ISSN 2616-5775 Vol. 7, Issue 8: 9-14, DOI: 10.25236/AJCIS.2024.070802

Research on image classification based on ResNet

Huimin Lu[1,*], Handong Song[2]

[1]*School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, 430074, China*
[2]*School of Mechanical Engineering, University of Jinan, Jinan, 250024, China*
[*]*Corresponding author: 2832326967@qq.com*

Abstract: *As one of the representative algorithms of convolutional neural networks, deep residual network has achieved very good results in computer vision, classification and other fields. Deep residual networks redefine the way networks learn, allowing networks to directly learn the difference between input and output information. This paper takes the deep residual network as the main research object, introduces the innovation of the deep residual network, and uses the deep residual network and other neural network models to train on different datasets. It is found that the highest accuracy of the test set of the deep residual network on the MNIST dataset is 0.555% higher than that of LeNet. 0.225% higher than AlexNet; The maximum accuracy of the test set on the CIFAR-10 dataset is 23.13% higher than LeNet and 2.46% higher than AlexNet.*

Keywords: *Deep residual network, Convolutional neural network, Image recognition, Neural network model*

1. Introduction

In today's era of rapid technological development, deep convolutional neural networks play a crucial role in the development of the computer vision field. With the increase in the amount of data and the improvement of computing power, deep learning technology has achieved great success in image recognition [1], object detection and other fields. Classical neural network models such as LeNet, AlexNet and VGG have achieved certain success. However, with the continuous development and complexity of neural network models, they all have common problems. That is, training difficulty and gradient disappearance of deep neural networks. This problem has led to a bottleneck in these traditional neural network models when dealing with deeper network structures, limiting their performance and application on large-scale data sets. Therefore, it is very important to find an effective method to solve the problem of network degradation during deep neural network training.

Network degradation refers to the phenomenon that the training error of the model increases when the number of network layers increases. This phenomenon is mainly due to the problem of information loss caused by gradient disappearance and gradient explosion. In traditional deep neural networks, as the number of network layers increases, the information gradually decreases in the process of transmission, which makes the optimization of the network difficult and makes the network performance reach saturation or decline. In order to solve this problem, various methods and structures have been proposed in the academic circle. It is under this background that the deep residual network comes into being.

In 2015, He Kaiming et al. proposed a network structure named deep residual network [2], which effectively solved the problem of network degradation by introducing the ideas of jump connection and residual learning. The core idea of residuals learning is to learn residuals maps by adding the content learned at each layer of the network to the residuals (i.e. residuals blocks) between the inputs, rather than learning the full mapping directly. This method makes it easier for the network to learn the identity mapping, avoids the problem of information loss, and also effectively reduces the impact of gradient disappearance and gradient explosion, so that deeper and more complex neural network structures can be trained. This key innovation enables deep residual networks to easily train thousands or even tens of thousands of layers of networks, greatly improving the model's performance and generalization ability, and is widely used in many fields today [3][4].

In this paper, deep residual network will be deeply discussed, and its application and advantages in image recognition, target detection and other tasks will be analyzed. At the same time, the performance of deep residual network, traditional deep neural network and other network models on different data

ISSN 2616-5775 Vol. 7, Issue 8: 9-14, DOI: 10.25236/AJCIS.2024.070802

sets will be compared, so as to comprehensively evaluate the performance and characteristics of deep residual network. Finally, the experimental results are summarized, and the direction of improvement and further research is proposed, in order to provide useful reference and inspiration for the development of deep convolutional neural networks in the field of computer vision.

2. Related work

Significant progress has been made in the field of deep learning over the past few decades, driven in large part by the continuous development and improvement of neural networks. As early as 1989, Yann LeCun and his colleagues published the results of convolutional neural networks. Convolutional neural networks (CNNS) are deep learning models specifically designed to process neural networks with grid-structured data. Its basic structure is composed of multiple layers, including convolutional layer, pooling layer and fully connected layer, as shown in Figure 1. In the convolutional layer, local features of the input image can be extracted through convolution operations, including edges and textures, etc., which have translation homogeneity. The pooling layer is usually followed by the convolution layer, which is used to reduce the size of the feature map and retain the main features, and has translation invariance. The fully connected layer is used to map the high-dimensional features to the class probability distribution and finally output the classification result. The structure of convolutional neural networks is inspired by the well-known Hubel-Wiesel biological vision model, specifically simulating the behavior of Simple and Complex cells in layers V1 and V2 of the visual cortex [5].

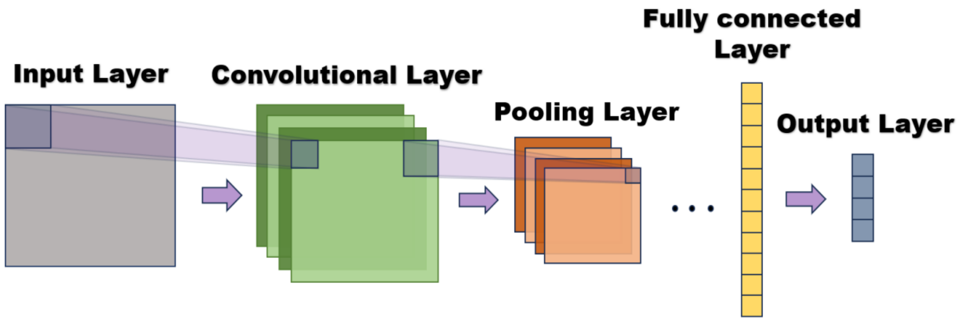


Figure 1: Basic structure of convolutional neural network

Among traditional convolutional neural networks, LeNet[6] can be regarded as one of the pioneers of deep learning. It was proposed by Yann LeCun et al in the early 1990s and is one of the earliest convolutional neural networks. It is mainly applied to handwritten digit recognition tasks [7]. Convolution and pooling operations are used to extract features and achieve spatial invariance. LeNet's success marked a new era for neural networks in computer vision, setting the stage for more complex network models.

Subsequently, AlexNet's proposal completely changed the landscape of deep learning. AlexNet[8], proposed by Alex Krizhevsky et al., in 2012, uses a deep convolutional neural network structure and introduces ReLU activation function [8] and Dropout technology to greatly improve the network's performance and generalization ability. AlexNet achieved great success in the ImageNet Image Recognition Challenge, leading a new wave of deep learning research, attracting more researchers to invest in the field of neural networks, and promoting the rapid development of deep learning technology.

As one of the important architectures, deep residual network has attracted the attention of many researchers. Through the comprehensive analysis of existing literature, we can see the rise and development of deep residual network [2] and its important role in solving problems such as gradient disappearance and gradient explosion in deep learning. It has fewer parameters, simpler network structure, and faster training speed. Due to the introduction of residual connections, gradients can propagate more smoothly, and the training process of deep residual networks is more stable and efficient. This allows deep residual networks to converge faster and achieve better training results on large data sets.

3. The basic fundamental of Deep residual network

Deep residual network is a kind of deep residual neural network proposed by Kaiming He[2] et al in 2015. The core innovation of deep residual networks is the introduction of residual connections, through

ISSN 2616-5775 Vol. 7, Issue 8: 9-14, DOI: 10.25236/AJCIS.2024.070802

which very deep neural networks can be easily trained, while effectively solving the problem of gradient disappearance during deep network training.

Residual connections allow information to cross directly between layers, rather than being cascaded layer by layer like a traditional network. This means that even if more layers are added to the network, the network can learn an identity map that passes the input directly to the output. If the added layer does not increase the performance of the network, then the network can choose to learn an identity map, which solves the problem of network degradation. Specifically, let's say our original input is x and we want to pass it to some layer's output $H(x)$. Traditional network will enter x F through a series of nonlinear transform (x) to get the output, namely $H\ (x) = F\ (x)$. In fact, existing neural networks have a hard time fitting the underlying identity mapping function, and in ResNet, instead of passing the input directly to the nonlinear transformation, we pass it through a residual unit $H(x) = F(x) + x$. This way, even though $F(x)$ learns a transformation close to zero, $H(x)$ can still be equal to x, which implements a residual connection and speeds up the network training process.

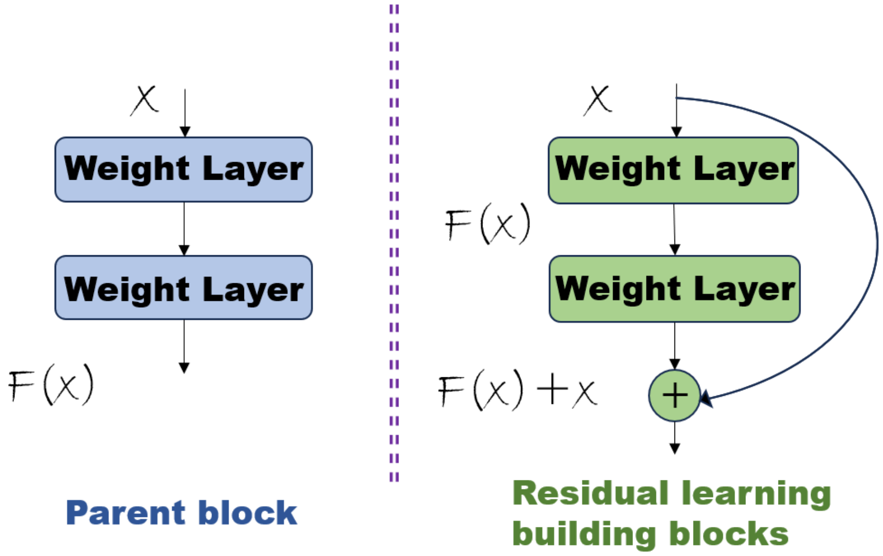


Figure 2: Residual structure compared with normal structure

As shown in Figure 2, the residual structure has more connection lines on the right named shortcut connection than the normal structure. Through the connection line, the output of the previous layer (or several layers) is added to the output of the current layer, and the sum result is input to the activation function to get the output of the current layer. If the identity map is an optimal model, it is only necessary to reset the weight of the weight layer in the above figure to 0. In addition, the residual connection adds neither parameters nor computational effort, and the model can be built quickly using current deep learning tools.

Its mathematical expression [2]:

$$y = F(x, \{W_i\}) + x \tag{1}$$

Where x and y respectively represent the input and output vectors of the residual block, and $F(x)$ is the residual mapping to be learned, shown in the figure above

$$F(x) = W_2\sigma(W_1x) \tag{2}$$

Where σ is the activation function, $W1$ and $W2$ represent two weight layers and the bias b is omitted for convenience.

4. Results

The experimental part aims to deeply analyze the advantages and innovations of deep residual network compared with traditional neural network architecture by comparing the performance of deep residual network with LeNet and AlexNet on data sets CIFAR-10 and MNIST respectively. We chose deep residual networks and AlexNet for comparison because they represent important milestones in the

ISSN 2616-5775 Vol. 7, Issue 8: 9-14, DOI: 10.25236/AJCIS.2024.070802

evolution of deep learning and differ significantly in structure and performance from deep residual networks.

We will first introduce the data set and experimental setup used, and then discuss the training results and performance of the deep residual network, LeNet, and AlexNet on this data set, respectively.

The MNIST dataset, which is widely used in the field of machine learning and deep learning, contains 60,000 training samples and 10,000 test samples, each of which is a 28 × 28 pixel grayscale image. Represents a handwritten digit between 0 and 9, and part of the handwritten digit image is shown in Figure 3:

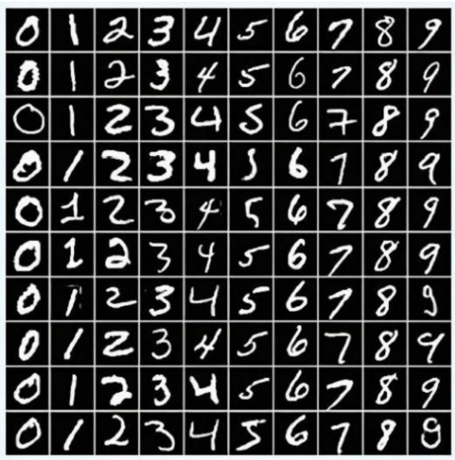
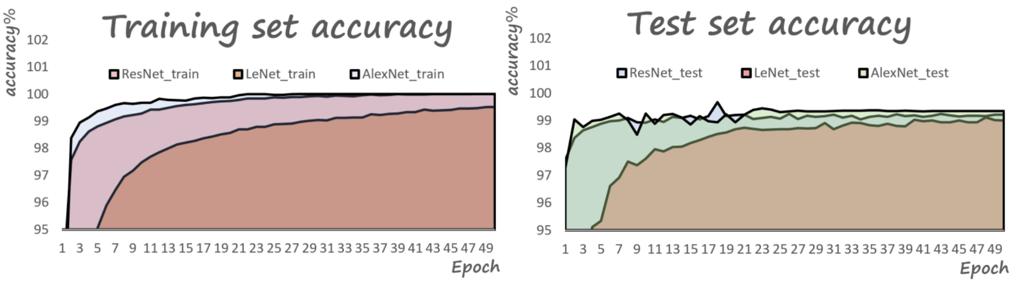


Figure 3: MNIST data set partial sample

LeNet, AlexNet and deep residual network training data were respectively used on the data set, and the accuracy results of the three kinds of networks on the training set and test set were finally obtained, as shown in Figure 4:



(Left side is training set accuracy, right side is test set accuracy)

Figure 4: Training set and test set accuracy on MNIST data sets

As can be seen from the result graph, the effect of deep residual network on the training set is in the middle, which is not as high as AlexNet's training accuracy. However, in the test set, the accuracy of deep residual network is similar to AlexNet's. This is because the deep residual network introduces the residual connection mechanism, which effectively alleviates the problem of gradient disappearance. The model can better learn and represent complex features. As a result, deep residual networks exhibit better generalization ability on test sets and are better able to adapt to previously unseen data.

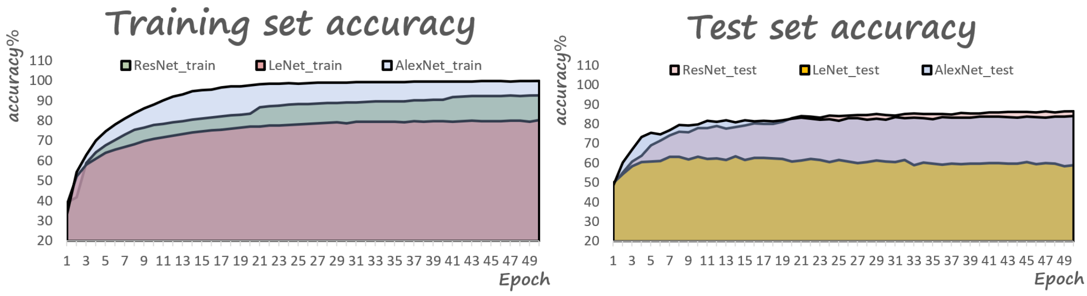
Deep residual networks and AlexNet perform better than LeNet on MNIST data sets because of the relative simplicity and standardization of MNIST data sets, which enables deeper and more complex models such as deep residual networks and AlexNet to better learn and represent features in the data.

The CIFAR-10 data set contains 10 categories of color images, each category has 6000 images, a total of 60,000 images, part of the image is shown in Figure 5, image size is 32 × 32 pixels, each pixel is composed of 3 channels (red, green and blue).

ISSN 2616-5775 Vol. 7, Issue 8: 9-14, DOI: 10.25236/AJCIS.2024.070802

Figure 5: Partial sample of CIFAR-10 data set

LeNet, AlexNet and deep residual network training data were respectively used on the data set, and the accuracy results of the three kinds of networks on the training set and test set were finally obtained, as shown in Figure 6:



(Left side is training set accuracy, right side is test set accuracy)

Figure 6: Training set and test set accuracy on CIFAR-10 data set

As can be seen from the result graph, the accuracy of the deep residual network on the training set is still in the middle, which is not as high as that of AlexNet on the training set. However, the accuracy of the deep residual network and AlexNet on the test set is about the same as that of AlexNet. The effect of the deep residual network and Alexnet on the training set and test set is better than that of LeNet.

After sorting out the experimental results on the MNIST and CIFAR-10 data sets, and taking the highest accuracy rate in the experimental process as the index for comparison, the results in Table 1 below can be obtained. The highest test accuracy of the deep residual network on the MNIST dataset is 0.555% higher than that of LeNet and 0.225% higher than that of AlexNet, respectively. The maximum accuracy of the test set on the CIFAR-10 dataset is 23.13% higher than LeNet and 2.46% higher than AlexNet.

Table 1: The highest accuracy on the training and test sets

	MNIST **(Training set)**	MNIST **(Test set)**	CIFAR-10 **(Training set)**	CIFAR-10 **(Test set)**
ResNet	100.000%	99.665%	92.890%	86.600%
LeNet	99.523%	99.110%	80.594%	63.470%
AlexNet	100.000%	99.440%	99.974%	84.140%

In conclusion, both the MNIST data set and the CIFAR-10 data set have achieved good image recognition effect.

ISSN 2616-5775 Vol. 7, Issue 8: 9-14, DOI: 10.25236/AJCIS.2024.070802

5. Conclusions

In general, the deep residual network achieves better results on the MNIST dataset and CIFAR-10 dataset than LeNet and AlexNet. This experimental result successfully reveals the superiority of deep residual network in image classification task, and shows its universality and generalization ability on different data sets and the effectiveness of deep residual connection mechanism, which provides a better solution for image classification task.

Although deep residual networks have achieved great success in image recognition, they still have some disadvantages. Deep residual networks require more computational resources. The model parameters and computation amount can be reduced by model compression, pruning and quantization techniques to achieve lightweight and acceleration of the model. Moreover, deep residual networks are usually trained for a single task, but in practical applications, it is often necessary to process multiple related tasks or transfer knowledge from previous tasks. Therefore, we can study how to combine multi-task learning and transfer learning techniques with deep residual networks to improve the efficiency and generalization ability of the model.

References

[1] Li X , Zhang J , Xue Y ,et al.Classification of hops image based on ResNet-ConvLSTM and research of intelligent liquor picking system[J].Measurement, 2022(194-):194. DOI:10.1016/j. measurement. 2022. 110955.

[2] Xu Z , Sun K , Mao J .Research on ResNet101 Network Chemical Reagent Label Image Classification Based on Transfer Learning[C]//2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT).IEEE, 2020.DOI:10.1109/ICCASIT50869.2020.9368658.

[3] Xu Jiaqin, Zhang Min, Wang Liuying, et al. Application of deep learning method based on residual neural network in ECG automatic recognition of arrhythmia [J]. Chinese Journal of Health Statistics, 2023, 40(3):345-348.

[4] Han Zhengjie, Niu Rongjun, Ma Zikui, et al. Bearing Fault Diagnosis Method based on improved residual neural network based on attention Mechanism [J]. Journal of Vibration and Shock, 2023, 42(16):82-91.

[5] Chen Chao, Qi Feng. Review on the development of Convolutional neural networks and their applications in computer vision [J]. Computer science, 2019, 46 (3): 11. DOI: 10.11896 / j.i SSN. 1002-137 - x. 2019.03.008.

[6] Zeng F , Hu W , He G ,et al.Imbalanced Thangka Image Classification research Based on the ResNet Network [J].Journal of Physics: Conference Series, 2021, 1748(4):042054 (8pp).DOI:10.1088/1742-6596/1748/4/042054.

[7] Deng Changyin, Zhang Jie. Handwritten digit Recognition based on improved LeNet-5 model [J]. Information and communications, 2018 (1): 4. DOI: 10.3969 / j.i SSN. 1673-1131.2018.01.047.

[8] Wang Z .Research on Garbage Image Classification Based on Neural Network[J].2022 International Conference on Computer Network, Electronic and Automation (ICCNEA), 2022:214-217.DOI: 10. 1109/ ICCNEA57056.2022.00055.