Exploration of Neural Network Optimization Methods Based on LeNet-5

Yifang Pang

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

2020080602018@std.uestc.edu.cn

ABSTRACT. In today's society, the application of artificial neural networks in the field of image classification is becoming increasingly widespread. However, the exploration of how to improve the classification accuracy of neural networks has never stopped. This paper is based on the most classic neural network LeNet-5 and proposes three methods to optimize the network, observing its classification performance on the image dataset. The three methods are to increase network depth, add dropout mechanism, and use CBAM attention mechanism. For the experimental indicators, this paper chooses to use the Loss function, accuracy and recall to verify the effect of image classification. After comparing the experimental results, this paper draws the corresponding Line chart to observe the change trend, and conducts visual clustering analysis of the accuracy of each category classification. Finally, this work found that all three corresponding optimization methods have an improvement effect on the network, with the dropout and Attention mechanisms being the most obvious.

Keywords: Deep Learning, LeNet-5, Dropout, Attention, Neural Network

1 Introduction

In today's society, the application of artificial neural networks is becoming increasingly widespread, and people's discussions and topics are also climbing. As the cornerstone of machine learning and deep learning, artificial neural network is an algorithm model that processes information in a distributed manner and also utilize neurons to learn features and weights [1]. It simulates a specific function through numerous and complex neurons, continuously fitting the correct output results while adjusting the relationships and parameters between neurons to achieve practical effects such as prediction and classification. Its massively parallel processing, distributed storage, elastic topology, high redundancy and nonlinear computing capabilities enable itself to show excellent accuracy and high efficiency in some classification, prediction and image processing tasks [2].

Although some neural networks with higher completion rates have performed well in some tasks, the exploration of using different methods to improve the network's

P. Kar et al. (eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*, Advances in Computer Science Research 108,
https://doi.org/10.2991/978-94-6463-300-9_75

performance is never stopped. Among many artificial neural networks, LeNet-5 is one of the most classic and simple. It was proposed by Yann LeCun in 1998 to solve the problem of handwritten digit recognition, and is considered to be one of the pioneering works of Convolutional Neural Network [3]. This network is one of the first neural networks widely used in the digital image recognition field. Although LeNet-5's network structure is relatively simple, it is still a crucial reference for studying neural network optimization methods. Through in-depth understanding of LeNet-5, this article uses three different optimization methods to modify the network structure, namely deepening the network depth of LeNet-5, adding dropout mechanism, and adding attention mechanism. Afterwards, compare the results with the original LeNet-5 to observe its performance on handwritten digit datasets.

2 Methods

To analyze the influence of adding different mechanisms and structures on the final classification performance of neural networks, this paper proposes three corresponding methods based on LeNet. Fig.1 shows the workflow of the research in this paper.

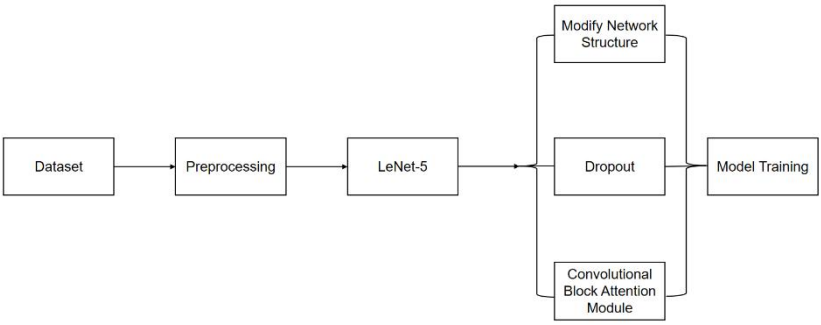

Fig. 1. Research Workflow（Photo/Picture credit: Original）

Before introducing the methods, it is very important to have a simple understanding of the LeNet-5. This experiment used an improved LeNet-5 structure. The basic structure of LeNet-5 includes a 7-layer network structure (excluding the input layer), which includes 2 convolutional layers, 2 down sampling layers (pooling layer), 2 fully connected layers, and an output layer. On this basis, the three specific design methods are as follows.

2.1 Modify Network Structure

The input layer receives handwritten digital images with a size of 32 * 32, including grayscale values of 0-255. Also, this experiment normalizes the pixel values to accelerate training speed and improves the accuracy of the model. The main component of LeNet is the convolutional layer. One of the most important advantages of convolutional layers is that they can maintain the shape of the input image

unchanged. Unlike fully connected networks that require the image to be flattened into a one-dimensional array form, convolutional layers can directly receive the 3D data form and output the original form to the next layer without losing key feature information. Therefore, in LeNet-5, it is possible to correctly understand data with shapes such as images.

The convolutional operation in convolutional layers is the reason why convolutional layers can perform feature extraction. Here, this paper considers the case of discrete multidimensional convolution, which is also the most common situation in the field of machine learning. The input is a multidimensional array, and the convolution kernel is also a multidimensional array, which is discrete in time. Therefore, infinite integrals become the sum of finite elements in a finite array:

$$H(i,j) = \sum_{m} \sum_{n} F(m,n)G(i-m,j-n) \tag{1}$$

Because the neurons in each convolutional layer are only connected to a portion of the neurons in the previous layer, the previous layer only transmits local information to that neuron. At the same time, the neurons in this layer will only be locally connected to the neurons in the next layer, so as to achieve the purpose of local Receptive field. LeNet-5 contains two convolutional layers: The convolutional layer C1 includes 6 convolutional kernels, each with a size of 5 * 5, the step size is 1, and the padding is 0. While the convolutional layer C2 includes 16 convolutional kernels, and the other parameters are consistent with the convolutional layer C1. Therefore, each convolutional kernel will generate a feature map with a size of 10 * 10.

The convolutional layer construction of LeNet-5 performs well when dealing with smaller datasets, but due to its small number of convolutional kernels and shallow network depth, it cannot handle more complex datasets well. Therefore, this work increases the depth of the network, raises the number of convolutional kernels (i.e., the number of output channels), and optimizes the construction of a deeper network. This experiment added new convolutional layers and hidden layers after maximizing pooling in the second convolutional layer C2. It also designed the new convolutional layer C3 with 16 input channels and 36 output channels, which greatly deepens the complexity of the network. Considering the feature map size of the input image, the convolution kernel size is chosen as 3 * 3. After the maximum pooling layer, this work set a hidden layer. The number of neurons in this hidden layer is 64.

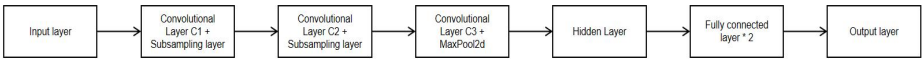

Fig. 2. Modified Network Structure (Photo/Picture credit: Original)

2.2 Dropout

In the training process of the original LeNet-5, due to the excessive number of neurons in the model, there were many hyperparameters that needed to be optimized, and the training data environment was too simple, resulting in overfitting after a

certain number of iterations. This phenomenon can cause experimental indicators such as accuracy to perform exceptionally well on the training set, while performing poorly on the test set, indicating that the model cannot perform classification tasks well. After deepening the depth of the network model, this phenomenon becomes more apparent. Therefore, this paper uses the dropout mechanism to reduce the overfitting degree of the model [4].

The Dropout mechanism mainly plays a role in reducing overfitting and enhancing generalization ability by stopping the work of some neurons. During each forward propagation process, dropout deactivates the neuron with a certain probability of p, thus losing some local features to a certain extent. This approach can effectively reduce the complex synergy and interdependence between neurons, as dropout may result in two neurons not necessarily being in the same neural network every iteration. Therefore, the update of weights will not be as rigid and strongly dependent on certain specific neurons as the original network, reducing the phenomenon of interdependent learning between weights and forcing the network to learn more robust and versatile features.

In this experiment, the dropout layer is added after the final maximum pooling layer and before the fully connected layer. This can not only simplify the neural network appropriately, but also maximize the retention of key information. It will not result in poor performance for a particular class or classes when outputting softmax results in the last fully connected layer. In this experiment, a dropout p of 0.5 was selected to ensure the strongest regularization effect (Fig.3).

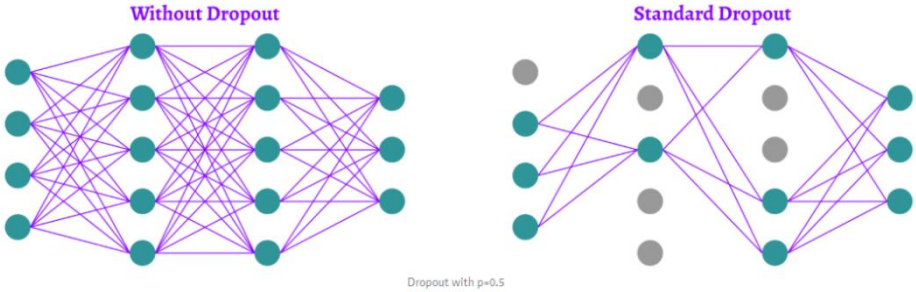

Fig. 3. Dropout with p=0.5 [4]

2.3 Convolutional Block Attention Module

The attention mechanism also performs well in image processing related tasks [5-6]. Given the good results of attention mechanisms in various fields of artificial intelligence today, this experiment adopts the Convolutional Block Attention Module (CBAM), which combines spatial attention and channel attention [7].

CBAM module constructs an attention module of forward Convolutional neural network with simple structure and good performance by combining channel attention and spatial attention mechanisms (Fig.4). This experiment inputs the intermediate output features of LeNet-5 as feature maps into the module to generate inferential attention maps corresponding to two dimensions. Finally, the generated inferential attention map is combined with the intermediate output feature map to enhance the

model's adaptability and generalization ability. Implement end-to-end training without affecting efficiency.

The attention module in the channel domain utilizes the inter channel relationships of features to generate channel attention maps [8]. The input is a C-dimensional feature map, and the output is a 1x1xC channel attention map. Firstly, this work preprocesses the input feature maps to a certain extent, using average pooling and maximum pooling methods to remove redundant data information and reduce the complexity of the input model. By integrating the information in this way, two C-dimensional pooled feature maps F_ Avg and F_ Max can be obtained. Then, this experiment will send the F_ Avg and F_ Max to the Multilayer perceptron containing the hidden layer and get two 1x1xC channel attention maps. Furthermore, so as to reduce the number of neurons and appropriately reduce the complexity of the model, C/r is set to the number of hidden layer neurons. Finally, this work adds the corresponding features of the two channel attention maps and gets the weight coefficient Mc through a sigmoid Activation function.

$$M_c(\boldsymbol{F}) = \sigma(\boldsymbol{W}_1(\boldsymbol{W}_0(\boldsymbol{F}^c_{avg})) + \boldsymbol{W}_1(\boldsymbol{W}_0(\boldsymbol{F}^c_{max}))) \tag{2}$$

The Spatial Attention Module generates a spatial attention map using the spatial relationships between features, with the input being Channel refined feature F 'and the output being a H*W spatial map. Firstly, for F ', after data preprocessing, this work uses the same operations as channel attention along the attention channel direction to get two feature maps F_ Avg and F_ Max with attributes of 1xHxW. By concatenating the two two-dimensional feature maps obtained, it was found that the concatenated feature maps reduced redundant information, making it easier for us to proceed with the next step of feature extraction. Then, for the obtained concatenated feature maps, this experiment uses a convolutional layer of size 7x7 to generate the combined spatial attention map Ms.

$$M_s(\boldsymbol{F}) = \sigma(\boldsymbol{f}^{7*7}([\boldsymbol{F}^s_{avg}; \boldsymbol{F}^s_{max}])) \tag{3}$$

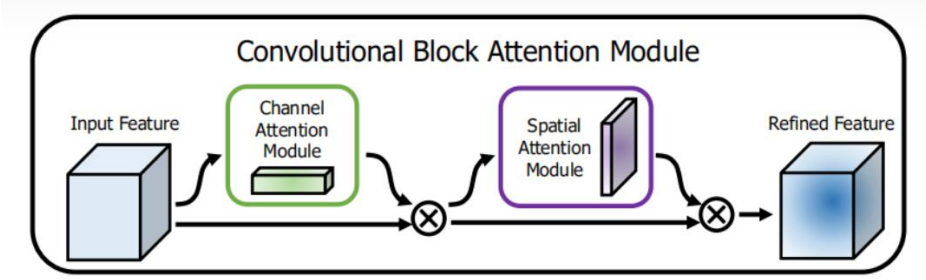

Fig. 4. Convolutional Block Attention Module Structure [7]

3 Results

3.1 Dataset introduction

The dataset used in this experiment is the CIFAR-10 dataset [9]. It has a total of 60000 samples, each of which is a 32 * 32 pixels RGB image, and each RGB image has three corresponding channels as the object of our feature learning. Also, it is used for supervised learning training, so each sample must be equipped with a tag value. Different types of objects use different tag values. There are 10 types of objects in CIFAR-10, and the tag values are distinguished according to 0~9. They are aircraft, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. Due to the limitations of the experimental hardware platform, this experiment selected 50000 pieces of data for training. Divide the training and testing sets in a 4:1 ratio, with a total of 40000 training data and 10000 testing data.

3.2 Experimental details

This experiment was run on a local computer platform. The CPU model is Intel Core i7-11800H fourteen cores. The graphics card model is RTX3060, and the graphics memory is 16GB. The configuration basically meets the requirements of the experiment.

The batch size for this experiment is set to 128. Due to the size of computer graphics memory, the batch size setting should not be too large. 128 can accelerate training speed and improve training efficiency while meeting graphics memory requirements. Then this work sets the Loss function as the Cross Entropy Loss function, and the optimization method is the small batch gradient descent. However, the Gradient descent cannot be set too large when setting the Learning rate. In order to prevent the parameters in the vertical direction from being updated too much, such a small Learning rate causes the parameters in the horizontal direction to be updated too slowly, so the final convergence is very slow. Therefore, this work adopts the momentum method, and each time it updates parameters, this method takes into account the previous velocity. The movement amplitude of each parameter in each direction depends not only on the current gradient, but also on whether the past gradients are consistent in all directions. Finally, the Learning rate is set to 0.01 and the parameter of momentum method is set to 0.9.

3.3 Comparison of experimental results

In this experiment, the loss function, accuracy rate and recall rate are used as experimental indicators. All methods use the same dataset, experimental parameters (batch size, learning rate, momentum method parameters, etc.), epochs (100), and a comparison table is drawn as follows:

Table 1. Model Evaluation Results

Model	Loss	Accuracy	Recall
LeNet-5	40.34	0.48	0.53
Deeper Network	32.91	0.51	0.55
Dropout	9.25	0.64	0.67
CBAM	8.78	0.72	0.85
Combination	22.14	0.55	0.61

From the table 1, it can be observed that all three methods and combinations are to some extent superior to the original model LeNet-5. The effect of deep network on classification is not significantly improved. In fact, the Loss function, accuracy and recall rate are slightly different from the original network structure. The possible reason is that the dataset itself is relatively simple, and the overly complex network leads to serious overfitting and network degradation. The learned parameters are too many, resulting in poor classification performance. The Dropout method effectively reduces the overfitting degree of the network by discarding some neurons, enhances generalization ability, and therefore results in more robustness. On the basis of the original LeNet-5 model, there has been a certain improvement in accuracy and recall. Afterwards, the attention mechanism multiplied the attention map and the input feature map for adaptive feature refinement, and learned corresponding weights for each parameter, thus achieving the best results in this image classification task. Finally, the results obtained by combining the corresponding three methods were not satisfactory. Performance is close to that of deep networks and LeNet-5. The speculated reason is that the model network structure is too complex, the dataset itself is relatively simple, and too many useless features have been learned, resulting in a decrease in the model's generalization ability, resulting in poor performance in the end.

The optimization of the model can also be clearly noticed through the classification accuracy curve of the training set and the classification accuracy curve of the test set. The dropout with the most obvious effect is selected for comparison with the original LeNet-5 to observe the improvement of generalization ability. Since the accuracy of the test set does not change significantly after 40 epochs, this paper only uses the first 40 iterations to obtain the most intuitive results (Fig.5-8).

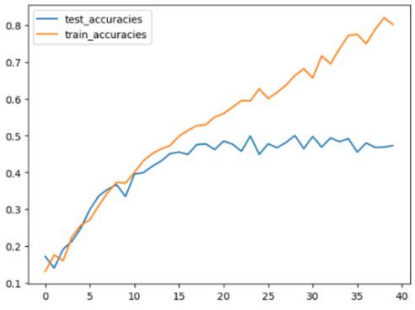

Fig. 5. LeNet-5 Loss (Original)

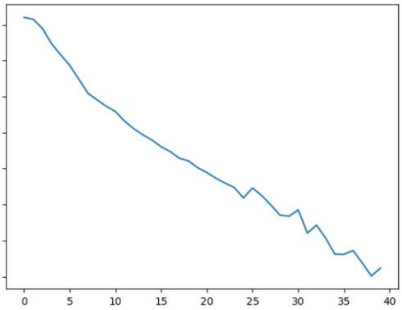

Fig. 6. LeNet-5 Accuracies (Original)

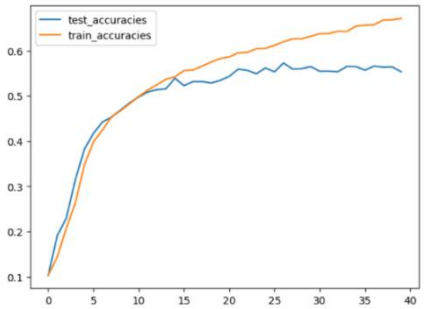

Fig. 7. Dropout Loss (Original)

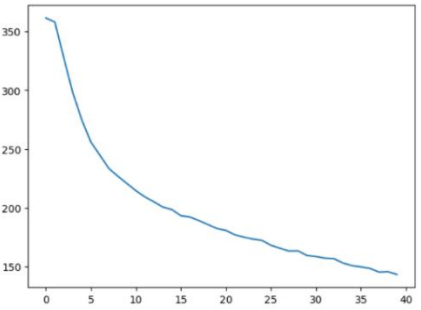

Fig. 8. Dropout Accuracies (Original)

As shown in the Fig.6,8, during the first 10 iterations, the classification accuracy of the training set and the test set both steadily increased, and their curves basically overlapped. After 20 iterations, the accuracy of the original LeNet-5 test set remained basically stable, fluctuating around 0.45. The accuracy of the training set is still steadily increasing, reaching around 0.79 in 40 iterations. The accuracy of the training set is much higher than that of the test set, resulting in severe overfitting. The dropout mechanism alleviates the overfitting phenomenon. During 40 iterations, the difference in classification accuracy between the test set and the training set is small, and that of the test set fluctuates around 0.5, which is also better than the original LeNet-5 network.

Finally, the best performing method is chosen for t-SNE visualization, converting the multidimensional dataset into a low-dimensional dataset, and observing the classification performance of each category of CIFAR-10 under this model [10].

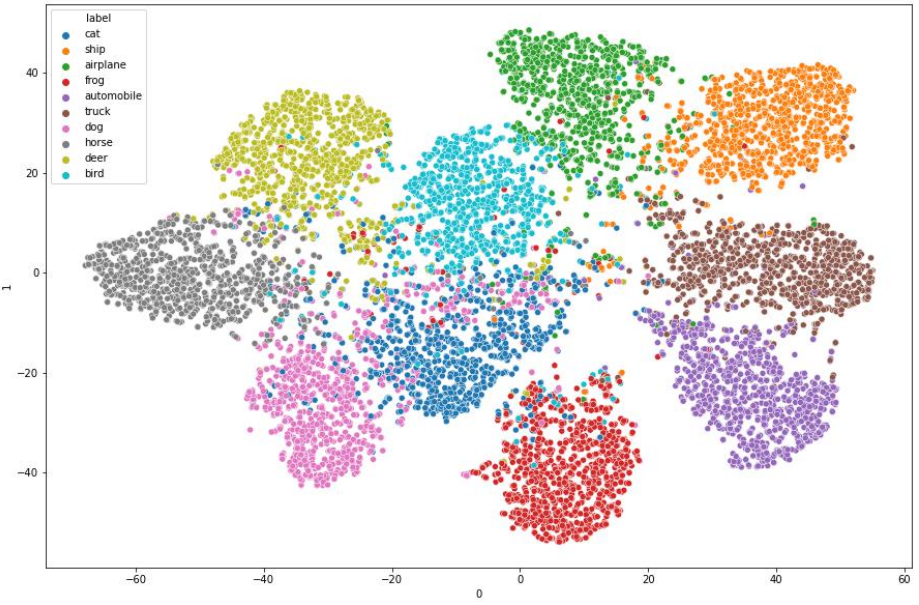

Fig. 9. t-SNE visualization results (Original)

It can be observed that the 10 categories of CIFAR-10 perform well in classification, and 10 different clusters can be clearly seen in visualization (Fig.9). Among them, there are some erroneous samples distributed at the edge boundary, manifested as different colors or categories appearing in the same clustering. It can also be found that the overlap area between pink and dark blue clusters is relatively large, indicating that the model cannot distinguish the categories of pink and dark blue very well. Observing the labels, the dark blue labels are for cats and the pink labels are for dogs, indicating that some features between cats and dogs may be repetitive, and image classification is prone to misjudgment, leading to confusion.

4 Conclusion

In summary, this paper first delves into the network structure of LeNet-5, understands its advantages and disadvantages, and makes appropriate improvements to the network structure, adding some mechanisms that were not included in the original network. Method 1 effectively improves the problem of relatively simple network structure by deepening the depth of the original network and increasing the hidden layer. However, the corresponding evaluation indicates that the improvement in standard accuracy and recall rate is not significant, and it is speculated that the reason is that the dataset is relatively simple, resulting in overfitting. Method 2 uses the dropout mechanism to effectively reduce overfitting issues by analyzing the accuracy change curve. Method 3 uses a CBAM module that combines spatial attention and channel attention, which performs best among all modules. Its accuracy reached 0.72 and recall rate reached 0.85. Finally, by combining all modules, it was found that the effect was almost identical to the original network. It is speculated that the dataset is too simple and the network structure is too complex, leading to network degradation. Improvement based on LeNet-5 is just a beginning. In the future, deeper networks and more diverse optimization methods will make neural networks more adaptable to new and massive datasets, unleashing their infinite potential.

References

1. Dong Jun, Zhou Feiyan, Jin Linpeng: Review of Convolutional neural network. The journal of Computer Science 06, 1229-1251 (2017).
2. Lu Hongtao, Zhang Qinchuan: Review of the Application of Deep Convolutional neural network in Computer Vision. Journal of Data Acquisition and Processing 01, 1-17 (2016).
3. Le Cun, Y., Boser, B., Denker, J.S.: Handwritten digit recognition with a back-propagation network. Neural Information Processing Systems 89, 396-404 (1989).
4. Srivastava, Nitish: Dropout: a simple way to prevent neural networks from overfitting. Journal of machine learning research 15.1, 1929-1958 (2014).
5. Yang Guanci, Yang Jing, Li Shaobo: Improved CNN Algorithm Based on Dropout and ADAM Optimizer. Journal of Huazhong University of Science and Technology (2018).
6. Zhu Zhengli, Wu Yuan: Research progress on attention mechanism in deep learning. Journal of Chinese Information Processing 33(6), 1-11 (2019).

7. Woo, S., Park, J., Lee, J.Y.: Convolutional block attention module. Proceedings of the European Conference on Computer Vision (ECCV), 3-19 (2018).
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Aidan, N.: Attention Is All You Need. Advances in Neural Information Processing Systems (2017).
9. Wu Zhengwen: Application of Convolutional neural network Based on CIFAR-10 Data Set in Image Classification. Computer Applications and Software, 1-9 (2016).
10. Maaten, L., Geoffrey, E.: Visualizing Data using t-SNE. Journal of Machine Learning Research 9(2605), 2579-2605 (2008).