A Lightweight Residual-Inception Convolutional Neural Network

Kangyu Gao, Qingyong Zhang*, Haoran Wang

School of Automation, Wuhan University of Technology, Wuhan, Hubei, 430070, China

*Corresponding author’s e-mail: qyzhang@whut.edu.cn

Abstract. Deep convolutional neural networks have become a powerful tool to solve practical problems, especially in the field of image recognition and machine learning. This paper introduces a lightweight convolutional neural network model which named RINet, based on a combination of the Inception structure and blocks inspired by ResNet. This model achieves high accuracy while reducing the number of parameters and increases the training speed. It has reached the correct rate of 93.7% on the dataset of ISIC and the recognition accuracy improves 6.3% and 1.5% compared to deeper networks called InceptionV1 and transfer learning of InceptionV3.

1. Introduction

Whether in machine learning or image recognition field, an increasing number of practical problems have been solved because of the application of convolution neural networks. Convolution neural networks have helped people make breakthrough in tasks including object detecting [1][2][3], language recognition [4] and image recognition [5][6][7]. Meanwhile, the technology is also widely used in the detection of materials such as [8], medical diagnosis [9] and other various fields. And in some areas [10], the application of deep convolution neural network makes artificial intelligence even be better than humans, with the development of AI technology and robotic industry, neural network convolution will penetrate into every corner of future life.

In the year of 2014, Szegedy et al. proposed Inception [6] model (also known as GoogLeNet), and achieved the top accuracy in ImageNet Large-Scale Visual Recognition Challenge 2014(ILSVRC14). Then on, the team has subsequently published a series of paper [11][12][13], continuous improving Inception model. Sparse expression module, Batch Normalization and other methods were introduced, making the model gain better performance. Then in 2016, He et al. proposed ResNet model [7], which greatly reduces the difficulty of deep network training and has made new breakthroughs in deep web gradient disappears on the issue, the model broke the record GoogLeNet made in ILSVRC14 and won the championship in ILSVRC15.

Although Inception model brought many fine and inspiring Inception blocks, which greatly improves the efficiency of the model. And the application of ResNet module greatly reduces the computational difficulty of deep network which enhance the training speed. After many experiments, with increase of the Inception and ResNet modules, it did not obtain the corresponding benefits [14]. In the dataset of ISCI, the transfer learning of InceptionV3 model only reaches 92.2% accuracy rate. Based on these above, this paper proposes a new deep convolutional neural network, combining the advantages of the Inception module and the ResNet module to build networks and test them.
Ultimately find out what we call RINet network which achieved the correct rate of 93.7% on the dataset of ISIC.

2. Related works

2.1 Inception models

20 years ago, LeCun, Y et al proposed LeNet-5 and achieved markable results. After that day convolution neural network has aroused widespread concern in various fields, and continue to produce new models [2][3][6][7] year after year in the relevant competitions. In 2014, InceptionV1 was born, while going deeper which like the other models these days and paying more attention on the width and sparse structure of the network. In general, the most straightforward approach to improve network performance is to increase the depth of the network, which means a huge number of parameters and much time and difficulty to train. However, a large number of parameters are prone to overfitting and greatly increase the amount of calculation. [6] proposed that the fundamental solution to the above two drawbacks is converting fully connected or even generally linear convolution into a sparse connection. On the one hand, the connection of the real biological nervous system is also sparse. On the other hand, the paper [15] shows that for large-scale sparse neural networks, it can be layer by layer by analysing the statistical properties of the activation values and clustering the highly correlated outputs. This suggests that a bloated sparse network may be simplified without loss of performance. And a lot of paper indicates that a sparse matrix may be clustered into a relatively dense submatrix to improve computing performance. Accordingly, [6] proposed a structure called Inception to achieve this purpose. And finally, GoogLeNet’s 22 layers structure beat the other competitors and won the ILSVRC14. Subsequently, the team also proposed a variety of improvements on Inception models, developing that continuously refreshing the correct rate record reached by the previous generation model constantly.

2.2 ResNet models

Later in 2015, ResNet was born, and won the ILSVRC15 competition. While Inception models focused on the width, ResNet paying attention on the link between layers. The main idea of ResNet is to add a direct connection channel to the network, which is the idea of the Highway Network. Prior to performance of the input network structure is made a non-linear transformation, and outputs Highway Network certain percentage before the network layer allows retention. ResNet’s idea is very similar to Highway Network, allowing the original input information directly to the back of the layer, as shown in figure1.

![Figure1. Highway Structure.](image)

Traditional convolutional networks or fully connected networks must have information loss during information transmission, and cause gradients disappearing or gradient explosions, resulting in deep networks’ hard training. ResNet solve this problem by directly entering information transmitted output bypass to protect the integrity of information, only to learn the entire network input, output difference that part, to simplify learning objectives and difficulty. The biggest difference with ResNet is that
there are a lot of bypasses that connect input directly to the back layer. This structure is also called shortcut or skip connections. That also plays a role in preventing the disappearance of the gradient while reducing the dimension and speeding up the training [7], which greatly inspires the subsequent development of the convolutional neural network. Then in 2016, InceptionV4 proposed [13], the model combined with features of ResNet and Inception, which is a very classic model with a great reference value, achieved 3.08% top-5 error rate on the dataset of the ImageNet competition.

2.3 Transfer learning
In order to ensure high reliability and accuracy of the trained classification models, there are two basic assumptions in classic neural network tasks: (1) training samples for learning and new test samples must meet the independent and identically distributed; (2) there must be enough available training samples. Only after satisfying the above two conditions can we learn an excellent classification model. But in fact, these two assumptions are often difficult to achieve in practical applications. Firstly, with the change of actual situations, the previously available labelled sample data may become unavailable, and the new profile of the test sample of the present nicking semantic distribution. In addition, it is very difficult to build a data set with a sufficient number of labelled samples, which usually takes a long time to accumulate and costs a lot of human and financial resources. These led out an important question, how to make good use of the existing small amount of training samples with labels or the source field data to create a reliable model, and use it to predict the target areas having different data distribution?

Transfer learning can solve this problem very well. Transfer learning is shipped with the knowledge there have been different but related fields to solve the problem of a new machine learning method. It stretches those two basic assumptions in traditional deep learning, with the goal of migrating existing knowledge to solve learning problems in the target domain with only a small number of tagged sample data or even no data. Transfer learning widely present in human activities, such as who can ride a bicycle learn to ride a motorcycle is not difficult. The more similar in two different areas, the easier to realize transfer learning. In recent years, there have been a considerable number of researchers into the field of transfer learning. There are many transfer learning paper published [16][17][18] each year in the field of machine learning and data mining. And, in order to compare the merits of the method results, many people also adopted the way of comparing self-build models with transfer learning of classical models on the same dataset.

Although the application of Inception and ResNet modules successfully enhance the accuracy of prediction, through research, net increase in the number of Res modules did not cause their benefits equivalent enhance [14]. This seems to be somehow trapped in the same predicament as the “deeper time”. Indeed, a simple concoction does not get significantly improved. Based on the traditional convolutional neural network, a new model is proposed which having the features of ResNet and Inception blocks by us. The adding of ResNet modules’ direct way in the model ensure high accuracy and reduce the dimension at the same time. Finally, we construct a multi-group model called RINet. And get 93.7% correct rate on the dataset of ISIC.

3. Architecture design
As previously mentioned, despite inception module has a compact structure and application of the model has achieved very good results in many areas. But for more residual module’s introduction will not bring more revenue, we use Inception and Residual module that has been modified accordingly to test, and build a series different model. Finally, we concluded a best-performance model which called RINet.

3.1 Inception structure
The Inception module used in this paper obeys the idea that is to expand the width while modifying the parameters. Model has been adjusted according to the dataset we use. A total of 3 Inception
structures are used. Meanwhile, in order to speed up training, the scale of the model has been reshaped. Its specific composition is as follows and can be seen in figure 2.

- Previous layer: DepthConcat or 3x3+2(S) MaxPool.
- Branch 1: 1x1+1(S) Convolution layer or a direct connection layer.
- Branch 2: 1x1+1(S) Convolution layer and next is a 3x3+1(S) Convolution layer.
- Branch 3: 1x1+1(S) Convolution layer and next is a 5x5+1(S) Convolution layer.
- Branch 4: 3x3+1(S) MaxPool layer and next is a 3x3+1(S) Convolution layer.
- All branches are contacted by the DepthConcat.

### 3.2 Residual structure

The main idea of ResNet is to add a direct connection channel to the network, which is the idea of the Highway Network. The previous network structure was a non-linear transformation of the performance inputs, while the Highway Network allowed a certain percentage of the output of the previous network layer to be preserved. The idea of ResNet is very similar to that of Highway Network, allowing the original input information to be passed directly to the later layers, as shown in the figure 1. So that this layer of neural network can learn without the entire output, but learn the residual of the previous network output. As used herein, residual blocks are added into the Inception structures with a total of two straight connections.

- First one is applied in the first inception layer’s branch 1.
- Second one is applied in the third inception layer’s branch 1.

### 3.3 The whole model

The specific structure is shown in figure 2.

### 4. Training strategy

This model uses the BN (Batch Normalization) structure proposed in the inceptionV2 model. This configuration can accelerate training, improve the generalization ability. In the BN layer, the sample mean is calculated first, then the sample variance, next the sample data is normalized, and finally the translation and scaling are performed. BN layer is essentially a normalized network layer, which can replace partial response normalization layer (LRN layer), meanwhile it can disrupt the order of the input samples, enhancing prediction accuracy. And in the training process, we choose the Adagrad optimizer, learning rate is set as 0.001, batch size is decided to 32. After 30000 steps, we achieve 93.7% correct rate and RINet converges in about 17000 steps. Approximately, it cost 13 seconds per hundred steps while InceptionV1 spent more than twice time.

### 5. Training strategy

This series of models were set to train in the ISCI dataset with different parameters, and ultimately come to the best model results, RINet, reached the correct rate 93.7% compared to the other classic models, the correct rates improved 6.3% and 1.5% with less training time and more stability. The specific comparison figures and table of the relevant experiments are as follows Table 1, figure 3 figure 4, figure 5 and figure 6.
Figure 2. RINet structure.
Table 1. Correct rate comparison.

| Model name                  | Correct rate on dataset of ISIC |
|-----------------------------|----------------------------------|
| RINet                       | 93.7%                            |
| InceptionV3 Transfer Learning | 92.2%                            |
| InceptionV1                 | 87.4%                            |

Figure 3. RINet Training Correct Rate.  
Figure 4. InceptionV1 Training Correct Rate.  
Figure 5. InceptionV3 Training Correct Rate.  
Figure 6. Correct Rate Comparison.

6. Conclusion
We introduces a lightweight convolutional neural network model which named RINet, based on a combination of the Inception structure and blocks inspired by ResNet. This model achieves high accuracy while reducing the number of parameters and increases the training speed. It has reached the correct rate of 93.7% on the dataset of ISIC and the recognition accuracy improves 6.3% and 1.5% compared to deeper networks InceptionV1 and transfer learning of InceptionV3.

Acknowledgments
This paper is financially supported by Wuhan University of Technology Independent Innovation Project Fund (2018-ZDH-A1-02).

References
[1] Schmidhuber, J. Deep learning in neural networks: An overview. Journal of Neural networks, 61:85–117, 2015
[2] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In In Advances in neural information processing systems, pp.1097–1105, 2012.
[3] Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. In arXiv preprint arXiv:1409.1556, 2014.

[4] Sainath, T.N, Vinyals, O., Senior, O., Sak H: CONVOLUTIONAL, LONG SHORT-TERM MEMORY, FULLY CONNECTED DEEP NEURAL NETWORKS

[5] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Computation 1(4), 541–551 (1989)

[6] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., and Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9, 2015.

[7] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016b.

[8] Girshick, R., Donahue, J., Darrell, T., and Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587, 2014.

[9] Bar Y, Diamant I, Wolf L, et al. Chest pathology detection using deep learning with non-medical training[C]//2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). IEEE, 2015: 294-297

[10] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of go without human knowledge[J]. Nature, 2017, 550(7676): 354

[11] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015

[12] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826

[13] Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In arXiv preprint arXiv:1602.07261, 2016a.

[14] Miikkulainen R, Liang J, Meyerson E, et al. Evolving deep neural networks[M]//Artificial Intelligence in the Age of Neural Networks and Brain Computing. Academic Press, 2019: 293-312

[15] Fung G P C, Jeffrey X Y, Lu H, et al. Text classification without negative examples revisit[J]. IEEE Transactions on Knowledge & Data Engineering, 2006 (1): 6-20

[16] Sanjeev Arora, Aditya Bhaskara, Rong Ge, and Tengyu Ma. Provably bounds for learning some deep representations. CoRR, abs/1310.6343, 2013.

[17] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on knowledge and data engineering, 2010, 22(10): 1345-1359

[18] H. Al-Mubaid and S. A. Umair, “A new text categorization technique using distributional clustering and learning logic,” IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 9, pp. 1156–1165, 2006

[19] K. Sarinnapakorn and M. Kubat, “Combining subclassifiers in text categorization: A dst-based solution and a case study,” IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 12, pp. 1638–1651, 2007