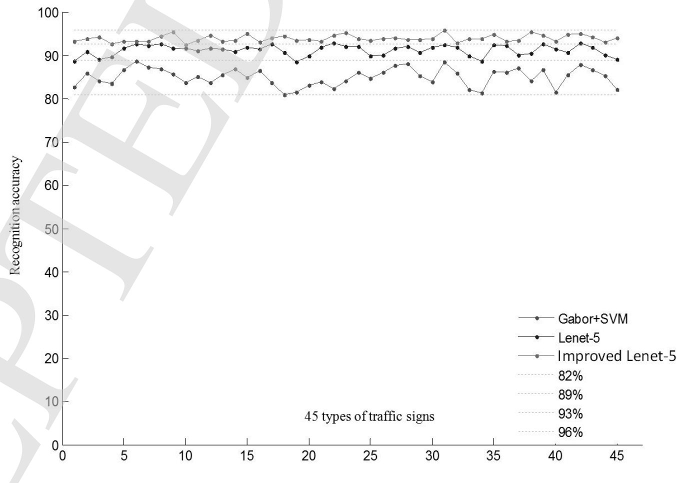


Figure 17. Comparison of traffic sign recognition rates

According to figure 17, the classification accuracy of SVM classifier[30] is between 82% and 89%. Lenet-5 convolution neural network is used to identify traffic signs. The accuracy of the method is obviously higher than that of Gabor SVM. The accuracy range of the method is between 88% and 93%, although the traffic sign recognition based on Lenet-5 has reached a relatively high accuracy.

However, compared with the improved Lenet-5 convolution neural network, the accuracy of Lenet-5 recognition is obviously lower. According to the broken line diagram, the improved Lenet-5 is the basis of recognition. The accuracy of Lenet-5 convolution neural network for the recognition of 45 kinds of traffic signs is between 93% and 96% on the test set. Overall, the improved Lenet-5 convolution neural network can be adopted to achieve a higher recognition rate of traffic signs.

The different algorithms are evaluated by recognition accuracy, and the real-time performance is analyzed by the running time of the algorithm. On the same experimental platform, 32110 expanded training pictures are used to train the above three recognition algorithms. Table 3 shows the training time of the three classification algorithms. After the three classifiers are trained, twenty RGB images of 32×32 size are recognized and all the mean recognition time is obtained. Table 4 shows the comparison of the average recognition run time of the three classification algorithms:

Table 3 Comparative analysis of training time of different classification algorithms

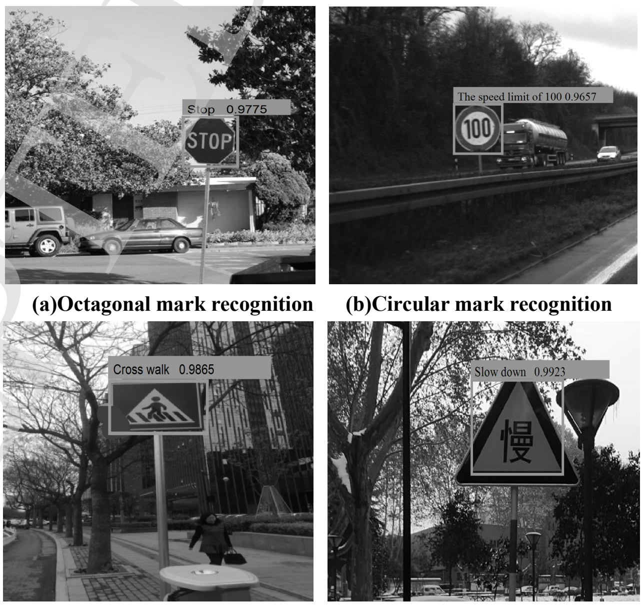
Sort algorithm	Improved Lenet-5	Lenet-5	Gabor+SVM
Training time (hours)	0.68	0.92	1.12

Table 4 Comparative analysis of recognition time of different classification algorithms

Sort algorithm	Improved Lenet-5	Lenet-5	Gabor+SVM
Identification time(seconds/sheet)	0.19	0.21	5.49

According to Table 3, the corresponding training time of three algorithms of improved Lenet-5, Gabor+SVM and convolutional neural network was 0.68, 0.92 and 1.12 hours respectively. And according to Table 4, it takes 5.49 seconds to identify an image by using Gabor SVM algorithm, which is the longest of the three algorithms. It takes 0.21 seconds to identify an image with Lenet-5, which is 0.02 seconds more than the algorithm we proposed. This is because the improved Lenet-5 network convolution layer uses BN normalization algorithm to speed up the operation of the network. In view of comprehensive recognition accuracy, the improved Lenet-5 network model has higher recognition accuracy for traffic signs and is more suitable for traffic sign recognition.

Finally, the real-time and accuracy of traffic sign recognition in Lenet-5 network are verified. The data set is annotated on GTSDB, and then the trained network model is effectively recognized in video by using the improved Lenet-5 convolution neural network training model. The recognition effect is shown in figure 18.



(a)Octagonal mark recognition (b)Circular mark recognition

(c)Rectangular mark recognition (d)Triangle mark recognition

Figure 18. Traffic sign recognition

6 Conclusion

In view of the problems faced by traffic sign recognition, we propose a traffic sign recognition method based on improved Lenet-5 network. This method is based on the Lenet-5 convolutional neural network model and make overall adjustments for the network. The internal relative correlation algorithm and parameters are optimized to get a new improved Lenet-5 network model. Through experiments with the method proposed by us, we find that the improved Lenet-5 is superior to the convolutional neural network and the classic Gabor+SVM classifier in accuracy and real-time performance. It improves the recognition accuracy and reduces the running time to realize the real-time performance. The system of traffic sign recognition has been realized and improved, and the expected target has been achieved.

7 Acknowledgement

This paper was supported by Shaanxi Key Laboratory of Integrated and Intelligent Navigation (SKLIIN-20180101), the authors would also like to express their deep gratitude for the support of the College of Mechanical Engineering, Xi'an University of Science and Technology.

Reference

1. Twelve major challenges to road traffic safety [J]. China Highway,2018(01):56.
2. Miurs J, Kanda T, Shirai Y. An active vision system for real-time traffic sign recognition[J].Intelligent Transportation Systems, IEEE, 2008:52-57.
3. Y. Yuan, Z. Xiong and Q. Wang. "An Incremental Framework for Video-Based Traffic Sign Detection, Tracking, and Recognition". in IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 7, pp. 1918-1929, July 2017.
4. Lituan Wang, Lei Zhang, Zhang Yi. "Trajectory Predictor by Using Recurrent Neural Networks in Visual Tracking". Cybernetics IEEE Transactions on, vol. 47, no. 10, pp. 3172-3183, 2017.
5. Mohsen Biglari, Ali Soleimani, Hamid Hassanpour. "A Cascaded Part-Based System for Fine-Grained Vehicle Classification". Intelligent Transportation Systems IEEE Transactions on, vol. 19, no. 1, pp. 273-283, 2018.
6. Mogelmose A, Trivedi M M, Moeslund T B. Vision-based traffic signs detection and analysis for intelligent driver assistance systems: Perspectives and survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(4): 1484-1497.
7. Chao Song, Weihua Chen. A method for automatic identification of traffic identifiers based on machine vision [J]. Journal of xiangfan vocational and technical college, 2011,10 (2) : 25-28.
8. Song Wenjie, Fu Mengyan, Yang yi. An efficient traffic sign recognition method for driverless cars [J]. Robot, 2015,37 (1) : 102-111.
9. Arcos-García, J.A. Álvarez-García, L.M. Soria-Morillo. Deep neural network for traffic sign recognition systems: an analysis of spatial transformers and stochastic optimisation methods. Neural Netw[J], 99 (2018), pp. 158-165.
10. Hao Bowen, Liang Yufeng, Li Wenqiang et al. Traffic sign detection method based on color space and template matching [J]. Intelligent computer and application, 2016, 6(4): 20-22.
11. Feng Chungui, Zhu Shiping, Wang Haijun, et al. Study on speed limit sign recognition based on improved template matching [J]. Journal of southwest university, 2013, 35(4): 167-172.
12. Yang Fei, Wang Kunming, Ma Xin et al. Application of BP neural network classifier to identify traffic signs [J]. Computer engineering, 2003, 29 (10) : 120-121.
13. SAHA S K, DULAL C M, BHUIYAN A A. Neural network based sign recognition [J]. International Journal of Computer Application, 2012, 50(10):3541-3546.
14. CIRESAN D, MEIER U, MASCI J, et al. Multicolumn deep neural network for traffic sign classification [J]. Neural Networks2012,32(2):333-338.
15. Huang Zhiyong, Yu Yuanlong, Gu J. An efficient method for traffic sign recognition based on extreme learning machine. IEEE Trans. Cybern. 2016, 47, 920–933.
16. Wang Yan, Mu Chunyang, Ma Xing. Traffic sign detection based on color standardization model and HOG feature [J]. Software guide, 2015,14 (3) : 142-144.
17. De La Escalera A, Moreno L E, Salichs M A, et al. Road traffic sign detection and classification[J]. Industrial Electronics, IEEE Transactions on, 1997, 44(6): 848-859.
18. Ruta, Yongmin Li, Xiaohui Liu. Real-time traffic sign recognition from video by class-specific discriminative features [J]. Pattern Recognition, 2010, 43(1): 416-430.
19. F.Zaklouta, B.Stanciulescu, Real-time traffic sign recognition in three stages[J]. Roboticsand Autonomous System, 2012, 32(8): 1-9.
20. Maldonado-Bascon S, Lafuente-Arroyo S, Gil-Jimenez P, et al. Road-sign detection and recognition based on

support vector machines[J]. IEEE Transactions on Intelligent Transportation Systems, 2007, 8(2):264-278.

21. Jin Xuhui. Detection and identification of traffic signs based on regional color segmentation [J]. Electrical automation, 2016, 03(38):14-16.

22. Vitabile S, Pollaccia G, Pilato G, et al. Road signs recognition using a dynamic pixel aggregation technique in the HSV color space[C]. International Conference on Image Analysis and Processing, 2001,572-577.

23. Chen L, Li Q,Li M, et al. Traffic sign detection and recognition for intelligent vehicle[C]. In Proceedings of IEEE International Conference on Intelligent Vehicles Symposium,IEEE Press,2011:908-913.

24. Song M L, Tao D P, Chen C, Li X L, Chen C W. Color to gray: visual cue preservation. IEEE Transactions on Pattern Analysis and Machine Intelligence,2010,32(9): 1537−1552.

25. Rafael CGonzalez. Digital image processing[M]. 2nd ed.bei-jing:Publishing House of Electronics Industry, 2006

26. Lin Kezheng, Xu Ying, Zhong Yuan. Using 2DGabor values and kernel fisher discriminant analysis for face recognition[C]//Proceed-ings of the 2nd International Conference on Information Science and Engineering, 2010：7624-7627.

27. Xiaojie Wang. Modular PCA based on Within-Class median for face recognition[C]//Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology, China, 2010：52-56.

28. Jianming Zhang, Qianqian Huang. Honglin Wu and Yukai Liu. A Shallow Network with Combined Pooling for Fast Traffic Sign Recognition. Information 2017, 8(2), 45.

29. He Yaoping, Xu Lizhen. Automatic identification algorithm of traffic signs in natural scenes [J]. Microcomputer information, 2010, 26 (4) : 161-163.

30. Chen Xing, Jia yinshan. Research on traffic sign recognition method based on support vector machine [J]. Science, technology and engineering, 2011, 11 (8) : 1751-1754.

31. X. Zhang, X. Song, L. Feng, L. Chen and M. Törngren. "A Case Study on Achieving Fair Data Age Distribution in Vehicular Communications". 2017 IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS), Pittsburgh, PA, 2017, pp. 307-318.

32. Samuele Salti, Alioscia Petrelli, Federico Tombari, Nicola Fioraio, Luigi Di Stefano. Traffic sign detection via interest region extraction[J]. Pattern Recognition, Volume 48, Issue 4, 2015.

Chuanwei Zhang received the Ph.D. degree from xi'an jiaotong university, is the dean, professor and tutor of xi'an university of science and technology. Chairman of the mining machinery branch of shaanxi mechanical engineering society, standing director of shaanxi automobile engineering society, expert member of China energy society, academic committee member of the university, head of the first-level discipline of mechanical engineering. Research direction: intelligent control of coal mine electromechanical system.

Rui Wang received the B.S. degree from Mechanical Design manufacture and Automation of xi'an university of science and technology in 2017, and is now studying in the graduate school of xi'an university of science and technology. Her research interests are traffic sign recognition based on Four-wheel independent drive electric vehicle.

Yupeng Ding reived the B.S. degree from Henan university of technology, Zhengzhou, China, in 2017. His research interests Wet brake vibration mechanism.

Xiangyang Yue received the B.S. degree from Mechanical Design manufacture and Automation of xi'an university of science and technology in 2017, and is now studying in the graduate school of xi'an university of science and technology. His research interests are deep learning and computer vision.

Niuniu Li received the B.S. degree from Mechanical Design manufacture and Automation of xi'an university of science and technology in 2017, and is now studying in the graduate school of xi'an university of science and technology. Her research interests are traffic sign recognition based on neural network.