ORIGINAL ARTICLE

Pre-IdentifyNet: An Improved Neural Network for Image Recognition

Shixiong Chen*, **Xinyang Du, Qingyue Xu, Xiaoru Chen, Jingle Li**

Shijiazhuang Tiedao University, Shijiazhuang 50043, China. E-mail: chenshixiongmail@163.com

Abstract: With the rise and development of artificial intelligence, image recognition and classification technology has received more and more attention as an important branch of its research field. Among them, the introduction of deep learning networks and the construction of neural network structures not only avoid a lot of the tedious work of manual extraction, but also improve the accuracy of image recognition. Convolutional neural networks have many advantages that conventional neural networks do not have. Therefore, image classification systems based on convolutional neural networks emerge in endlessly, but there is still much room for improvement in terms of recognition accuracy and recognition speed. Based on this, this paper proposes an improved deep convolutional neural network to improve the accuracy of the network by changing a series of parameters such as the number of channels of the convolution layer, the size of the convolution kernel, the learning rate, the number of iterations, and the size of the small batch with speed. In this paper, three data sets were selected, namely sewage, animals and the Simpson Family. Comparing the improved convolutional neural network network with the existing SqueezeNet and GoogleNet. It is found that the accuracy of the network is maintained while maintaining a similar speed. Both F1-score and F1-score have been improved with a higher recognition rate and better recognition effect in image recognition classification.

Keywords: Convolutional Neural Network; Image Recognition and Classification; Pre-IdentifyNet

1. Introduction

Image recognition is the next important technology in the current information age. Due to the increase in computer operation speed, image recognition technology has developed tremendously in recent years. Some classic classification algorithms, for example, image classification algorithms based on SVM classifier[1] have been gradually replaced by convolutional neural networks[2]. Among them, the convolutional neural network plays an important role in image recognition. The features extracted by convolution focus more on the locality. It is not necessary for each neuron to perceive the global image. It only needs to perceive the locality, and then the higher layer integrates the local information to obtain the global information. It can share parameters, which greatly reduces the amount of calculation. Therefore, models such as AlexNet[3] and LeNet-5[4] were born, making it possible for image recognition to be widely used in real life.

In recent years, many countries have applied convolutional neural networks[5-7] for image recognition. Kai Zeng *et al*. proposed a multi-convolutional neural network to automatically obtain a local metric map for defocus blur detection[8]. Foo Chong Soon *et al*. used convolutional neural networks based on principal component analysis to extract the main features from existing vehicles for the recognition of vehicle models.

Copyright © 2020 Shixiong Chen *et al*.
doi: 10.18686/esta.v7i2.137
This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous development of artificial intelligence technology, artificial intelligence methods for building deep network structures of multi-level feature learning have achieved great success in the field of large-scale picture classification[11]. In order to improve its speed and accuracy through image recognition, this paper attempts to apply the improved convolutional neural network to image recognition. The Pre-IdentifyNet proposed in this paper accelerates the recognition of images, which improves the accuracy and timeliness of image recognition and processing with less manpower, material resources and time. It also reduces the limitations of manual recognition to realize the automation of finding images.

The rest of the article is organized as follows. The second part briefly summarizes the convolutional neural network and it introduces the improved network in detail. The third part mainly introduces experiments on the designed network, including the data set used in the experiment, image preprocessing, training network and performance evaluation, which gives the experimental results. Finally, the fourth part summarizes the conclusion of this article.

2. Network design

2.1 Convolutional neural network

Taking neurons as a model, convolutional neural network (CNN) is a biologically-inspired artificial intelligence algorithm that obtains input from a layer of cells, performs mathematical transformations, and provides the output to the next group of neurons, which is very suitable for computers visual task.

GoogleNet is carefully prepared by the Google team to participate in the ILSVRC competition, and it has been learned and used by many researchers. GoogleNet proposed that the most direct way to improve deep neural networks is to increase the size of the network, including width and depth. Depth is the number of layers in the network, and width refers to the number of neurons used in each layer. To this end, GoogleNet proposed a new structure, called inception. The entire inception structure is composed of multiple inception modules connected in series. There are two main contributions of the inception structure: one is to use 1x1 convolution to perform up-down dimension; the other is to perform convolution and aggregation on multiple sizes at the same time.

SqueezeNet was proposed by UC Berkely and others in 16 years and it is one of the current mainstream convolutional neural networks. The network is a network model that can reduce the input volume for the advanced AlexNet and VGG Net models with increasing parameters. The core part of the model is the Fire Module. The structure is divided into squeeze and expand structures. Squeeze contains S 1×1 convolution kernels, and the expand layer contains E1 1×1 kernels and E3 3×3 convolution kernels, and the model satisfies $S < (E1 + E3)$. The model reduces the size of the network convolution kernel on the basis of the AlexNet network, and it replaces the fully connected layer with an average pooling layer, thereby maximizing the calculation speed, but it may reduce the model accuracy and model parameters.

2.2 Network design

In this study, MATLAB's Deep Net Designer toolbox was used to synthesize SqueezeNet and GoogleNet. The main job is to test the pre-trained neural network with different numbers of modules to determine the number of fire modules and inception modules.

In order to reduce the computational complexity, this study only analyzes and designs the shallower network. The 8 combinations of designs are shown in **Table 1**. According to the data in the table, when a fire module is combined with two inception modules, the accuracy of picture recognition is the highest.

Fire Module	Inception Module	Accuracy
0	1	65.26%
1	0	57.85%
1	1	86.13%
1	2	95.83%
1	3	92.04%
2	1	83.38%
2	2	89.93%
3	1	87.69%

Table 1. Effect of different number of fire module and inception module on recognition accuracy

Subsequently. Based on the number of fixed modules, a fire module and two inception modules are arranged and combined, and it is found that the fire module is in front of the two inception modules with the highest accuracy rate, and the improved neural network Pre-IdentifyNet is obtained. The specific structure is as follows **Figure 1.**

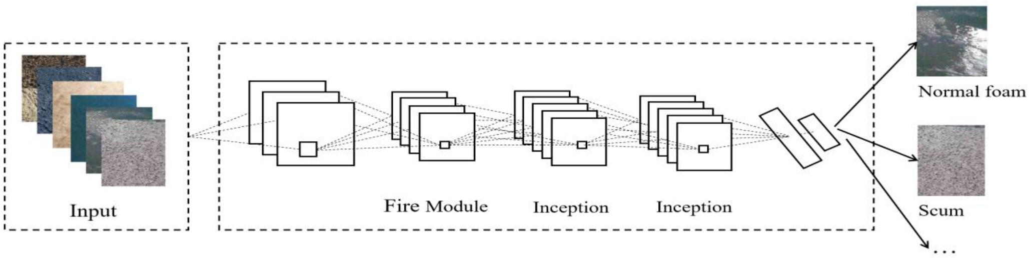

Figure 1. The structure of a neural network.

3. Experimental design

3.1 Data set

In order to accurately obtain the recognition situation of the network in the face of different types of pictures, this study used three completely different types of data sets, namely simpsons_dataset, animals_dataset and sewage dataset. Among them, simpsons are images captured by different animated characters of the Simpsons at different angles, animals are images collections of different living conditions of various animals in nature, and sewage datasets are images of abnormal working conditions of the water treatment process obtained from sewage treatment plants set.

The images of abnormal working conditions in the water treatment process are divided into six categories: foam, normal foam, normal working condition, scum, scum sludge, silt swelling, simpsons_dataset and animals_dataset are also divided into six categories in the above manner, and they are divided into 8:2 ratios respectively training set and test set.

In order to further improve the accuracy of the built convolutional neural network, this study also uses image processing software to segment the sample image, vertically flip left and right, random translation and other operations, so as to obtain additional training data with these image pairs of the pre-trained network performs secondary training to improve the accuracy of pre-recognition.

3.2 Training network

The next step is to use the pre-processed image to retrain the network. In the designed network, the freeze weights function is used to freeze these initial layers to suppress overfitting during training.

The gradient descent method is used to train the designed convolutional neural network. In addition, through experiments, the Mini Batch Size is adjusted to 100, which reduces the training time and improves accuracy. The relationship between Validation Frequency and Mini Batch Size is shown in Equation 1. In **Table 2**, and the setting values of other training parameters are given.

$$\text{ValidationFrequency}=\left\lfloor \frac{\text{numel(augimdsTrain.Files)}}{\text{MiniBatchSize}} \right\rfloor \quad (1)$$

Parameter	Setting
Initial Learn Rate	0.00006
Mini Batch Size	100
Max Epochs	6

Table 2. Network training parameter settings

3.3 Performance evaluation

This section calculates the accuracy, precision, recall, and F1-score of the network by introducing four indicators: TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative).

Among them, precision refers to the proportion of samples that are positive in the positive prediction:

$$precision = \frac{TP}{TP + FP} \tag{2}$$

Recall is the proportion of samples that are predicted to be positive in the overall sample:

$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1\text{-}score = 2\frac{precision \cdot recall}{precision + recall} \tag{4}$$

The Simpsons	accuracy	precision	recall	F1-score
Pre-IdentifyNet	95.83%	0.96	0.95	0.95
GoogleNet	79.67%	0.77	0.83	0.79
SqueezeNet	71.00%	0.76	0.70	0.72

Table 3. Performance comparison of different networks in the Simpsons

Animals dataset	accuracy	precision	recall	F1-score
Pre-IdentifyNet	94.79%	0.94	0.96	0.94
GoogleNet	80.62%	0.82	0.79	0.81
SqueezeNet	73.51%	0.74	0.75	0.74

Table 4. Performance comparison of different networks in animals dataset

Abnormal condition	accuracy	precision	recall	F1-score
Pre-IdentifyNet	96.31%	0.96	0.97	0.95
GoogleNet	79.67%	0.73	0.82	0.80
SqueezeNet	75.35%	0.77	0.76	0.76

Table 5. Performance comparison of different networks in abnormal condition image

In the same test environment, the Simpsons, animals, and sewage treatment abnormal working conditions datasets were used to train the Pre-IdentifyNet for many times, they are compared with the use of GoogleNet and SqueezeNet. The results are shown in **Tables 3~5**. It can be seen that the accuracy and F1-score of the design network are higher, which has achieved better results than GoogleNet and SqueezeNet. **Figure 2** shows the recognition effect of four randomly selected pictures in the sewage treatment dataset. The accuracy of using Pre-IdentifyNet is above 95%, which can effectively improve various performance indicators of image classification to achieve good experimental results.

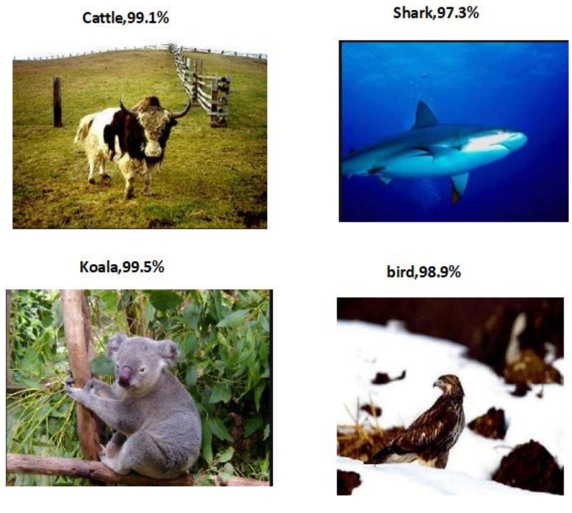

Figure 2. Animal recognition results.

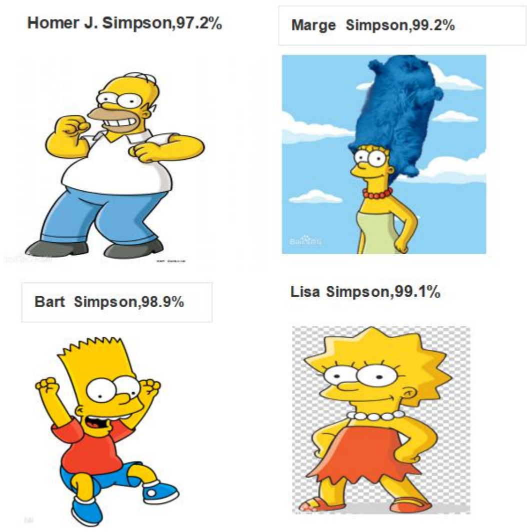

Figure 3. Simpson's recognition results.

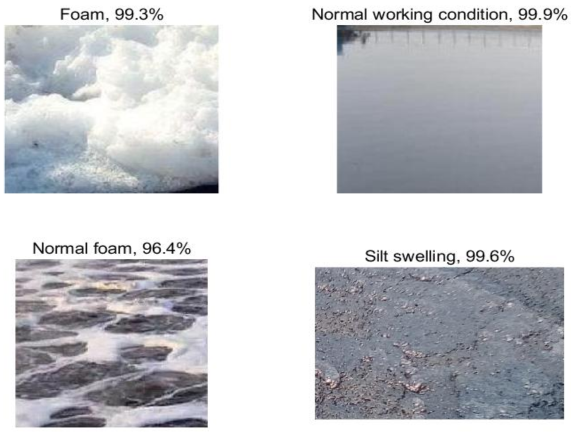

Figure 4. Identification results of abnormal sewage conditions.

4. In conclusion

Based on SqueezeNet's fire module and GoogleNet's inception, this paper constructs an improved convolutional neural network---Pre-IdentifyNet, which is used to identify animals, cartoon images of the Simpson family, and abnormal working conditions in sewage. During the training process, an attempt is made to gradually approach and determine a series of parameter values such as the learning rate, the number of iterations, and the size of the small batch. On the sample dataset, training results superior to the original network are obtained.

In addition to identifying abnormal conditions in wastewater treatment, the Pre-IdentifyNet designed in this study can also be applied in various fields such as image recognition, item classification, and behavior recognition, pose estimation, and other fields.

In the following research, further researches need to be focused on, including improving the accuracy of the network and the ability of network feature extraction by constructing a more scientific and reasonable structure, further optimizing relevant network parameters to simplify the network structure, and trying to optimize the Pre-IdentifyNet network used in more scenarios.

References

1. Liu L, Ma Y, Zhang X, *et al*. High discriminative SIFT feature and feature pair selection to improve the bag of visual words model. IET Image Processing 2017; 11(11): 994-1001. doi: 10.1049/iet-ipr.2017.0062.
2. Jiang Y, Chi Z. A CNN model for semantic person part segmentation with capacity optimization. IEEE Transactions on Image Processing 2018; 28(5): 2465-2478. doi: 10.1109/TIP.2018.2886785.
3. Alex K, Ilya S, Geoffrey H. ImageNet classification with deep convolutional neural networks. NIPS 2012. doi: 10.1145/3065386.
4. Mopuri KR, Garg U, Babu RV. CNN fixations: An unraveling approach to visualize the discriminative image regions. IEEE Transactions on Image Processing 2017; 1-11. doi: 10.1109/TIP.2018.2881920.
5. Ge L, Liang H, Yuan J, *et al*. Real-time 3D hand pose estimation with 3D convolutional neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 2018; 1-14.
6. Hou Q, Cheng M, Hu X, *et al*. Deeply supervised salient object detection with short connections. IEEE Transactions on Pattern Analysis and Machine Intelligence 2018; 1-14.
7. Zeng K, Wang Y, Mao J, *et al*. A local metric for defocus blur detection based on CNN feature learning. IEEE Transactions on Image Processing 2019; 28(5): 2107-2115. doi: 10.1109/TIP.2018.2881830.
8. Soon FC, Khaw HY, Chuah JH, *et al*. PCANet-Based convolutional neural network architecture for a vehicle model recognition system. IEEE Transactions on Intelligent Transportation Systems 2019; 20(2): 749-759. doi: 10.1109/TITS.2018.2833620.