ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0031

Detection and Recognition of Traffic Signs based on Improved Deep Learning

Anliang Kou

School of Information Engineering, Shanghai Maritime University, Shanghai 201306, China.

Abstract

Facing various special environments, combining image preprocessing and deep learning neural network, a road traffic sign recognition algorithm is proposed. This method uses image segmentation to detect traffic signs, and uses a convolutional neural network model to more accurately identify road traffic signs. First, preprocess the data by adjusting the lighting effects, feature matching, normalization, and data enhancement to form a data set that meets the requirements; then, combined with the idea of AlexNet and the residual network structure, fully train your own convolution Neural network model; Finally, the optimized network model is used for road traffic sign recognition to compare with the AlexNet model. The experimental results show that the accuracy of the AlexNet model is improved, and the recognition accuracy can reach 94.50%.

Keywords

Traffic signs; Image segmentation; Convolutional neural network; Data enhancement; Normalization; AlexNet.

1. Introduction

The detection and recognition of road traffic signs has received extensive attention in the rapidly developing field of intelligent transportation, which contains prompt information such as warnings, prohibitions, and instructions. Through in-depth research on the identification and detection of road traffic signs, it has very important practical significance and application value for the development of social economy and the safety of people's lives.

There are many detection methods for road signs. For example, using different color spaces [1, 2] for matching detection methods, shape and contour [3, 4] segmentation and extraction methods, or fusion shape and color features [5] detection methods, or use deep learning, neural network related Methods of knowledge detection [6, 7]. The deep learning model can be regarded as composed of multiple artificial neural network layers. By constructing a neural network model with multiple hidden layers, the low-level features are transformed by layer-by-layer nonlinear feature combination to form a more abstract high-level feature expression. This paper combines the principles of the AlexNet model and the residual network model, taking their respective advantages, and using a better number of network layers to achieve a higher level of recognition.

Therefore, in response to the above-mentioned problems, the work done in this article is:

1) A method of combining HU invariants and multi-scale color space for processing is proposed to reduce the influence of illumination and complex background, and reduce the probability of useless features to be learned;

2) Through data enhancement and normalization operations, small data sets are enhanced to meet network training requirements and improve recognition accuracy;

3) Combining the structure of AlexNet network and residual network to construct its own network model. At the same time, by adjusting the corresponding parameters to adapt to the characteristics of the traffic sign image, a high recognition accuracy rate is ensured.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0031

2. Image Preprocessing

2.1 Color Space Processing Of The Data Set

The influence of environmental factors will bring great difficulty to the image collection of traffic signs, and thus cause identification difficulties. This paper uses threshold-based segmentation method for all pixels in the image from two different color spaces of RGB and YCbCr [8], so as to realize the preprocessing of the image and reduce the interference of environmental factors. The RGB color space model expresses the color image by constructing a Cartesian rectangular coordinate system with three channels of red, green, and blue. The value range of the channels is [0,1].

Because the color of the road traffic sign image is special, we can achieve the effect of enhancing the image by changing the brightness value of the corresponding color space. The YCbCr color space can display the overall brightness information of the image compared to the RGB color space. The specific conversion formula [9] for converting RGB image into YCbCr space representation is

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

After space conversion, the YCbCr color space information of the pictures in the road traffic sign data set is obtained, and the Y value information is taken, that is, the brightness information. Through a large number of experiments, it is relatively normal for the image brightness threshold to be within Y $\in$ [85,200], and the out-of-threshold images are processed through image enhancement and other methods. Image enhancement is mainly carried out by superimposing or subtracting RGB color space. The processing formula for underexposed images is $R = R + (1-R).\times R*\mathrm{k}$. The overexposure image processing formula is $R = R - (1-R).\times R*k$. Where k is an adjustable parameter.

2.2 Feature Matching

In digital image processing, segmentation by extracting contour boundaries and shape features is a method often used in the field of image recognition. This paper integrates the common shape features of traffic sign images for feature extraction, reducing the useless feature learning of the neural network during data set training. Traffic sign shape feature map, see Fig. 1.

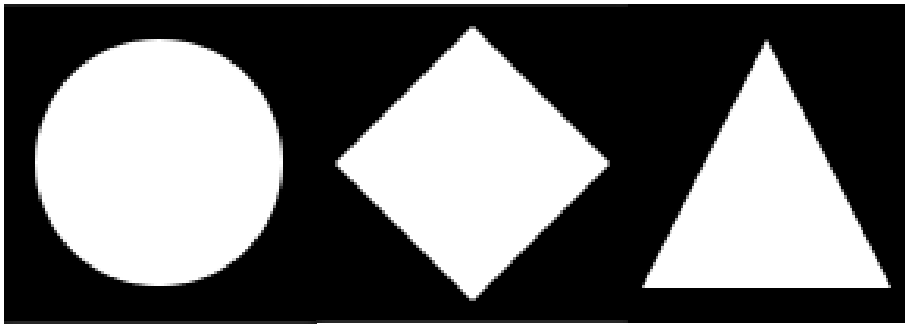

Fig 1 The shape feature map of traffic signs

Aiming at the problem of noise pollution generated in the image processing process, the noise in the image is reduced through the opening and closing operations of morphology, and then the maximum between-class variance method is used to find the image threshold, and the threshold value is used to convert the gray image into a binary image; finally used The hole filling extracts the characteristic shape of the image. By matching the feature vector of the traffic sign image with the standard contour image, a mask is formed to extract the features of the traffic sign.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0031

2.3 Normalization and Data Enhancement

Normalize the input image and process the input features into a similar range, making the optimization of the cost function easier and faster. In this project, the training set and test set are normalized to a size of 128 × 128.

In the process of traffic sign recognition training, the original data is trained, and the results show that the accuracy of the training set is quite high, but the accuracy of the verification set is low, showing overfitting. When over-fitting occurs, the high accuracy of the training set indicates that the algorithm has fully learned the characteristics of the original data; the low accuracy of the validation set indicates that the characteristics of the original data are not sufficient, which makes the algorithm perform poorly in the new validation set.

For small-sized pictures, according to the characteristics of the deep learning model, the image can be enhanced by geometric transformations such as translation, rotation, scale stretching, contrast adjustment and color transformation; for large-sized pictures, the mean reduction is used to reduce the image size. This article mainly uses rotation transformation to expand the data set with less data.

3. Deep Learning Algorithm

3.1 Convolutional Neural Network

With the deepening of the deep learning network and the increase in the dimensions of the input data (the image size becomes larger), the original neural network and the fully connected network of the autoencoder will generate a huge amount of network connection parameters, making the deep learning network huge. It's even difficult to train. The convolutional neural network [10] can share the convolution kernel, which has no pressure on high-dimensional data processing. Secondly, the convolutional neural network does not need to manually select features. If the weights are trained, the feature classification effect is good.

Convolutional neural network is mainly composed of:Input layer, CONV layer,ReLU layer, Pooling layer, FC layer, etc.

Convolutional network is essentially a kind of input to output mapping. It can learn a large number of mapping relationships between input and output without requiring any precise mathematical expressions between input and output, as long as the known The model trains the convolutional network, and the network has the ability to map between input and output pairs

3.2 Residual Network Model

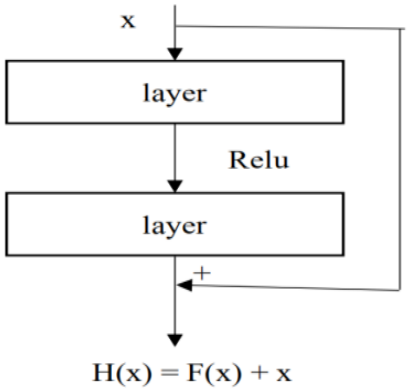

Fig. 2 Residual model structure diagram

In order to improve the generalization ability of the network and reduce the degree of overfitting, the residual network is introduced, which can not only deepen the performance of the network, but also minimize the performance degradation problem. The structure of the residual model, see Fig. 2.In the residual network, after the input x passes through the multilayer network, the output H(x) is the addition of F(x) and x as the new residual structure. H(x) = F(x) + x, where F(x) is the residual. In

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0031

order to reduce the gradient explosion problem, simplify the training and learning process, use the mathematical operation to convert the formula H(x) = F(x) + x into H(x) = F(x) - x, where F(x) = 0 when H(x) = x; if F (x) ≈ 0, only the difference between H(x) and x is considered, and the residual network uses this idea.

3.3 AlexNet Network

In order to improve the training speed, AlexNet [11] uses ReLU instead of Sigmoid, which can train faster, and at the same time solves the problem of sigmoid's gradient disappearance in a deeper network, or gradient dispersion [12]. Structure diagram of AlexNet, see Fig. 3.

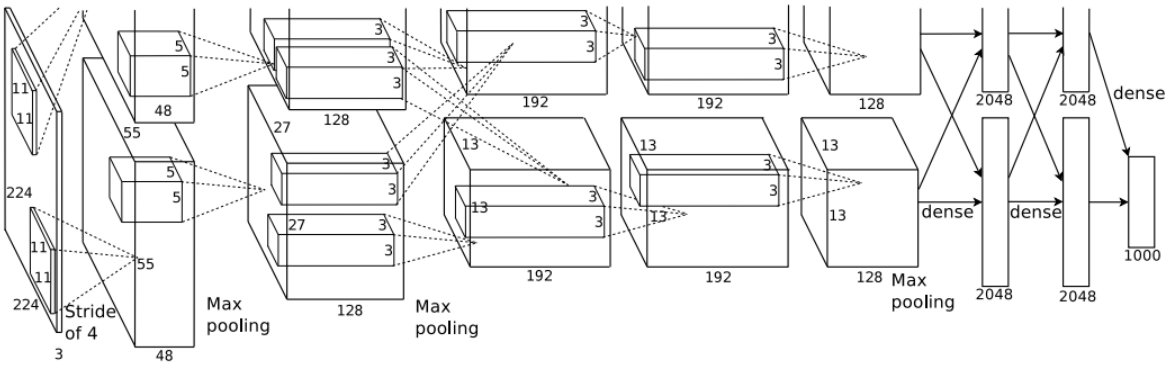

Fig. 3 AlexNet structure diagram

The total number of layers of the network is 8 layers, 5 layers of convolution, and 3 layers of fully connected layers. Each convolution layer contains the excitation function RELU and local response normalization (LRN) processing, and then after downsampling (pool processing)).

AlexNet has many features such as a deeper network structure, the use of stacked convolutional layers, that is, convolutional layer + convolutional layer + pooling layer to extract image features, use of Dropout to suppress overfitting, and use of data to enhance Data Augmentation suppression Overfitting, using Relu to replace the previous sigmoid as the activation function, multi-GPU training.

3.4 Improved Traffic Sign Recognition Algorithm Based on Deep Learning

By integrating the above theories and formulas, worrying about the network structure, adjusting the model parameters, and combining the idea of combining AlexNet and residual network, the algorithm of this paper is formed. Structure diagram, see Fig. 4.

1) Data preprocessing, feature extraction, normalization and data enhancement.

2) Combining the ideas of AlexNet algorithm and residual network, constructing an improved network model by adjusting parameters, and fully training and learning the training set.

3) Detect and recognize traffic signs with the trained model to verify the effectiveness of the algorithm.

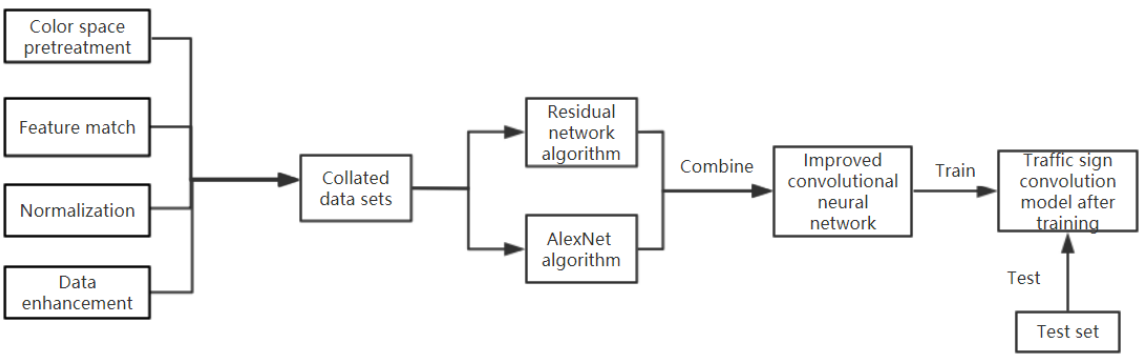

Fig. 4 Structure diagram of improved traffic sign algorithm

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0031

4. Experimental Results And Analysis

The experiment uses the German Traffic Sign Standard Database (GTSRB), which contains 39,209 training set road sign images and 12,630 test set road sign images. There are 43 categories; the size of the images in the data set ranges from 15 × 15 to 250 × 250. The training set is normalized and data enhanced, and the image size is unified to (128 × 128). The initial learning rate is set to 0.001, the training set image processing volume is set to 100, and the test set image processing volume is set to 10. The final output of the algorithm, see Table 1.

Table 1 Comparison of the recognition rate of the training set and the test set of the two models

Model name	Training set accuracy	Test set accuracy
AlexNet	0.912	0.912
This model	0.962	0.945

It can be seen from Table 1 that the improved algorithm has a higher recognition rate and can be used for the recognition of traffic signs.

5. Conclusion

This paper proposes an improved traffic sign recognition algorithm based on deep learning. The algorithm first extracts the area of the traffic sign using features such as edges, colors, and shapes, and then uses an improved deep learning algorithm to identify the database. Through experiments on the German standard traffic sign map database, the results show that the algorithm can effectively understand traffic sign information, can make full use of the advantages of neural network's strong learning ability, and make the recognition algorithm of traffic signs have higher recognition accuracy. In addition, because the German standard database is used in the article, there is a certain difference between the traffic sign image information obtained in the actual environment, and the data collection and preprocessing need further study and research. At the same time, in response to the increase in data volume, The real-time effect of the recognition algorithm needs further analysis and research.

References

[1] KHAN J. F. ADHAMI R. R. et al. Traffic Sign Recognition Based on Pictogram Contours[J]. IEEE Ninth International Workshop on Image Analysis for Multimedia Interactive Services. 2011, 1(1):83-96.

[2] YANG H M, LIU C L. et al. Traffic Sign Recognition in Disturbing Environments[C]//Lecture Notes in Computer Science, Maebashi City, Japan. 2003, 2871:252-261.

[3] BARNES Z A. FLETCHER L. S. Real-Time Speed Sign Detection Using the Radial Symmetry Detector[J]. Intelligent Transportation Systems, IEEE Transactions on. 2008, 9(2):322-332.

[4] PAULO C P L. Traffic Sign Recognition Based on Pictogram Contours[J]. IEEE Ninth International Workshop on Image Analysis for Multimedia Interactive Services. 2008, 5(7-9):67-70.

[5] GAO X W L. PODLADCHIKOVA D. et al. Recognition of traffic signs based on their colour and shape features extracted using human vision models[J]. Journal of Visual Communication and Image Representation, 2006, 17(4):675-685.

[6] STALLKAMP J. SALMEN J. IGEL C. Man vs. computer:Benchmarking machine learning algorithms for traffic sign recognition[J]. Neural Networks. 2012, 02:323-332.

[7] PHILIPP Seebock Zsfassung. Deep learning in medical image analysis[J]. Neural Networks. 2012, 02: 333-338.

[8] ISLAM Kh TOHIDUL, RAJ Ram Gopal. Real-Time (Vision-Based) Road Sign Recognition Using an Artificial Neural Network[J]. PubMed periodicals, 2017, 14(4):853.

[9] ZHANG Yi, LIU Xu, LI Hai Feng. The Precision of RGB Color Space Convert to YCbCr Color Space[J]. Journal of Shouthem Yangtze University(Natural Science Edition). 2007, 6(2):200-202. (In Chinese)

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0031

[10] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]. Advances in neural information processing systems. 2012:1097-1105.

[11] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. lmagenet classification with deep convolutional neural networks[C]//Advances in neural Information Processing Systems. Lake Tahoe, USA, 2012:1097-1105.

[12] ZHANG Ling. The Relationship Between Kernel Functions Based SVM and Three-Layer Feedforward Neural Networks [J]. Chinese Journal of Computers, 2002, 44(6):25.(In Chinese)