Scientific
Research
Publishing

Optimization of Convolutional Neural Network for Recognition of Vehicle Frame Number

Haiming Li, Yongxue Liu, Yong Wang

School of Computer Science and Technology, Shanghai University of Electric Power, Shanghai, China
Email: lhm@shiep.edu.cn, 407843800@qq.com

Abstract

With the development of the economy and the surge in car ownership, the sale of used cars has been welcomed by more and more people, and the information of the vehicle condition is the focus information of them. The frame number is a unique number used in the vehicle, and by identifying it can quickly find out the vehicle models and manufacturers. The traditional character recognition method has the problem of complex feature extraction, and the convolutional neural network has unique advantages in processing two-dimensional images. This paper analyzed the key techniques of convolutional neural networks compared with traditional neural networks, and proposed improved methods for key technologies, thus increasing the recognition of characters and applying them to the recognition of frame number characters.

Keywords

Frame Number Recognition, Convolutional Neural Network (CNN), Feature Extraction, Pooling

1. Introduction

In the 2018 Boao Forum for Asia (BFA) annual meeting, auto industry has been mentioned twice during the speech given by China Chairman, Mr. Xi Jinping. With the development of the auto car science and the rising oil price, the second hand market is also rapidly developing. During the procurement of auto cars and with the applying of frame number identification, it will be quick to match the data of service provider and make all the information of the said cars clear, including accidents and illegal driving. The buyer could know the statues of the cars anytime and anywhere after the auto vehicle database is established [1].

In the current character recognition algorithms, it may turn out to be under

fitting when multi-layer neural network is applied due to the insufficient layer of the traditional neural network. With the addition of Feature Learning to the multi-layer neural network and the unique advantages of image recognition, the Convolutional Neural Network (CNN) technology is widely studied in the field of computer vision, model matching and pattern recognition. By the CNN character recognition, not only has been the extract features processing mechanisms in the human visual system simulated to extract underlying image feature information, but also the end-to-end feature extraction has been utilized to enhance the generalization ability and avoid the inaccuracy due to the image morphing.

2. The Problems in Traditional Multi-Layer Neural Network

The traditional multi-layer neural network includes an input layer, an output layer and many hidden layers between them. Every layer consists of some nerve cells. In the neighboring layer, every cell in the latter layer connects with every cell in the former layer respectively. In the image identification, input layer stands for feature vector and every cell of this layer stands for every eigenvalue.

In the vehicle frame number character identification, every nerve cell in the input layer stands for pixel gray values of every frame number image. But several problems exist. On the one hand, space structure is not considered while the position of the frame number is hidden and identification efficiency is limited. On the other hand, there are too many nerve cells in the neighboring layers resulting in a limited training speed.

3. Key Technology of Convolutional Neural Network (CNN)

CNN can be used to solve the problems in the traditional multi-layer neural network. The CNN can be trained faster due to its special structure aiming at the image identification. Thanks to the high efficiency, it makes the multi-layer neural network easier and multi-layer has a great advantage in the accuracy rate of identification. In this way, the usage of CNN, as one of the deep learning algorithms, in the field of frame number and license number identification and traffic management is attracting attention and updating and improving.

3.1. The Structure of CNN

The CNN consists of input layer, hidden layer and output layer while the hidden layer consists of several convolutional layer and subsampling layer. The structure of the CNN is as Figure 1.

In Figure 1, the image of frame number is input from the Input layer and CNN can process it and extract image feature information automatically without human operation. Multi-Kernels in the first convolutional layer dispose the image respectively before they can generate the corresponding convolutional feature map. The second pooling layer will extract local feature from the convolutional feature image in C1 and generate the corresponding subsampling feature
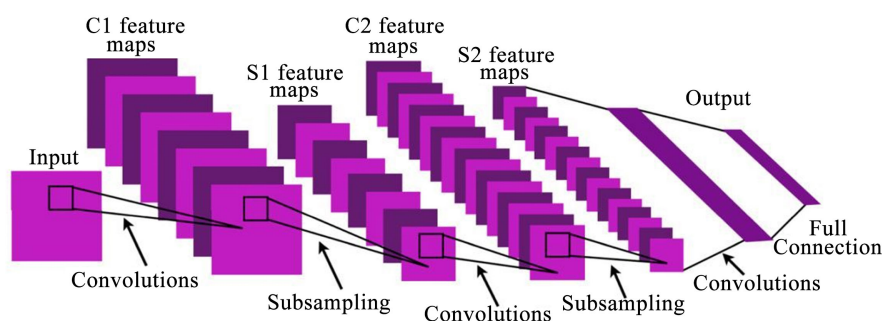
Figure 1. Structure of CNN.

map. C2 and S2 will repeat the operation of C1 and S1. That the hidden layer of CNN has all the same structure of C1 and S1 as well as the structure of convolutional and pooling layers is used to extract features lowers the resolution, increases the quantity of feature map generated and obtains more feature information. The vehicle information can be obtained by the 1D matrix, expanded from the last pooling feature map and fully connected with the out-layer [2].

3.2. Convolutional Layer

Comparing to the creature visual cell's local sensory, the convolution will be made by local filter, or the output result will be taken as the dimensionality corresponding value of the convolutional output matrix. In the CNN's structure, in addition to the size of convolutional kernel, not only will the quantity of convolutional kernel affect the precision but also the selection of the activation function will determine the time efficiency of algorithm. To give a better representation of data, the convolutional layer usually provides more than one such local filter and forms multi-output-matrix. The size of every matrix is $(n-m+1)$ and the details of arithmetic process is as Formula (1):

$$x_k^l = f\left(\sum\nolimits_{i \in M_k} x_i^{l-1} * H_{ik}^l + b_k^l \right) \tag{1}$$

PS: 1 stands for the number of model layer; H stands for convolutional kernel; M_k stands for the k-th feature map of 1 − 1 layer: b stands for the biasing of output plots; f stands for activation function.

In theory, the smaller convolutional kernel is, the more exquisite the extraction feature information. However, the practical image may be polluted in different levels by noise, that is to say, some rambling information can be easily extracted if the kernel is too small [3].

3.3. Pooling Layer

The function of the pooling layer is to lower the matrix dimensionality and does not damage to the inner link of the data. This layer can be established by the way of average or maximum value. The input of this layer come for the former convolutional layer and the output can be treated as the input of the latter convolutional layer [4].

Suppose input feature map matrix is F, subsampling pooling condomain is $c \times c$ matrix P, biasing is b_2, the subsampling feature map is S and c stands for pooling movement step length. The dimensionality of average pool is being lowerd by local average value calculating. The details are as Formula (2):

$$S_{ij} = \frac{1}{c^2}\left(\sum_{i=1}^{c}\sum_{j=1}^{c} F_{ij}\right) + b_2 \tag{2}$$

The maximum pooling calculating formula is as Formula (3):

$$S_{ij} = \max_{i=1, j=1}^{c}\left(F_{ij}\right) + b_2 \tag{3}$$

PS: $\max_{i=1, j=1}^{c}\left(F_{ij}\right)$ means taking out the maximum element from the $c \times c$ size pooling codomain of input feature map F.

4. Optimization of CNN for Recognition of Vehicle Frame Number Character Based on Key Technologies

4.1. Extension of Network Structure

The traditional Convoluational Neural Network directly takes the gray scale image as the original data and inputs them into the network for training and recognition. Mr. HeSF's team proposes an improved plan with multipath input [5]. Before the training of superpixel by CNN division, the input image should be multi-scale pixel segmented. Then the serial number for 3 path, being applied to recover the space structure matrix and scope structure matrix among superpixel contextual information, is input into the CNN for training.

Carry out the following experiment: select 3000 training pictures of the frame number and 6000 test pictures. The single-path CNN includes two convolutional layers and two pool layers, input the normalized frame number grayscale image. The multi-path CNN input the image calculated by the Sobel operator, the three CNN models are the same as the single-path model. The error rates of the two algorithms are shown in Figure 2.

It can be seen from the experimental results that the multi-path CNN has a lower recognition error rate than the single-path in frame character recognition.
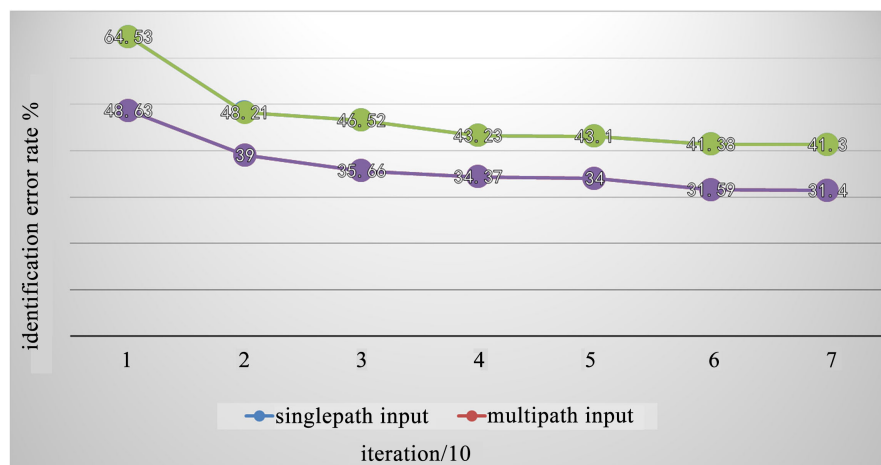

Figure 2. Comparison of multi-path and single-path.

4.2. Activation Function Improvement

As an important part of the CNN, the function of activation function is to adjust the output of convolution layer and make the feature extraction result of every layer meet the requirement of human vision. Sigmoid function and hyperbola tangent function are common and they can only adjust the output rang [6]. As the popularity of sparse representation, human vision system prefers it, or better sparse output can be obtained when Rectified Linear Units (ReLU) is taken as activation function. ReLU is linear correction and as Formula (4):

$$h^{(i)} = \max\left(w^{(i)T}x, 0\right) = \begin{cases} w^{(i)T}x & w^{(i)T}x > 0 \\ 0 & w^{(i)T}x \le 0 \end{cases} \tag{4}$$

The value of convolution calculation will be assigned to be 0 if it is less than 0 while stays the original value. Though this method forces some data to be 0, the experiment has proved that CNN can fully suit such sparse constraint and identification efficiency gets significantly improvement, which proves ReLU is able to guide sparse in a certain level.

4.3. Pooling Model Improvement

The average pooling model and maximum pooling model are the most common classical model, feature extraction with which will damage the precision and presentation of the global feature. Based on the maximum pooling algorithm, a dynamic self-adaption pooling model is proposed to improve the pooling model [7]. Such model can adjust the pooling procedure dynamically according to the different feature map and auto-amend the pooling weight according to the data of every pooling codomain. If there is only one value in the pooling codomain, then this value is the maximum and the representation feature. If the eigenvalues are all same, then its maximum value can be expressed as the eigenvalue. Therefore, on the basis of maximum pooling algorithm and the interpolation principle, a model is made to simulate the function [5]. Suppose μ is pooling factor, then the dynamic self-adaption algorithm can be explained as the Formula (5):

$$S_{ij} = \mu \max_{i=1, j=1}^{c}\left(F_{ij}\right) + b_2 \tag{5}$$

This is basic representation and its nature is to use μ to optimize the maximum pooling algorithm.

For μ, there is a Formula (6) as follows:

$$\mu = \rho \frac{a\left(v_{\max} - a\right)}{v_{\max}^2} + \theta \tag{6}$$

The a is the average of the element in pooling codomain except the maximum value; $v_{\max}$ is the maximum value in pooling codomain; θ is alignment error factor, ρ is characteristic coefficient:

$$\rho = \frac{c}{1 + \left(n_{epo} - 1\right)c^{n_{epo}^2 + 1}} \tag{7}$$

ρ is determined by side length of pooling codomain c and iterations n_{epo}.

When the pooling codomain and iteration cycle remain unchanged, μ can value automatically. μ can dynamically adjust to be the optimal according to the different iterations n_{epo} when it is faced with the same pooling codomain. $\mu \in (0,1)$, this gives consideration to maximum and average pooling and can keep the precision when processes the pooling codomain with a clear maximum feature. What's more, CNN can guarantee the precision of feature extraction when it is faced with different iterations n_{epo} and different pooling codomain as it can weaken the effect of the maximum pooling when it processes the rest co-domains [1].

4.4. Output Layer Improvement

The frame number is generally composed of 10 digits about 0 - 9 and 23 letters about A - Z (except I, O, and Q), there are 33 characters to be recognized, so set the number of output layer nodes to 33.

5. Experiment and Analysis

Selected 5000 pictures as the training set from the frame number database. Firstly, used the principal component analysis technique to reconstruct the image. Then used BP neural network, traditional LeNet-5 CNN and the improved CNN to analyse the same character sets. The experimental results are shown in Figure 3.

It can be seen from the above results that the LeNet-5 CNN has obvious advantages in recognition rate compared with the traditional BP Neural Network. Although the Improved CNN and the LeNet-5 CNN are similar, the former improved the recognition rate at all.

6. Conclusion

In summary, the principle of CNN is briefly introduced. Then the key technical points of CNN are analysed, and the optimized scheme is proposed respectively.
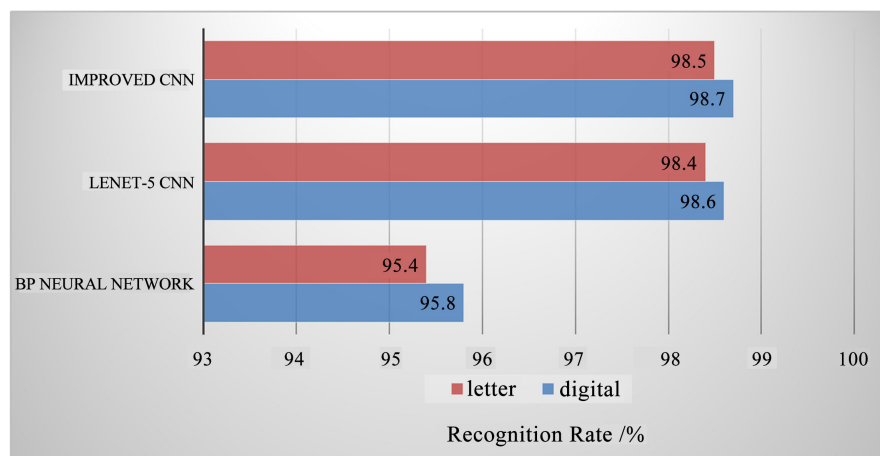

Figure 3. The comparison of recognition rate.

In the frame character recognition, one of them can be used to establish the system, or combine these improvements to create a more perfect model. As one of the widely used deep learning algorithms, the excellent performance in the field of image identification has already assured the ability in feature extraction [8]. In the background that CNN is being improved and updated until the imperfection including the monotonous structure and low rate of convergence is being overcome, the hardware platform develops, the training time is greatly reducing and efficiency boosts by the GPU accelerating, the CNN application in the vehicle frame number character recognition will be more mature and stable.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Liu, W.J., Liang, X.J. and Qu, H.C. (2016) Learning Performance of Convolutional Neural Networks with Different Pooling Models. *Journal of Image and Graphics*, **21**, 1178-1190.

[2] Tian, W.J., Shao, F., Jiang, G.Y. and Yu, M. (2016) Blind Image Quality Assessment for Stereoscopic Images via Deep Learning. *Journal of Computer-Aided Design & Computer Graphics*, **28**, 968-972.

[3] Chen H.C., Cheng, Y. and Zhang, C.Y. (2017) Fine-Grained Vehicle Type Recognition Based on Deep Convolution Neural Networks. *Journal of Hebei University of Science and Technology*, **38**, 564-569.

[4] Wang, Z.M., Cao, H.J. and Fan, L. (2016) Method on Human Activity Recognition Based on Convolutional Neural Networks. *Computer Science*, **43**, 56-58.

[5] He, S.F., Lau, R.W.H., Liu, W.X., *et al.* (2015) SuperCNN: A Superpixelwise Convolutional Neural Network for Salient Object Detection. *International Journal of Computer Vision*, **115**, 331. https://doi.org/10.1007/s11263-015-0822-0

[6] Liu, H.L., Li, B.A., Lv, X.Q. and Huang, Y. (2017) Image Retrieval Based on Deep Convolutional Neural Network. *Application Research of Computers*, **34**, 3817.

[7] Li, Y.D., Hao, Z.B. and Lei, H. (2016) Survey of Convolutional Neural Network. *Journal of Computer Applications*, **36**, 2515-2516.

[8] Hu, Z.P., Chen, J.L., Wang, M. and Zhao, S.H. (2015) New Progress of Convolutional Neural Network Classification Model in Pattern Recognition. *Journal of Yanshan University*, **39**, 284-286.