Research on image classification algorithm based on depth residuals shrinkage network in Commercial Image Library

Jiantao Zhao, Wenxin Chen*
(School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)
chenwenxin@eplian.com

Abstract With the advent of the age of multimedia technology and image data, commercial image library services have emerged. We need to use image classification to parse into content that the computer can understand. But when recognizing image features for classification, image data will inevitably produce noise, so in order to make image classification more accurate, we need to remove information that is irrelevant to noise. The Deep Residual Shrinkage Network (DRSN) proposed in this paper can effectively denoise images, and then learn features to classify images. The Deep Residual Shrinkage Network (DRSN) combines the Deep Residual Network (ResNet) and the SENet commonly used in the attention mechanism, and adds a soft threshold function to it. The image data set used in this experiment was downloaded from Panorama Network and Baidu Pictures and then processed. In the experiment, noise was first manually added to the data set, and the deep residual network and the deep residual shrinkage network were used for comparison experiments. It is concluded through experiments that the improved DRSN network has higher accuracy when applied to image classification.

1. Introduction
Because of the standardization of Internet technology, copyright awareness has gradually increased, Image acquisition channels also need to be standardized, so commercial image gallery services came into being. Sophisticated image classification can analyze the content of the image and provide feedback in sentences that humans can understand. Convolutional neural network plays an important role in image classification, and it is very helpful for image recognition to extract image features. The deep residual network can solve the problem of gradient disappearance on the basis of deepening the number of network layers to make the classification more accurate. But for noisy picture data, this article will improve the depth residual network after the deep residual shrinkage network to improve the accuracy of image classification more. This network combines deep residual network (ResNet) and SENet, which is commonly used in attention mechanism, and adds a soft threshold function to it, which greatly improves the accuracy of image classification.

2. Research theories and methods

2.1. Deep Residual Network
Deep Residual Networks (ResNet) is a feed-forward neural network, which has a good effect in image recognition. Compared with the Deep Neural Network (DNN), it reduces the complexity and scale of the model and is one of the most commonly used CNNs.
Theoretically speaking, increasing the number of network layers can increase the network's presentation ability and improve classification accuracy. However, in practice, this is not the case due to the vanishing and exploding gradients. When the network layer is deepened, the gradient in the transmission process will explode and disappear due to the gradual decrease. The original intention to continue to adjust the weight of the previous network layer cannot be effectively implemented [1].

At this time, a deep residual network appears, and ResNet can effectively solve this problem through residual blocks, thereby greatly increasing the number of network layers [2]. At the same time, ResNet accelerates the training speed of deep networks. While increasing the depth of the network, it can handle short-term sequences better than traditional CNN models. It can contain quite deep convolutional layers with "residual connections", while alleviating the problem of vanishing gradients in traditional CNN architectures.

The basic structure is shown in Figure 1 below. It can be seen that the picture has a jump structure. Through the "shortcut connections" method, the input is x, and the output is H(x) = F(x) + x. In this way, the residual F(x) = H(x) - x can be obtained. Therefore, the following training goal is to approach the residual result F(x) to 0, and achieve the identity mapping between the target value and the input value, so that as the network deepens, the accuracy rate does not decrease.

![Residual module diagram](image)

**Figure 1** Residual module

2.2. **Soft threshold function**

Soft thresholding is a very common concept in signal noise reduction [3]. It refers to shrinking the value of a segment of signal toward the "zero" direction. As shown in Figure 2, the horizontal axis x represents input, and the vertical axis y represents output. Then, compared with the input signal, the output signal develops toward "zero" and shrinks.

![Soft threshold value output signal](image)

**Figure 2** Soft threshold value output signal
This method of noise reduction has a premise. That is, the part close to zero is noise, or in other words, it is not important and can be eliminated. However, in fact, for many signals, the part close to zero may contain a lot of useful information and cannot be eliminated directly. Therefore, it is usually not directly processed by soft thresholding of the original signal.

In view of the above problem, the traditional way of thinking is to transform the original signal into other forms of representation. Than soft thresholding is used to process the converted characterization[4]. Finally, the characterization after soft thresholding is reconstructed to obtain the signal after noise reduction.

However, this method of signal noise reduction has some unresolved problems. First of all, in wavelet analysis, how to construct the wavelet function that is most suitable for the current signal, or filter, local filter operator, has always been a very difficult problem. Secondly, how to set the soft threshold is also a very difficult problem.

2.3. Attention mechanism

Attention has become one of the most important concepts in the field of deep learning. Because we often have information redundancy when processing data, and the data is noisy, the attention mechanism can effectively help us reduce this problem. It uses limited processing resources to quickly select high-value information from the massive amount of information.

SENet is the abbreviation of Squeeze-and-Excitation Networks. It is one of the commonly used algorithms of the attention mechanism. The function of SENet is to add weight coefficients to a feature channel to make the feature focus rate different. Its weight coefficients are learned through learning.

As shown in Figure 3, SENet and residual module can be integrated. For each output channel, first perform a global average pool, each channel will have 1 scalar, C channels have C scalars, and then go through a network layer such as fully connected layer-ReLu activation function-fully connected layer-Sigmoid threshold function The structure obtains C scalars, the scalar range is between 0 and 1, as the weight of the channel, the original output channel of each output channel is weighted with the corresponding weight (the channel corresponding to each element is multiplied by the weight), and the new weight is obtained characteristic. The attention mechanism is actually to learn a weight distribution, and then apply this weight distribution on the original features to recalibrate the features.

![Figure 3 Integration of SENet and residual module](image-url)
3. Deep residual shrinkage network model construction

3.1. Data overview and display

The image dataset used in this article is a collection of images from QuanJing website and Baidu image crawler. QuanJing website is a commercial image library website, and Baidu image is a free image library. There are 10 categories of reptile pictures, including dogs, sea, maple leaves, trees, cats, elephants, basketballs, flowers, Building, and cars. Each category contains 1000 image feature data, and all 10 categories have a total of 10,000 image data. These image categories are mutually exclusive, and the class name label ranges from 0-9 numbers. The 10,000 image data is divided into 1,000 test data and 9,000 training data. Before the experiment, the image is processed first, and its size is processed into a 32*32 size in batches, and the processed image is then used for the experiment. The corresponding relationship between the image label and the actual category is shown in Table 1.

| label | category   | label | category   |
|-------|------------|-------|------------|
| 0     | dog        | 5     | Elephant   |
| 1     | The sea    | 6     | basketball |
| 2     | Maple leaf | 7     | Flowers    |
| 3     | tree       | 8     | Building   |
| 4     | cat        | 9     | car        |

Because the main idea of this article is to use the residual shrinkage model that is fused with soft thresholding and attention mechanism to remove image noise and then classify, so during the experiment, the noise will be manually added to the processed In the image data set, a new feature image data set is formed, and then the noise is reduced through the constructed model, and finally the classification effect is achieved.

During the use of the data set, it is divided into four training batches and one test batch. Each batch contains 10 categories and a total of 2000 images. Each time an experiment is conducted, 1,000 feature images of each category are randomly selected for experimentation.

3.2. Construction of Deep Residual Shrinkage Network

When identifying image features for classification, image data may inevitably produce noise due to unclear photos, excessive content in the image, and so on. Therefore, in order to make image classification more accurate, we need to remove information that is irrelevant to noise. The deep residual shrinkage network draws on the sub-network structure of the deep learning model of the attention mechanism, and adaptively sets the threshold for the feature channel, which makes the image classification more accurate.

The deep residual shrinkage network constructed in this paper includes an input layer, a relu activation function, 4 residual shrinkage modules (residual_shrinkage_block), a batch normalization layer (batch_normalization), convolution layer (conv_2d), and a global average pooling (global_avg_pool) And a fully connected layer (fully_connected) and output layer[5].

The network structure is shown in Table 2.

| Network structure | type           | Network structure | type           |
|-------------------|----------------|-------------------|----------------|
| input_data        | Data input layer | RSBU-CW           | Residual shrinkage module |
| conv_2d           | Convolutional layer | batch_normalization | Batch standardization |
| RSBU-CW           | Residual shrinkage module | global_avg_pool | Global pooling |
| RSBU-CW           | Residual shrinkage module | fully_connected | Fully connected layer |
| RSBU-CW           | Residual shrinkage module | Output layer | |

First, the input layer image feature size of the input layer is set to 32*32*3, where 32*32 is the image size, and 3 represents the three RGB channels of the image. In the convolution layer, the step size of the
convolution kernel is set to 2, and the feature image The width is halved to achieve image compression. The number of convolution kernels is set to 1, and the number of channels of the feature image is still 3.

The residual shrinkage network mainly includes 4 RSBU-CW residual shrinkage modules. Compared with the ordinary residual module, the advantage of RSBU-CW is that the threshold can be set adaptively. The threshold it sets is not a value, but a vector, that is, each channel of the feature map corresponds to a shrinkage threshold. Using the relu activation function is more conducive to learning and extracting the nonlinear feature relationship of the feature image, and the accuracy of classification is improved. After global average pooling, the structure of the entire network is regularized to prevent overfitting. The fully connected layer uses the softmax function classifier to output the probability value of a certain image belonging to a certain category. The larger the value, the greater the probability of belonging to a certain category. Finally, the output layer outputs the image type.

4. Experimental results and analysis

After processing the image data set obtained by the crawler, perform experiments according to the training data set 10*100 and the test data set 10*900, epoch=100, a total of 100 trainings, that is, a total of 100 trainings for all samples. The batch_size is 100, that is, the size of each batch of data is 100, so a total of 100 iterations are required.

The deep residual shrinkage network is improved on the basis of the deep residual network. First, the deep residual network is used for experiments to obtain the success rate of the training set and the test set.

The results are shown in Table 3:

| Methods                  | Deep residual network |
|--------------------------|-----------------------|
| Training accuracy        | Test accuracy         |
| 87.78%                   | 83.99%                |

The deep residual shrinkage network after the improvement of the residual block, after 100 trainings with the same parameters, the running results obtained are shown in Figure 4:

![Figure 4 Operation results of the deep residual shrinkage network](image)

The success rate of the training set and test set is shown in Table 4:

| Methods                  | Deep residual shrinkage network |
|--------------------------|---------------------------------|
| Training accuracy        | Test accuracy                   |
| 94.72%                   | 94.90%                          |

After the experiment, the loss rate and success rate of 100 training results are plotted into a graph, and the graph shown in Figure 5 is obtained. As the number of training increases, the loss rate gradually decreases and the accuracy rate gradually increases. After 100 trainings, the loss rate was reduced to 20.38%, and the correct rate reached 94.61%. As shown in Figure 5:
Figure 5  Loss rate and success rate of deep residual shrinkage network

It can be clearly seen from Picture 5 that the deep residual shrinkage network has great advantages over the ordinary residual network. After adding a soft threshold function to the attention mechanism, it is easier to identify noise, thereby reducing noise and facilitating image classification. Similarly, for pictures without noise data, the deep residual shrinkage network can better learn and classify the features of the pictures.

5. Conclusion

With the advent of multimedia technology and the Internet era, image data has become a widely demanded data. But in real life, image data may inevitably produce noise due to unclear photos, too many elements in the image, etc. Many image data have interference information and noise. In order to eliminate noise and facilitate image classification, This paper proposes an image classification method based on deep residual shrinkage network.

This deep residual shrinkage network combines the deep residual network RSN and the SENet, which is commonly used in the attention mechanism, and adds a soft threshold function to it, which greatly improves the accuracy of image classification. The image data set used in this experiment was downloaded from the QuanJing website and Baidu pictures, and then the image data was batch-processed into a fixed size, and the data set was manually added with noise. After the deep residual network RSN Through comparative experiments, it is found that the accuracy of the deep residual shrinkage network has been greatly improved, and it has broad application space in many places.

Acknowledgments

This work was supported by the Institute of Network and Information Systems of North China Electric Power University.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for im-age recognition,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR), Seattle, W A, USA, Jun. 2016, pp. 770–778.
[2] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in Computer Vision—ECCV 2016 (Lecture Notes in Computer Science 9908), B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham, Switzerland: Springer, 2016, pp. 630–645.
[3] D. L. Donoho, “De-noising by soft-thresholding,” IEEE Trans. Inf. Theory, vol. 41, no. 3, pp. 613–627, May 1995.
[4] K. Isogawa, T. Ida, T. Shiodera, and T. Takeguchi, “Deep shrinkage convolutional neural network for adaptive noise reduction,” IEEE Signal Process. Lett., vol. 25, no. 2, pp. 224–228, Feb. 2018.
[5] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, “Deep Residual Shrinkage Networks for Fault Diagnosis,” IEEE Transactions on Industrial Informatics, 2019, DOI: 10.1109/TII.2019.2943898