

International Journal of Dynamical Systems and Differential Equations

ISSN online: 1752-3591 - ISSN print: 1752-3583

<https://www.inderscience.com/ijdsde>

Technical evaluation on improved Hough transform for monitoring traffic sign images under depth algorithm

Jie Ding, Guotao Zhao

DOI: [10.1504/IJDSDE.2025.10068898](https://doi.org/10.1504/IJDSDE.2025.10068898)

Article History:

Received:	11 May 2024
Last revised:	06 September 2024
Accepted:	02 December 2024
Published online:	27 June 2025

Technical evaluation on improved Hough transform for monitoring traffic sign images under depth algorithm

Jie Ding

College of Technology,
Hubei Engineering University,
432000, Hubei, China
Email: dingjie@hbeu.edu.cn

Guotao Zhao*

School of Foreign Languages,
Hubei Engineering University,
432000, Hubei, China
Email: Zhaogt0712@126.com

*Corresponding author

Abstract: The experiment drew the accuracy curve, training loss curve, and test loss curve of different traffic sign sizes and the network model after feature fusion, and compared the algorithm in this paper with the artificial neural network (ANN) algorithm and the random forest algorithm. The research results showed that the classification accuracy of the multi-scale model combining the three branch networks was as high as 99%. The classification accuracy of the extreme learning machine (ELM) classifier was better than that of the softmax and support vector machine (SVM) classifiers. Compared with the ANN algorithm and the random forest algorithm, the classification time of the algorithm in this paper was reduced by 125 ms and 155 ms, respectively. The classification accuracy reached 99%, which lays the foundation for the innovation and development of traffic sign image processing technology.

Keywords: traffic sign image; improved Hough transform; deep algorithm; multi-scale feature fusion; convolutional neural network.

Reference to this paper should be made as follows: Ding, J. and Zhao, G. (2025) 'Technical evaluation on improved Hough transform for monitoring traffic sign images under depth algorithm', *Int. J. Dynamical Systems and Differential Equations*, Vol. 14, Nos. 1/2, pp.82–99.

Biographical notes: Jie Ding received her MS degree from Wuhan University, Wuhan, China in 2012. She is currently an Associate Professor in the College of Technology of Hubei Engineering University. Her research interests include computer vision and image processing.

Guotao Zhao holds a degree from Computer Science and Technology from Hubei Engineering University, China in 2004. He is currently an Associate Professor in the Hubei Engineering University. His research interests include computer networks and intelligent image recognition.

1 Introduction

The number of locomotives in China is increasing every year. While promoting people's convenient travel, traffic safety accidents are also increasing yearly, such as congestion caused by large traffic flow during the morning rush hour on weekdays, and traffic accidents caused by drivers fatigued driving or not complying with traffic regulations. These problems not only threaten people's lives and property safety but also hurt the social economy. In addition, the complexity of vehicle types and the rise in vehicle numbers have created significant challenges for vehicle supervision tasks like hit-and-run, case pursuit, and parking lot management. As a result, relevant departments are continuously looking for more effective technical solutions and management strategies. Traffic signs, as the core part of traffic image recognition, play a vital role in traffic management and safety. However, traffic sign recognition faces many difficulties in practical applications. Traditional manually designed feature extractor methods are often slow and have low accuracy, and are difficult to cope with complex traffic environments. With the increase in the number of cars and the increases in traffic complexity, accurate and real-time traffic sign recognition is of great significance for reducing traffic accidents, improving road safety, and supporting intelligent driving. Especially in the context of the rapid development of intelligent transportation systems (ITS) and autonomous driving technologies, the accuracy and speed of traffic sign recognition technology have become key factors affecting the overall performance of the system. To address these challenges, researchers are committed to developing more advanced algorithms and technologies to improve the accuracy and efficiency of traffic sign recognition. This paper studies a deep algorithm based on improved Hough transform, aiming to monitor traffic sign images more accurately and efficiently.

In this paper, the research on the use of improved Hough transform to monitor traffic sign images is mainly based on the depth algorithm, which explores the processing process of traffic sign images from a novel perspective, to achieve innovation. This paper also compares the algorithm in this paper with the ANN algorithm and the random forest algorithm. As a result, the classification accuracy of the algorithm in this paper is up to 99%. The innovation of this paper is

- 1 using an improved Hough transform to monitor traffic sign images
- 2 the algorithm in this paper is compared with the ANN algorithm and the random forest algorithm.

This paper improves the traditional Hough transform and proposes a weighted voting mechanism and adaptive threshold algorithm based on edge strength, which improves the accuracy and robustness of feature detection. In the architecture design of CNN, the combination of the convolutional layer and pooling layer is optimised, which enhances the feature extraction capability of the network and improves the generalisation performance of the model while ensuring computational efficiency. Using GPU parallel computing technology, the calculation process of the Hough transform is accelerated and optimised, which significantly shortens the running time of the algorithm. These innovations solve some of the technical problems in existing methods and provide new ideas and methods for subsequent research.

2 Related work

With the increase in car ownership, transportation has become an area of attention. In recent years, many scholars have studied traffic sign images. Scholar Shi et al. (2016) proposed a hardware accelerator based on FPGA for traffic sign detection based on cascaded classifiers. Because traditional computer vision recognition technology cannot meet the real-time accuracy requirements, Zhang et al. (2020) proposed a traffic sign recognition algorithm based on the improved LeNet-5 algorithm. Akbar (2021) presented traffic sign recognition in Indonesia using convolutional neural networks (CNN). He et al. (2021) proposed a novel semi-supervised learning method that combined global and local features for traffic sign recognition in IOT-based traffic systems. Traffic sign recognition occupies a key position in computer vision applications, and Khayeat et al. (2020) proposed a multilingual traffic sign detection and classification method. The above studies have studied traffic sign images from many aspects, but the above studies only use existing algorithms and technologies to study traffic sign images and do not use new technologies to study traffic sign images. The contribution of previous research lies in the application of existing algorithms such as cascade classifier, LeNet-5 improved algorithm, and CNN, which improves the real-time and accuracy of traffic sign recognition. However, the shortcoming is that new technologies and multi-dimensional methods are not fully utilised, which limits further performance optimisation and innovation. Therefore, it is necessary to study image processing methods from other dimensions.

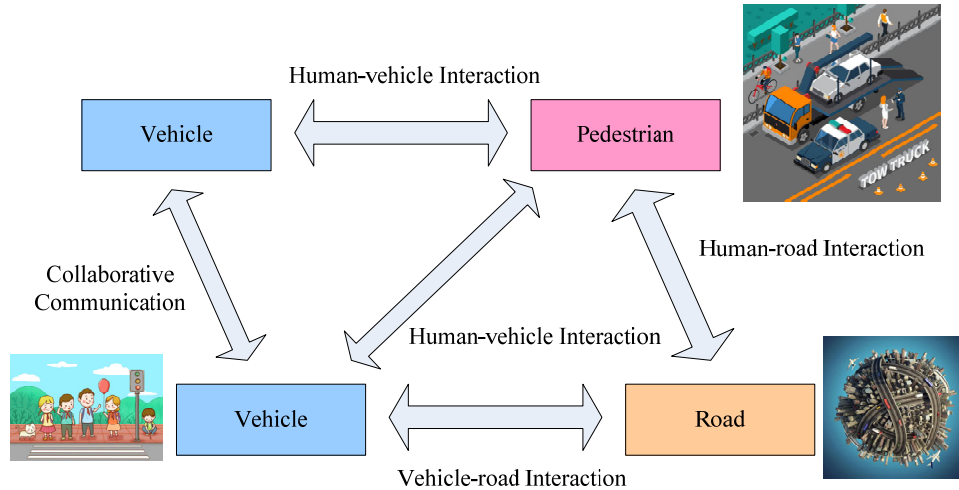
Aiming at the above problems, this paper will research improving the Hough transform to monitor traffic sign images under the depth algorithm. These two methods have long been widely used in other fields. To simplify the lane line detection algorithm based on the Hough transform, scholar Zheng et al. (2018) proposed an algorithm to directly identify lane lines in Hough space. Aiming at the disadvantage that the conventional Hough transform line detection algorithm is prone to false detection, Li et al. (2017) proposed a line detection algorithm based on the improved Hough transform. Yura et al. (2021) evaluated the performance of commercial deep learning algorithms to automatically measure subsolid lesions. Han et al. (2018) used a deep learning algorithm to classify clinical images of 12 skin diseases. Farooq and Bazaz (2021) employed deep learning to propose an artificial neural network (ANN)-based online incremental learning technique for developing an adaptive and non-invasive analytical model of the COVID-19 pandemic to analyse the temporal dynamics of disease spread. The above research shows the application of deep algorithms and improved Hough transform in the medical field, but the research and analysis of traffic sign image processing are relatively lacking. Therefore, this paper will carry out the research and analysis of improved Hough transform monitoring traffic sign images under the depth algorithm, provide a theoretical direction for its subsequent development, and lay a foundation for future practical applications.

3 Traffic sign recognition based on improved Hough transform under deep algorithm

3.1 Traffic sign images

Figure 1 is a schematic diagram of real-time communication between vehicles, pedestrians, and roads.

Figure 1 Schematic diagram of human-vehicle-road collaborative interaction (see online version for colours)



There are various types of traffic signs, and accurate and real-time identification of traffic sign information can reduce the incidence of traffic accidents and road traffic violations. In actual traffic scenes, human damage, weather illumination, and aging are all difficult problems that affect traffic sign recognition. In addition to being interfered with by the complexity of natural scenes, traffic sign recognition would also reduce network performance due to changes in image size. Therefore, the value brought by accurate identification is very meaningful. It can not only reduce the incidence of traffic accidents, and promote functions such as unmanned driving, intelligent navigation systems, and assisted driving, but also be used in auxiliary traffic management departments to scientifically manage and maintain damaged traffic signs.

3.2 Types of Hough transform

The Hough transform is a parameter estimation technique that uses the voting principle. The biggest advantage of Hough transform is that it has strong anti-interference. Hough transform is a technique based on parameter space voting, which is used to detect geometric shapes in images, such as lines and circles. By transforming the points in the image space into curves in the parameter space, the intersection points represent the parameters of the shape, and the detection is realised. Even if there are small disturbances, gaps, or even dotted lines on the curve, after the Hough transform, obvious peaks can still be formed in the parameter space. It is one of the basic methods for

identifying geometric shapes from images. Therefore, the Hough transform is widely used in road recognition and traffic sign recognition in the field of image processing (Al, 2021; Xu et al., 2019).

3.2.1 Hough transform

In the known image space X-Y, all straight lines passing through the point (x, y) satisfy the formula $y = ax + b$. Among them, a and b represent the slope and intercept of the straight line, respectively, and are the parameter values that need to be calculated. The formula of a straight line is expressed as:

$$b = -ax + y \quad (1)$$

This formula represents a straight line through a point (a, b) in the parameter space b - a , and the slope and intercept are determined by the parameters x and y .

Any three points online $y = a_0x + b_0$ are assumed to be (x_1, y_1) , (x_2, y_2) , and (x_3, y_3) . The two lines in the parameter space corresponding to (x_1, y_1) and (x_2, y_2) are:

$$b = -x_1a + y_1 \quad (2)$$

$$b = -x_2a + y_2 \quad (3)$$

Therefore, the intersection of these two lines can be obtained as:

$$a_1 = \frac{y_1 - y_2}{x_1 - x_2} = a_0 \quad (4)$$

$$b_1 = \frac{x_1y_2 - x_2y_1}{x_1 - x_2} = b_0 \quad (5)$$

Similarly, the intersection of the two straight lines corresponding to points (x_2, y_2) and (x_3, y_3) can be obtained as:

$$a_2 = \frac{y_3 - y_2}{x_3 - x_2} = a_0 \quad (6)$$

$$b_2 = \frac{x_2y_3 - x_3y_2}{x_2 - x_3} = b_0 \quad (7)$$

From this, it can be obtained that if there are n points on the line $y = ax + b$, then these points correspond to a line cluster in the parameter space, and all lines intersect at one point. This property can be used to detect collinear points. If polar coordinates are used for the mapping calculation, then n points on the straight line correspond to n sinusoids in the polar coordinate space, and all sinusoids intersect at one point, forming a sinusoid cluster with a clear trend of change.

The standard Hough transform has significant limitations. The specific manifestations are a large amount of calculation; large storage space required; and poor grasp of quantitative scale. Whether the quantisation is too large or too small would have a bad influence on the detection of the straight line.

3.2.2 Improvement of Hough transform

To overcome the defects brought by the standard Hough detection line, an improved Hough transform is adopted. Not only can the straight line be detected quickly, but also the position of the end point of the straight line can be determined. The processing time is short and the storage space required is small, and false straight lines are not detected. This paper first optimises the parameter space of the Hough transform and uses the polar coordinate system to reduce the dimension of the parameter space, thereby reducing the computational complexity. At the same time, to solve the problem of equal voting weights in the traditional Hough transform, this paper introduces a weighted voting mechanism based on edge strength to more accurately reflect the real information of feature points by adjusting the voting weights. In addition, to improve the accuracy of peak detection, this paper adopts an adaptive threshold algorithm, which can dynamically adjust the detection threshold according to the characteristics of different images, thereby improving the robustness and accuracy of detection. Finally, this article also optimises the entire Hough transform process by using GPU parallel computing technology, which significantly improves the execution speed of the algorithm.

Extraction of the outer contour of the document image: the edge represents the important information in the image. It is the place where the pixels in the image have step-like changes or roof-like changes, representing important information such as the position, contour, and orientation of the image. In many cases, the edges of objects need to be acquired for advanced analytical processing.

Screening statistics: the contour points found in the contour scanner are mapped to polar coordinates. Here, the upper left corner of the traffic sign image is no longer taken as the coordinate origin. To facilitate the calculation and reduce the number of operations, its centre point is the origin of the coordinates, and the filter conditions are set. If the calculated α is greater than 0, it would be accumulated, otherwise it would not be counted. β is expanded to 0 to 360 degrees, and the local maximum value of the accumulator of the collinear point of the contour point and the value of α_0 and β_0 at this time is found.

The values of $\alpha\beta$ and corresponding to the previous step are found. According to the corresponding relationship between the rectangular coordinate system and the parameter coordinate system and the geometric characteristics of the traffic sign image, the adjacent straight lines must intersect with the same point, and the coordinate position of the vertex can be calculated according to the formula.

$$\alpha_0 = a_0 * x + b_0 * y \quad (8)$$

The solution of the formula yields the coordinates of one of the endpoints as shown in formulas (9) and (10):

$$x = (\alpha_0 * b_1 - \alpha_1 * b_0) / (a_0 * b_1 - a_1 * b_0) \quad (9)$$

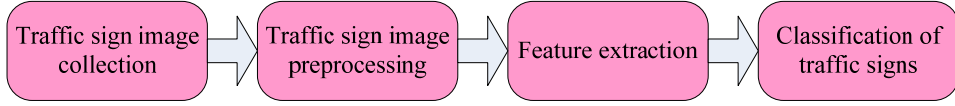
$$y = (\alpha_1 * a_0 - \alpha_0 * a_1) / (a_0 * b_1 - a_1 * b_0) \quad (10)$$

3.3 Depth algorithm

Traffic signs are an important recognition object in traffic images, which can provide basic data for vehicle-road coordination and unmanned driving. The traffic sign image is

based on the real road data captured by the vehicle-mounted camera. The main factors affecting the recognition effect are the different sizes of the traffic sign image, the change of illumination in different periods, the traffic sign being artificially damaged, and the blurred shooting caused when the data is collected. The basic steps of traffic sign recognition in real road scenes are shown in Figure 2.

Figure 2 Basic steps of traffic sign recognition (see online version for colours)



There are two methods for traffic sign feature extraction. One is to design feature extraction descriptors to extract features, and the other is to extract features through network self-learning, mainly in the following two categories.

Traffic sign recognition based on the machine learning method is to recognise traffic signs by machine learning method. In feature extraction, feature extraction operators with strong adaptability to image deformation can be used to extract traffic sign features, which can make the algorithm more robust in the face of complex deformed images, and then use classifiers for classification (Saif et al., 2019).

The traffic sign recognition based on the deep learning method is that the CNN has shown good results in image processing. The algorithm obtains the feature information of the input target task through layer-by-layer mapping, convolution calculation, and pooling processing, which is very helpful for traffic sign recognition (Soetedjo and Somawirata, 2018).

3.3.1 Theoretical basis of convolution

Convolution layer: CNN performs convolution operation on the input image data to obtain the feature information of the data. Due to the different number of convolution kernels, each convolution layer can obtain a large number of feature maps with different specifications. The convolutional layer has the characteristics of weight sharing, and the neurons between each feature map in the feature extraction process can share the weight (Natarajan et al., 2018). Formula (11) is the formula for the convolution operation:

$$S(p, q) = (X * K)(p, q) = \sum_m \sum_n x(p + m, q + n)k(m, n) \quad (11)$$

In the formula, $*X$ represents the convolution operator and the input target data, respectively, K represents the convolution kernel used for the convolution operation and $S(p, q)$ represents the final output feature vector.

Figure 3 shows the process by which the convolution layer extracts the features of the target data through the convolution operation. It can be seen from the figure that the parameters of the convolution kernel are set through the network, and the convolution calculation result between the input image and the convolution kernel is obtained according to the calculation formula, to obtain the image feature vector.

Pooling layer: after convolution calculation and nonlinear mapping, the mapped features are sent to the pooling layer. The main role of the pooling layer is to downsample the incoming feature vector to reduce dimensionality (Junjia et al., 2020).

The main function of the pooling layer is to reduce the spatial dimension of the feature map through downsampling, thereby effectively reducing the computational complexity and memory consumption of the network. At the same time, the pooling layer in CNN also plays a role in enhancing the network's translational invariance to the input data. For example, maximum pooling can extract the most significant features of the local area, making the network more responsive to small-scale translation, rotation, or scaling. Strong robustness. In addition, the pooling layer also has the effect of suppressing noise. By aggregating the features of local areas, the pooling layer can effectively filter out small noises in the image, thereby enhancing the significance of the feature map. These characteristics work together to make the pooling layer reduce the complexity of the model in CNN, enhance the generalisation ability of the network, and reduce the occurrence of over-fitting phenomena.

Figure 3 Schematic diagram of convolution operation (see online version for colours)

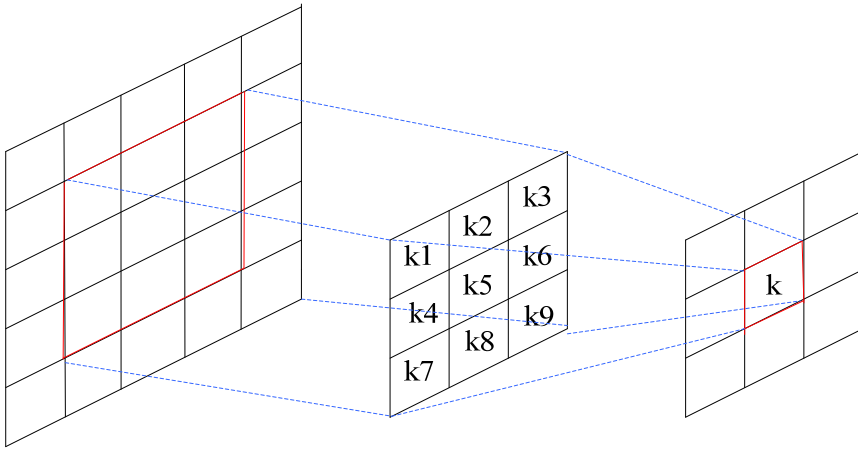
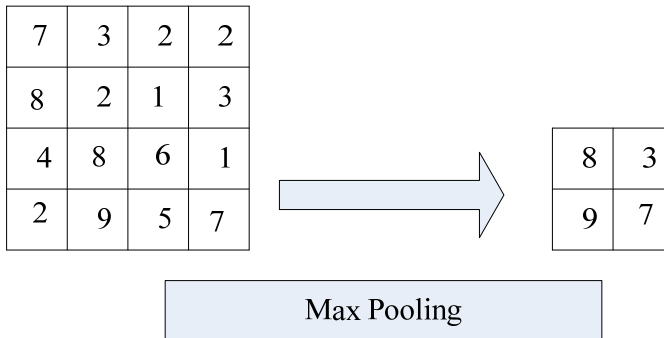


Figure 4 Schematic diagram of max pooling operation (see online version for colours)



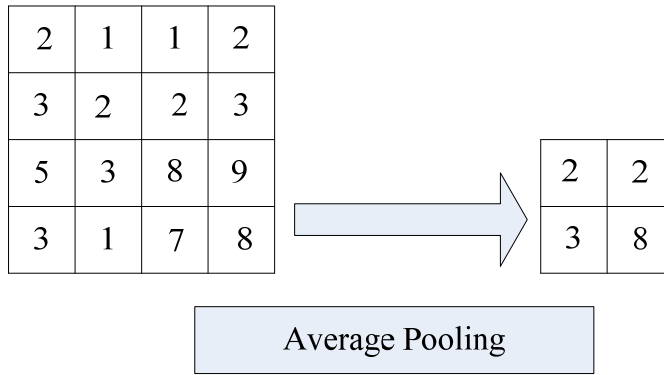
A setting of downsampling with a stride of 2 and a pooling kernel size of 2 performs max pooling, as shown in Figure 4. The average pooling is shown in Figure 5.

Activation function: the purpose of setting the activation function is to map the features extracted by the convolutional layer to the activation function to enhance its nonlinear ability (Gong et al., 2017).

The Sigmoid activation function is:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (12)$$

Figure 5 Schematic diagram of the average pooling operation (see online version for colours)



The Sigmoid function has a powerful nonlinear mapping capability, which enables neural networks to process and learn complex nonlinear relationships, thus breaking through the limitations of linear models. The output value of the Sigmoid function changes smoothly between 0 and 1. This feature is particularly useful in processing probability output or binary classification tasks because it can directly interpret the output as the probability of a certain category. In addition, the derivative of the Sigmoid function is easy to calculate and can efficiently update the network weights during backpropagation. Its smooth derivative curve makes the gradient changes of the network more stable, which helps to accelerate the convergence of the network. Finally, the Sigmoid function simulates the activation characteristics of biological neurons. This biologically inspired nonlinear characteristic makes the network more expressive when processing complex tasks.

The Tanh activation function is:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (13)$$

The ReLU activation function is:

$$f(x) = \max(0, x) \quad (14)$$

3.3.2 Multi-scale feature fusion and ELM combination algorithm

- **Data pre-processing:** The traffic sign recognition set is selected as the experimental data set. As shown in Figure 6, the pictures in the data set have images under the conditions of illumination changes, artificial damage, motion blur, and shooting angle changes, and the image quality of the datasets varies widely. Image pre-processing is performed on the dataset in two steps. The first step is to perform feature enhancement on images with poor quality, and the second step is to perform size normalisation on the enhanced data.

Figure 6 Traffic signs for the database (see online version for colours)



- Image enhancement: The traffic sign image is collected from the actual road conditions, so it is greatly affected by environmental factors. Therefore, it is necessary to enhance the image features. As shown in Figure 7, the method of histogram equalisation can reduce the influence of the environment on the image quality and enhance the image features, and the enhancement is performed on the colour image, so that the colour features of the traffic signs are also well preserved.

Figure 7 Original image versus histogram equalisation (see online version for colours)



- Image size normalisation: When the vehicle camera collects traffic sign images, the image sizes obtained due to the change of distance would be different, which would lead to poor recognition accuracy of the network. Through multiple sets of comparison experiments, it is finally determined that the bilinear interpolation method is used to normalise the image size because the bilinear interpolation method does not lose more image features while requiring less computation.
- Multi-scale CNN feature extraction: In the task of traffic sign recognition, the image-enhanced data is normalised into three sizes 32×32 , 64×64 , and 128×128 , and three sets of data of different sizes are input into CNNs with different convolution kernels and network depths for pre-training. Through the experimental comparison, the specific parameters of the final designed network are shown in Table 1. The network performs the convolution operation to obtain the feature map corresponding to the traffic sign. A feature map corresponds to a set of feature information, and its calculation formula is:

$$x_i^l = f \left(\sum_{j \in N_i} x_j^{l-1} * k_{ji}^l + a_i^l \right) \quad (15)$$

In the formula, k_{ji} and a_i represents the convolution kernel and bias corresponding to the convolutional layer, respectively, and N_i represents the extracted traffic sign feature.

After the traffic sign features are obtained through the convolution operation, the ReLU nonlinear mapping function is used to map the obtained features to the next layer, and the formula is:

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (16)$$

The pooling layer does not affect the final number of feature maps when downsampling, which is a unique invariance of the pooling layer, and also has the effect of enhancing the robustness of the network, and has strong adaptability to image distortion and scaling. The maximum pooling method is selected when pooling the traffic sign features. The calculation formula is:

$$x_i^l = f(\alpha_i^l b(x_i^{l-1}) + a_i^l) \quad (17)$$

Table 1 Multi-scale CNN model parameter settings

<i>Sub-network</i>	<i>Type</i>	<i>Filter size</i>	<i>Pooling size</i>	<i>Pooling model</i>	<i>Layer output</i>
Size-32	Conv1/pooling1	$5 \times 5/1$	$2 \times 2/2$	Max	$48 \times 15 \times 15$
	Conv2/pooling2	$3 \times 3/1$	$2 \times 2/2$	Max	$64 \times 8 \times 8$
	Conv3	$3 \times 3/1$	$2 \times 2/2$	Max	$64 \times 8 \times 8$
Size-64	Conv1/pooling1	$5 \times 5/1$	$2 \times 2/2$	Max	$84 \times 32 \times 32$
	Conv2/pooling2	$5 \times 5/1$	$2 \times 3/2$	Max	$128 \times 15 \times 15$
	Conv3	$3 \times 3/1$	-	-	$384 \times 15 \times 15$
	Conv4	$3 \times 3/1$	$2 \times 2/2$	Max	$384 \times 4 \times 4$
Size-128	Conv1/pooling1	$6 \times 6/1$	$2 \times 2/2$	Max	$96 \times 31 \times 31$
	Conv2/pooling2	$5 \times 5/1$	$2 \times 2/2$	Max	$128 \times 8 \times 8$
	Conv3	$5 \times 5/1$	-	-	$256 \times 8 \times 8$
	Conv4	$3 \times 3/1$	-	-	$384 \times 9 \times 9$
	Conv5/pooling3	$3 \times 3/1$	$2 \times 2/2$	Max	$384 \times 4 \times 4$

In the formula, b represents the function that performs the pooling operation, and α_i , a_i represents the weight and bias coefficients, respectively.

After adding the dropout layer to the fully connected layer of the three network branches, its coefficient is set to 0.5.

3.3.2.1 Multi-scale feature fusion

Conv represents the convolution layer of each network branch. After the convolution layer, there is a ReLU function for nonlinear mapping. Conv/pooling means that the module needs to be followed by a pooling operation after the convolution operation. The network cascade adopts the cascade method of serial fusion. The cascade formula is as follows:

$$F_{MCNNFC} = \sum_{j=1}^m F_j \quad (18)$$

In the formula, F_j is the fully connected layer feature of each branch network, and m is the number of cascaded networks. The branch network is spliced on the channel through a Concat layer. The formula is:

$$Z_{concat} = \sum_j^c X_j * K_j + \sum_j^c Y_j * K_{j+c} \quad (19)$$

In the formula, Z_{concat} represents the splicing of one of the three channels, K is the convolution kernel of the network to be spliced and $X_j Y_j$ represents any two channels of any two networks respectively. The formula is as follows:

$$\begin{cases} F_1 = M \times C_1 \times G \times V \\ F_2 = M \times C_2 \times G \times V \\ F_3 = M \times C_3 \times G \times V \\ F_{MCNN} = M \times (C_1 + C_2 + C_3) \times G \times V \end{cases} \quad (20)$$

In the formula, F_{MCNN} represents the multi-scale features obtained by splicing network branches on the channel; F_1, F_2, F_3 and represents the fully connected layers of the three network branches; M is the number of obtained traffic sign feature maps; G represents the height of the feature map; V represents the width of the feature map.

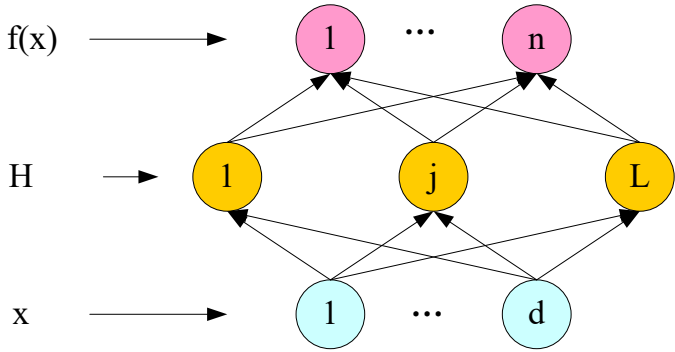
3.3.2.2 ELM classifier

The ELM classifier can randomly initialise the weights and biases of the input features, and the adjustment process of network thresholds and weights does not require repeated calculations, which simplifies the calculation steps, reduces the training time of the classifier, reduces the parameters to a certain extent, and is beneficial to the real-time recognition of traffic signs by the multi-scale fusion model (Chen and Yang, 2017). Its structure is shown in Figure 8.

3.3.3 Other algorithms

- ANN algorithm: The ANN algorithm is composed of a large number of interconnected nodes by simulating the structure and function of biological neural networks, and can be used to model complex relationships between data (Kim et al., 2019).
- Random forest algorithm: Random forest is a tree-based algorithm. It is a collection of many different kinds of random trees. The final value of the model is the average of all predictions produced by each tree (Medeiros et al., 2019).

Figure 8 ELM classifier structure (see online version for colours)



4 Experiment and evaluation of traffic sign image recognition under deep algorithm

4.1 Sample data

The experimental environment parameters of the verification and comparison experiments are shown in Table 2.

Table 2 Specific parameters of the experimental environment

Serial no.	Name	Parameter
1	Processor	I5-7500 3.4GHz frequency
2	Memory	32 GB
3	Graphics card	Nvidia GeForce GT730
4	Operating system	Ubuntu 16.04
5	Deep learning framework	Caffe
6	Programming language	Python 2.7

Evaluation index: considering that the classification time gap is small and the experimental hardware is equivalent, the classification accuracy is selected as the evaluation index, and the calculation is formula (21):

$$Accuracy = \frac{\sum_{j=1}^M m_j}{M} \times 100\% \quad (21)$$

In the formula, M represents the total amount of the traffic sign test set and m_j represents the number of correct classifications.

Bayesian optimisation is used to systematically explore the optimal parameter configuration and improve the performance and stability of the model.

4.2 Experiment results and evaluation of traffic sign image recognition technology

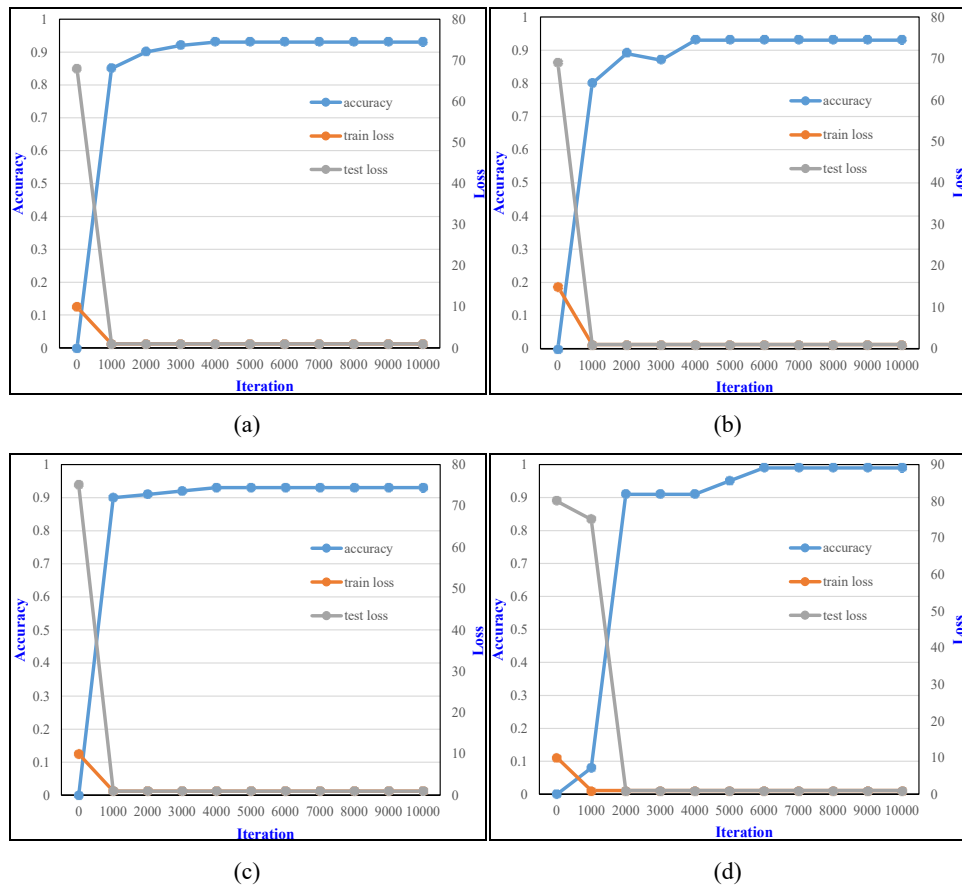
4.2.1 Experiment 1

The datasets of three sizes are input into three pre-trained networks, and the experimental hyperparameters are shown in Table 3.

Table 3 Experimental hyperparameter settings

Serial no.	Name	Parameter
1	Base learning rate	0.02
2	Weight decay	0.0005
3	Learning rate change factor	0.2
4	The maximum number of iterations	10,000
5	Learning rate adjustment policy	Step
6	Step size	4,000
7	Momentum	0.8

Figure 9 Loss value and classification accuracy, (a) size-32 (b) size-64 (c) size-128 (d) multi-scale (see online version for colours)



The experimental environment includes a computer with NVIDIA GTX 1080 Ti GPU and 16 GB RAM, with Ubuntu 18.04 as the operating system and TensorFlow 2.4 as the deep learning framework. The experiment was carried out with the support of CUDA 11.0 and cuDNN 8.0.

To ensure that the pre-training network can be used for model fusion, the loss value curves of training and testing of the pre-training network and the recognition accuracy curve of the pre-training network are drawn to judge the effectiveness of the pre-training model. As shown in Figures 9(a), 9(b), 9(c) and 9(d) respectively draw the accuracy curve, training loss curve, and test loss curve of the network model with input traffic sign sizes of 32×32 , 64×64 , 128×128 , and after feature fusion.

Figure 9(a) is a model drawing of size 32; Figure 9(b) is a model drawing of size 64; Figure 9(c) is a model diagram of size 128; Figure 9(d) is a multi-scale model diagram. It can be seen from Figures 9(a), 9(b) and 9(c) that the network classification accuracy of the three pre-trained models increases rapidly in the first 2,000 iterations, between 3,000 and 10,000 iterations. The classification accuracy changes are no longer obvious and tend to be stable. From the analysis of the two curves of training and classification loss values, the loss values decrease rapidly in the first 50 iterations, and are unchanged in the subsequent iterations, indicating that the model learning has been completed, the network model is stable, and the generalisation ability is strong. The multi-scale model integrating the three branch networks gradually stabilised when the loss value fell to 2,000 times, but the classification accuracy of the model was higher, reaching 99%.

4.2.2 Experiment 2

To verify the superiority of using the ELM classifier to classify the fusion network, a comparative experiment is designed, and the classification accuracy and time are used as two basic evaluation indicators to compare the performance of each classifier, as shown in Table 4.

Table 4 Performance comparison of different classifiers

<i>Classifier</i>	<i>Classification accuracy/%</i>	<i>Classification time/ms</i>
Softmax	98.56	50
SVM	98.25	60
ELM	99	45

According to the table, the parameters of the ELM classifier can be initialised randomly, and the adjustment process of the network threshold and weight does not require repeated calculation, which simplifies the calculation steps and reduces the training time of the classifier. Therefore, the feature learning of the ELM classifier is faster, the parameters are less, the classification time is shorter than the other two classifiers, and the classification accuracy is slightly improved than that of the softmax and support vector machine (SVM) classifiers. ELM classifier initialises the parameters randomly, without adjusting the weights and thresholds repeatedly, which greatly simplifies the calculation process. Therefore, its feature learning speed is faster, its parameters are less, and its classification time is shorter than other classifiers.

4.2.3 Experiment 3

To further verify the advantages of the algorithm in this paper, the algorithm in this paper is compared with the ANN algorithm and the random forest algorithm. In the data set, 30 traffic sign images with blur, serious occlusion, and extremely poor lighting conditions are selected to form a traffic sign test set in a complex environment. The performance of the algorithm under complex conditions is tested in single-layer feature and multi-layer feature networks respectively, and the results are shown in Figure 10.

Figure 10 Comparison of recognition and classification time and classification accuracy in complex environments, (a) ANN algorithm compared with the algorithm in the text (b) comparison of random forest algorithm and algorithm in the paper (see online version for colours)

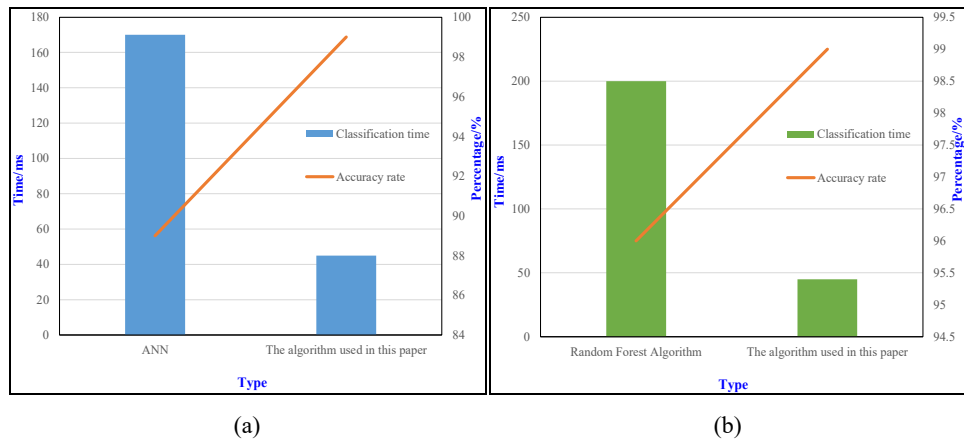


Figure 10(a) is the comparison between the ANN algorithm and the algorithm in the text; Figure 10(b) is the comparison between the random forest algorithm and the algorithm in the text. As can be seen from Figure 10, the algorithm proposed in this paper reduces the classification time by 125 ms and 155 ms respectively compared with the ANN algorithm and the random forest algorithm, and the classification accuracy is up to 99%. The reason is that the multi-scale feature fusion network divides the data into three sizes, large, medium, and small, and inputs it into three different network branches for training. This approach is equivalent to a threefold increase in the dataset. In addition, the multi-scale traffic sign features are obtained in the experiment. The fusion of multi-scale features is helpful for the update process of network parameters and the solution of algorithms. Therefore, the data quality is greatly affected by the environment, especially the blurred images caused by shooting. The performance of the algorithm in this paper is better than other algorithms in the face of this complex environment. Through multi-scale feature fusion, the algorithm in this paper improves the classification accuracy to 99% and significantly reduces the classification time. This method enhances the robustness of the network and the efficiency of parameter updating, especially when dealing with blurred images.

5 Conclusions

This paper studied the technology of improving the Hough transform to monitor traffic sign images under the depth algorithm, which can provide a more scientific and efficient method for traffic sign image processing and recognition. In this paper, the traffic sign data was firstly enhanced by the histogram equalisation method, and then the enhanced image was normalised into three sizes of large, medium, and small by bilinear interpolation method; the network structure and convolution kernel size adapted to the image size were constructed for feature extraction, the multi-scale features were fused in the fully connected layer of the overall recognition network, and the fully connected layer was fine-tuned; finally, an extreme learning machine (ELM) was used to classify the traffic sign images. The classification accuracy of the multi-scale model integrating the three branch networks was higher, reaching 99%, and the performance of the algorithm in this paper was better than other algorithms in the face of this complex environment. However, technology is always improving, and there are still problems that cannot be overcome at present. It is hoped that the traffic sign image processing and recognition technology will continue to innovate and make breakthroughs in the future. The research in this paper is mainly carried out in the laboratory environment, and its effect in the real and complex traffic environment has not been fully verified. Moreover, the model has a high complexity, and it may face the limitation of computing resources in practical applications. In the future, the model will be further optimised to reduce the consumption of computing resources and will be widely tested in real and complex traffic environments. In addition, explores the use of new technologies and multi-dimensional methods to improve the performance and application scope of traffic sign recognition.

Acknowledgements

This work was supported by the Teaching Research Project of Hubei Provincial Department of Education (No. 2020806); The Science and Technology Research Project of Hubei Provincial Education Department (No. B2022506); The Teaching Research and Scientific Research Project of College of Technology, Hubei Engineering University (No. 2020JY04, No. 2022KY07).

References

- Akbar, M. (2021) 'Pengenalan rambu lalu lintas menggunakan convolutional neural networks', *Jurnal Teknologi dan Sistem Komputer*, Vol. 9, No. 2, pp.120–125.
- Al, A. (2021) 'Traffic sign classification using convolutional neural networks and computer vision', *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, Vol. 12, No. 3, pp.4244–4250.
- Chen, Y. and Yang, Y. (2017) 'Two improved algorithms for ellipse detection based on Hough transform', *Bandaoti Guangdian/Semiconductor Optoelectronics*, Vol. 38, No. 5, pp.745–750.
- Farooq, J. and Bazaz, M.A. (2021) 'A deep learning algorithm for modeling and forecasting of COVID-19 in five worst affected states of India', *Alexandria Engineering Journal*, Vol. 60, No. 1, pp.587–596.

- Gong, X., Gao, K., Wang, Y., Lin, J. and Zhu, J. (2017) 'A novel linear target detection method based on improved probability hough transform in remote sensing imagery', *Yingxiang Kexue yu Guanghuaxue/Imaging Science and Photochemistry*, Vol. 35, No. 2, pp.162–167.
- Han, S.S., Kim, M.S., Lim, W., Park, G.H., Park, I. and Chang, S.E. (2018) 'Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm', *Journal of Investigative Dermatology*, Vol. 138, No. 7, pp.1529–1538.
- He, Z., Nan, F., Li, X., Lee, S.J. and Yang, Y. (2020) 'Traffic sign recognition by combining global and local features based on semi-supervised classification', *IET Intelligent Transport Systems*, Vol. 14, No. 5, pp.323–330.
- Junjia, J., Jianxin, S., Zhe, Z. and Peng, H. (2020) 'Research on rotary drum assembly and adjustment technology based on improved probabilistic Hough transform', *Journal of Applied Optics*, Vol. 41, No. 2, pp.394–399.
- Khayeat, A.R.H., Abdulmunem, A., Al-Shammari, R.F.N. and Sun, X. (2020) 'Traffic sign detection and classification based on combination of MSER features and multi-language OCR', *Webology*, Vol. 17, No. 2, pp.394–403.
- Kim, H., Kim, Y., Han, B., Jang, J.Y. and Kim, Y. (2019) 'Clinically applicable deep learning algorithm using quantitative proteomic data', *Journal of Proteome Research*, Vol. 18, No. 8, pp.3195–3202.
- Li, H., Zhu, W., Tan, M. and Zheng, X. (2017) 'Line detection algorithm based on improved Hough transform', *Bandaoti Guangdian/Semiconductor Optoelectronics*, Vol. 38, No. 4, pp.603–608.
- Medeiros, F.A., Jammal, A.A. and Thompson, A.C. (2019) 'From machine to machine: an OCT-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs', *Ophthalmology*, Vol. 126, No. 4, pp.513–521.
- Natarajan, S., Annamraju, A.K. and Baradkar, C.S. (2018) 'Traffic sign recognition using weighted multi-convolutional neural network', *IET Intelligent Transport Systems*, Vol. 12, No. 10, pp.1396–1405.
- Saif, A.S., Paul, P., Zubair, K.M., Shubho, S.A. and Nandi, D. (2019) 'Traffic sign detection based on color segmentation of obscure image candidates: a comprehensive study', *International Journal of Modern Education and Computer Science*, Vol. 10, No. 6, pp.35–46.
- Shi, W., Li, X., Yu, Z. and Overett, G. (2016) 'An FPGA-based hardware accelerator for traffic sign detection', *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 25, No. 4, pp.1362–1372.
- Soetedjo, A. and Somawirata, I.K. (2018) 'An efficient algorithm for implementing traffic sign detection on low cost embedded system', *An Efficient Algorithm for Implementing Traffic Sign Detection on Low Cost Embedded System*, Vol. 14, No. 1, pp.1–14.
- Xu, X., Jin, J., Zhang, S., Zhang, L., Pu, S. and Chen, Z. (2019) 'Smart data driven traffic sign detection method based on adaptive color threshold and shape symmetry', *Future Generation Computer Systems*, May, Vol. 94, pp.381–391.
- Yura, A., Min, L.S., Na, N.H., Wooil, K., Jooae, C., Kyung-Hyun, D. et al. (2021) 'Use of a commercially available deep learning algorithm to measure the solid portions of lung cancer manifesting as subsolid lesions at CT: comparisons with radiologists and invasive component size at pathologic examination', *Radiology*, Vol. 299, No. 1, pp.202–210, <https://doi.org/10.1148/radiol.2021202803>.
- Zhang, C., Yue, X., Wang, R., Li, N. and Ding, Y. (2020) 'Study on traffic sign recognition by optimized LeNet-5 algorithm', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 34, No. 1, p.2055003.
- Zheng, F., Luo, S., Song, K., Yan, C.W. and Wang, M.C. (2018) 'Improved lane line detection algorithm based on Hough transform', *Pattern Recognition and Image Analysis*, Vol. 28, No. 2, pp.254–260.