Convolutional Neural Networks Towards Diagnosis of Dermatosis

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. In the field of machine learning and image recognition, deep convolutional neural networks have become a powerful tool for solving practical problems. Based on convolutional neural network’s fast speed and high accuracy, we applied it to the diagnosis of pigmented skin in the medical field. By training the models on the dataset of ISIC, we eventually achieved up to 93.1% accuracy rate. If applied as an actual auxiliary tool, the diagnostic accuracy of related illness will be greatly improved.

1. Introduction

Due to factors including sun exposure, water loss, sebum secretion or genetic inheritance, human skin may produce various acne, pigmentation and other diseases, which are collectively referred to as skin diseases. One of the common diseases: pigmented skin disease. The disease is caused by abnormal production of melanocytes and melanin. Pigmented skin disease, such as freckle, perioral streak, usually with irregular pigment blocks, which brings people's daily life great inconvenience, though some of them will be more serious, even be life-threatening. Clinically, the diagnosis of skin diseases is mainly divided into two methods: visual judgment based on experience and pathological biopsy. In the actual diagnosis, due to lack of clinical experience, fatigue caused by long-term work and difficulty in distinguishing between various skin diseases, the visual diagnosis method has disadvantages including misdiagnosis and missed diagnosis; the biopsy diagnosis technique requires a part of the patient’s skin to be cut. Pathological biopsy, although high accuracy, is easy to leave a scar that brings inconvenience to one’s daily life and more emphasis on the risk of infection, which makes it even worse.

In the field of machine learning and image recognition, more and more people are beginning to use convolutional neural networks to solve engineering problems and have achieved great success. Convolutional neural networks are used to identify objects [1][2][3], speech recognition [4], and image recognition [5][6][7], which have achieved state-of-art accuracy. Meanwhile, the technology is also widely used in the detection of materials [8], medical diagnosis [9] and other various fields. Especially in medicine, due to a convolution neural network’s high accuracy and stability, applying that technology can greatly reduce the risk of fatigue because of lack of experience or lead to misdiagnosis of the mentioned. Based on that, we propose a pigmented skin diagnostic system based on convolution neural network, using LeNet-5[5], InceptionV3[12] model and train them on the dataset of ISIC, eventually set up to verify 93.1% accuracy rate, already can be a reference for diagnosis-related disorders.

2. Related works
2.1 Dataset of ISIC
ISIC (International Skin Imaging Collaboration) is an international skin imaging organization that promotes the use of computer technology in the diagnosis and treatment of dermatological diseases to help reduce the mortality of pigmented skin diseases. In addition, ISIC has developed and been expanding the open source public access archive of skin images, which has become a public resource for the development and testing of instructional images and automated diagnostic systems. In this paper, we have used three of various skin diseases, including Melanoma, Nevus and Seborrheic keratosis, each class corresponding to 2,000 tagged images, total 3 categories, 6000 sheets. Some samples of the dataset are shown as figure 1.

![Figure 1. Images from dataset of ISIC.](image)

2.2 Convolutional Neural Networks
Since 20 years ago LeCun, Y et al propose LeNet-5 and made great achievements, the convolution neural network has aroused widespread concern in all fields, and continued to produce more and more new models[2][3][6][7] since that day, which refreshing records year after year in related competitions. In the year of 2014, Szegedy et al. proposed famous Inception[6] model (also known as GoogLeNet). In the Large-Scale Visual Recognition Challenge 2014 (ILSVRC14), GoogLeNet has achieved the most advanced correct rate. The team has subsequently published a series of paper[11][12][13] later, improved the Inception models continuously. They introduced sparse expression module, Batch Normalization and other methods, making the model with better and better performance. The compact model also used the term “deeper”, at the same time, more focused on the width of the extension and development of the network “sparse structure”. Generally speaking, the most direct way to enhance network performance is to increase the depth and width of the network, but that means a huge number of parameters. However, the huge number of parameters prone to be overfit easily and also excessive large increase in the amount of calculation. [6] thought that the fundamental solution to the above two drawbacks is fully connected to a generally linear convolution which was converted into a sparse connection. Meanwhile realistic biological neural system connected aspect is also sparse. On the other hand, some research[15] showed that: for large scale sparse neural network, you can build an optimal network 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. Finally, 22 layers GoogLeNet beat the other competitors and won the champion of ILSVRC14. Subsequently, the team proposed a variety of improvements on Inception structures to develop the model continuously.

2.3 Transfer learning
In the conventional classification learning in order to ensure that the trained classification model having high accuracy and reliability, has two basic assumptions: (1) Training samples for learning and new test samples should satisfy independent and identical distribution; (2) must have available sufficient training samples in order to learn to get a good classification model. However, in practical, two assumptions are often difficult to achieve. First, over time, the previously available tagged sample data may become unavailable, with semantics and distribution gaps in the distribution of new test samples. In addition,
some tagged sample data is often scarce and difficult to obtain. This raises another important issue in machine learning: how to use a small amount of training samples with the label or the source field data to create a reliable model to predict the target areas having different data distribution. The transfer learning is a good solution to this problem. Transfer learning is shipped with the knowledge that there have been different but related fields to solve the problem of a new machine learning method. It relaxes two basic assumptions in traditional machine 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. Transfer learning is widespread in human activities, and the more factors shared by two different domains, the easier it is to transfer learning. In recent years, there have been a considerable number of researchers into the field of transfer learning, and there are many paper [16][17] are published every year in the top annual meeting in machine learning and data mining. Also, the method of comparing the new model with the classic one which is transferring on same dataset to evaluate performance is also adopted by more and more people.

If it is applied as an auxiliary diagnosis, it will greatly improve the correct rate of diagnosis of related diseases. Based on convolutional neural network recognition speed and high accuracy, we use a convolutional neural network and transfer learning, trained LeNet-5 inceptionV3 on the dataset of ISIC, respectively reached 91.2% and 93.1% correct rate.

3. Process

3.1 Inception structure
Firstly, we set all of the image data size as 299x299 pixel, and uniformly named them by tag name, meanwhile in order to make the training dataset more diverse, we randomly selected 30% of the image to rotate and fold in the training dataset, which corresponds to the practical application of the non-professional photography, which can help improve the robustness of the model. Part of processed images can be seen in figure 2. After that, the whole dataset is processed into tfrecord format to facilitate subsequent training process for image retrieval and the tag read.

3.2 Model structure
We mainly use two models for training. The first is LeNet-5, for the present dataset, we adjust the structure appropriately to obtain a better performance, the adjusted model have a total of 6 layer, the number of the input node is 299x299, and that of output node is 3. Its specific structure is as follows, the structure diagram is shown in figure 3.

- The first layer is the convolution layer, the input is the original image pixel, the size of the first convolution filter is 5x5, the depth (convolution kernel type) is 6, the full 0 padding is used, and the step size is 1.
- Second cell layer is a pooling layer, the filter is a 2x2 size, length and width of the step is 2.
- The third layer is a convolution layer, filter size is 5x5, a depth of 16. This layer does not use full 0 padding, and the step size is 1.
The fourth layer is the pooling layer, which uses a filter size of 2x2 and a step size of 2.

The fifth layer is a convolutional layer with a filter size of 3x3 and a step size of 1.

The sixth layer is a fully connected layer, which ultimately outputs the predicted results.

The second model is the InceptionV3 model’s transfer learning. Firstly, the inceptionV3 model will be pre-training on ImageNet. After reaching a higher correct rate, which means it has a strong feature extraction capability, and after bottleneck layer of the pre-trained model, two fully connected layers are followed, and the next layer output the prediction. Its main structure is shown in figure 4.

4. Training strategy
This model uses the BN (Batch Normalization) structure proposed in the inceptionV2 model. 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. The two parameters of $\gamma$ and $\beta$ are introduced for training, and the parameters $\gamma$ and $\beta$ can be learned, so that the network can learn to recover the feature distribution to be learned by the original network. The specific implementation process is as follows.

Input: Values of $x$ over a mini-batch: $\beta = \{x_1...m\}$;
Parameters to be learned: $\gamma$, $\beta$
Output: $\{y_i=BN_{\gamma, \beta}(x_i)\}$

\[
\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^{m} x_i
\]

(1)

\[
\sigma_B^2 \leftarrow \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2
\]

(2)

\[
x_i^* \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}
\]

(3)
\[
y_i \leftarrow y \gamma \beta + \beta \equiv BN_{\gamma, \beta}(x_i)
\]  

(4)

This configuration can speed up training, improve the generalization ability, while 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. Therefore, a higher learning rate is adopted when we training the network.

5. Experiment results
The two models are trained at different learning rates on the dataset of ISIC, and then compared under the same conditions. The modified LeNet-5 achieved a correct rate of 91.7% and the transfer learning of the inceptionV3 model yielded a correct rate of 93.1%. At the same time, compared with the modified LeNet-5, InceptionV3 also has better stability. The specific data is as shown in the following table 1, figure 5 and figure 6.

| Model name         | Correct rate on dataset of ISIC |
|--------------------|---------------------------------|
| LeNet-5_test_3     | 91.2%                           |
| InceptionV3-Transfer| 91.2%                           |

Table 1. Correct rate comparison.

![Figure 5. LeNet-5(adjusted) training process](image1)

![Figure 6. inceptionV3-Transfer training process](image2)

6. Conclusion
Based on convolutional neural network’s fast speed and high accuracy, we applied it to the diagnosis of pigmented skin in the medical field. By training the models on the dataset of ISIC, we eventually achieved up to 93.1% accuracy rate with InceptionV3-Transfer model. If applied as an actual auxiliary tool, the diagnostic accuracy of related illness will be greatly improved.

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 A, et al. Convolutional, long short-term memory, fully connected deep neural networks[C]//2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015: 4580-4584

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

[11] 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

[12] 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.

[13] Fung G P C, Yu J X, Lu H, et al. Text classification without labeled negative documents[C]//21st International Conference on Data Engineering (ICDE'05). IEEE, 2005: 594-605

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

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

[16] 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

[17] 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