Analysis of the Topological Structure of the Convolution Neural Network Model RESNET

Tao Shang and Ying Song
Beijing Advanced Innovation Center for Materials Genome Engineering, Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing Information Science and Technology University
Email:18713535718@163.com,songying@bistu.edu.cn

Abstract. With the advent of the era of big data and the development of artificial neural network, a large number of different neural network models have come out. In the field of image recognition, the application proportion of convolutional neural network model is very high, and the different structure levels of the same convolutional neural network model are also very different, and the recognition effect in solving different problems is not the same. Taking the application of image recognition as an example, this paper analyzes and compares the influence of different topological structures in RESNET model on recognition accuracy. Through the experimental comparison and analysis, it is concluded that the structure of shortcut_6 in RESNET convolutional neural network is the turning point. When the length of shortcut is less than 6, the recognition efficiency is on the rising trend with the deepening of the network. When it is equal to 6, the recognition efficiency is basically fixed, and when it exceeds 6, the recognition accuracy is reduced.

1. Introduction
Since the deep convolution neural network model alexnet was proposed by Alex in 2012 until 2019, it has made a qualitative leap in neural network learning which has been in the low and bottleneck period, and has also improved the algorithms in related fields, effectively promoting the deep learning in natural language processing, image recognition, speech recognition and many other fields. In addition, the tensorflow [1] deep learning development platform constantly improved and updated by Google gives people the opportunity to quickly build a complex and deep convolutional neural network layer[2]. In recent years, with the rapid development of convolutional neural network model in various fields, especially in the field of image recognition, many convolutional neural network models have been produced. Analyzing the advantages and disadvantages of these convolutional neural network models has become an important research direction. In this paper, by analyzing the characteristics of the mainstream convolutional neural network model RESNET, the experimental comparison is mainly divided into four parts In the part of this paper, it not only gives an overview, related work, the characteristics of RESNET neural network model and the comparative analysis of different topological structures, and draws conclusions, but also uses the open-source framework of tensorflow in Google to build, train and compare these models.

2. Related Work
Due to the rapid promotion of convolutional neural network, many domestic scholars have made comparative analysis on the convolutional neural network model. In March 2018, Xing Zhixiang[7]
and others compared and analyzed the pedestrian head detection method models based on convolutional neural network, such as perception, RESNET and mobilenet, and found that different feature extraction networks and other factors will also affect the detection effect of the target detection model. Features With the increase of the number of layers of extraction network, the experimental accuracy of the target detection model will increase, but the slower the detection speed is, for example, the detection effect of fast r-cnn\cite{2} target detection structure under different feature extraction networks of RESNET 50 and RESNET 101 is different.

The convolutional neural network model RESNET (recurrent neural network) proposed by Kaiming he and other four Chinese in Microsoft Research Institute is successful in using RESNET Unit has trained 152 layers of convolutional neural network and won the championship in 2015 \cite{7} ilsvrc competition, but the error rate on top 5 is no more than 4%, which is widely concerned due to its excellent performance. The parameters used in this model are small, but the performance advantage is obvious. Through comparative analysis, it can be seen that the biggest feature of RESNET structure is that it can accelerate the training of neural network very quickly, and the accuracy of the model has been greatly improved compared with the previous, which is also a good advantage for generalization, and even can be directly used in the inceptionnet network. One of the advantages of RESNET network in the application process is that the original input information can be directly transmitted to the next layer, as shown in Figure 1.

![Figure 1. ResNet thought](image)

The origin of the neural network model RESNET is introduced in the front, and then the residualnetwork structure of RESNET is described, RESNET itself can extend the network layer to a deeper level through this residual network structure. At present, the number of network layers for this kind of research at home and abroad has reached as high as 1000 layers, and the final experimental results have also achieved good results. The concept of residual network refers to the high-speed network (highway) RESNET \cite{5} uses the form of identity mapping to replace the original weight of residual term. In the original type, assume that the input of a certain neural network is x, the expected output is h (x), and then H (x) can be obtained It is the expected value of complex potential mapping \cite{6}, but for this kind of model, the training is too difficult. But if it is found that the error in the lower layer becomes larger, that is to say, when the relatively saturated accuracy is learned, then the next training learning goal becomes the form of identity mapping. In order to keep the later layer without reducing the accuracy, to make the input x infinitely close to h (x) In the residual network structure, the initial result is directly represented by input X through shortcut (Shortcut connection), which is publicized as: H (x) = f (x) + X, and the expression of identity map is h (x) = x when f (x) is zero, Just like this, RESNET is equivalent to changing the learning objective, not learning a complete output, but the difference between the objective values H (x) and X, so RESNET hopes to approach the residual result to 0, with the deepening of the network, the higher accuracy is the training objective. This paper analyzes RESNET's shortcut, compares RESNET with other topologies, and draws experimental conclusions.

3. A Comparative Analysis of RESNET Neural Network Topology
This part mainly focuses on the comparative analysis of various topological structures in RESNET neural network model, first describes the experimental environment and describes the experimental
data, then the observation results of training prediction by constructing RESNET with different structures in multiple convolution layers, and corresponding shortcut RESNET, shortcut 6 resnet, shortcut 9 are established for RESNET with different structures. RESNet and so on were tested and trained respectively. Finally, the experimental conclusion was drawn by comparing and analyzing the data.

3.1. Comparative Test Design And Experimental Environment
This paper uses tensorflow, Google's latest open source neural network framework, to build, and through Python to train and test the model under pycharm. Training data set and test data set use Python crawler technology to crawl the required training pictures in Baidu picture network. The training data set and the test data set are divided by the ratio of eight to two.

3.2. Design and Analysis of RESNET Model Test
RESNET in the premise of depth known, this paper is to analyze some newly established network performance, so as to get the highest recognition accuracy rate by accurately skipping several convolution layers. Therefore, it is necessary to determine the depth of the network first, and then use RESNET with different structures under the premise of network depth known, and then deepen the network depth, so as to achieve RESNET network performance of various structures To verify. Influenced by the difference of network depth, the corresponding training time is also different. Generally, the larger the network depth is, the longer the training period is required. However, under some shortcut conditions, it is necessary to analyze the influence of multiple connection modes of shortcut on the network training time. According to the analysis, RESNET's depth and a variety of shortcut are two variables. Two different functions should be defined to represent the accuracy and training time respectively, as shown in Figure 2 and Figure 3 below:

$\text{True}=D(a,b;x)$  \hspace{1cm} $\text{Time}=T(a,b;x)$

**Figure 2.** Accuracy function  \hspace{1cm} **Figure 3.** Training time function

In the formula, $a$ represents network depth, $B$ represents shortcut length, true represents recognition accuracy, time represents network training time, and $X$ represents other parameters needed for neural network training, such as learning efficiency, batch, momentum and time. In the experimental stage, other parameters remain unchanged. In this formula, two variable parameters are $a$ and $B$. Generally, for convenience, the length $B$ of shortcut can be set when $Ba$ represents the network depth $a$.

In the experiment, when facing the network of shortcut 9 ResNet, shortcut 6 ResNet, shortcut 18 ResNet and other structures, the construction method is $6n+2$ layer network structure, in which $n$ is greater than or equal to 3.

Based on the above theoretical analysis, in order to analyze the RESNET network performance of various shortcuts in different depth networks, the value selected here $n$ is represented by the following: $3, 7, 11, 20$, so that the RESNET of the following depth can be obtained: layer 20, layer 44, layer 68, layer 122.

3.3. Experimental Process And Parameters
This section defines parameter $x$ in the formula in the previous section and describes the details through the following four processes. First, L2 regularization method can avoid over fitting, and the attenuation regularization weight coefficient is 0.002. Secondly, during the parameter updating process, the last updated value should be readjusted. If the corresponding gradient direction of the two parameters remains the same, then it is necessary to multiply the amount of change at the previous time with momentum $B$ to speed up the change speed of gradient direction. If the corresponding gradient direction is not unified, then the newly generated direction should be offset, and the gradient
direction should be slowed down. The selected value in this test is 0.9. The third is to train one epoch and 250 epochs for each group of laboratory training, which represents the data of all training sets at a time. The last step is to set the initial learning rate to 0.1, train epochs, the number of which is 200, and adjust the momentum to one percent or one thousandth during the experiment.

As a common training method, CNN training phase batch gradient descent method is also chosen in this paper. The higher the data batch used, the higher the corresponding memory utilization. In this experiment, the network depth is deepened slowly, and the network depth also directly affects the memory utilization. When training networks of layers 20, 44, 68 and 122, the corresponding batch number is 128. And define train_number as the number of training data. The relationship among batches, epochs and iterations is shown in Figure 4 below:

\[ \text{Iterations} = \frac{\text{train}\_\text{number}}{\text{batch}\_\text{number}} \times \text{epochs} \]

**Figure 4.** Batch, epochs and iterations

### 3.4. Experimental Result
In the above section, RESNET of various topological structures is listed, and various shortcuts are analyzed from different depths. The number of network layers used is 6N + 2. The corresponding experimental results with different values of N are shown in Figure 6.

**Figure 5.** Experimental results of RESNET with different topological structures in different depths

### 4. Concluding Remarks
From the above experimental data analysis and comparison of RESNET’s various different residual network topologies, it can be concluded that the network deepening effect will decrease after the increasing of shortcut. In RESNET, the deepening of network layers will increase the corresponding image recognition rate, while the length of shortcut structure extracted from network layers 68 and 122 is basically the same. Through the comparative analysis of this layer, it can be concluded that the value of 122 layer network is shortc The accuracy of the network is lower than that of the 68 layer.
This shows that the size of the shortcut directly affects the recognition results, and the structure of shortcut_6 in RESNET is the turning point. When the length of shortcut is less than 6, the recognition efficiency will rise with the deepening of the network. When it is equal to 6, the recognition efficiency is basically fixed, and when it is more than 6, the recognition accuracy will decrease.

5. Acknowledgments
This work is Supported by the National Natural Science Foundation of China (Grant No.61872043), Qin Xin Talents Cultivation Program, Beijing Information Science & Technology University(No. QXTCP B201904), the fund of the Beijing Key Laboratory of Internet Culture and Digital Dissemination Research (Grant No.ICDDXN004), and the Key Scientific Research Projects of Henan Higher School (Grant No.19A520043).

6. References
[1] Hua Huang, Shan Lin. WiDet: Wi-Fi based device-free passive person detection with deep convolutional neural networks[J]. Computer Communications, 2020, 150.
[2] Cornel W. du Toit. An immanent approach to death: Theological implications of a secular view[J]. Hervormde Teologiese Studies, 2009, 65(1).
[3] Maevin,M,Seymour,P. Perceptrons: An Introduction to Computational Geometry[M]. 1969:68-69.
[4] Ma Xiaolei, Dai Zhuang, He Zhengbing, Ma Jihui, Wang Yong, Wang Yunpeng. Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction.[J]. Sensors (Basel, Switzerland), 2017, 17(4).
[5] J. Justin Gooding, D. Brynn Hibbert, Wenrong Yang. Electrochemical Metal Ion Sensors. Exploiting Amino Acids and Peptides as Recognition Elements[J]. Sensors, 2001, 1(3).
[6] Ming Zhu. A Novel Deep Model for Image Recognition[C]. IEEE Beijing Section. Proceedings of 2014 IEEE 5th International Conference on Software Engineering and Service Science. IEEE Beijing Section: IEEE BEIJING SECTION. 2014:369-372.
[7] Mhaskar H.N.. Function approximation with zonal function networks with activation functions analogous to the rectified linear unit functions[J]. 2019, 51.