Cancer Auxiliary Detection System Based on ResNet

Bingsen Wang[1]

[1]School of Software and Microelectronics, Peking University, No.5 Yiheyuan Road, Beijing, 100871, Beijing, China.

Contributing authors: nevermore@ss.pku.edu.cn;

Abstract

In recent years, as the accuracy of deep learning algorithms in image classification tasks exceeds that of the human brain, AI auxiliary diagnosis systems have attracted more and more attention. We used a deep convolutional neural network model ResNet with excellent performance in all aspects to complete the task of auxiliary diagnosis of cancer, and equipped the auxiliary diagnosis system with a user interface and a voice broadcast system. Finally, we achieved a classification accuracy of 97.54% on a dataset consisting of 270 thousand cancer pathological photos.

Keywords: Cancer Detection, Deep Learning, ResNet, Image Classification

1 Introduction

With the rapid development of artificial intelligence, especially deep learning, computers have been used more and more in the field of medical diagnosis, such as cancer diagnosis,and have begun to gradually replace the role of medical experts.As one of the representative algorithms of deep learning, the rapid development of CNN(convolutional neural network)[1] and the substantial increase in GPU computing power have directly led to the epic progress in the field of image processing[2, 3] after 2012.

Experiments[4, 5] show that the network depth is crucial for improving the performance of the model, and when the network depth is less than 30, deeper network models tend to have better performance. Deep network models[6, 7] have achieved outstanding performance both on the ImageNet dataset[8]

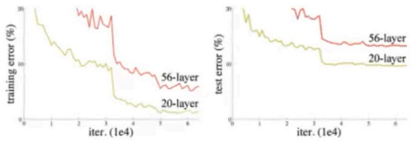

Fig. 1: Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error.

and on some other important image classification tasks[9–13]. However, as the network depth increases more than 30, a degradation problem has been exposed: with the network depth increasing, accuracy gets saturated (which might be unsurprising) and then degrades rapidly[14]. Network degradation is different from overfitting[15, 16], which is that a model does not perform as well on the test set as it does on the training set. While the network degradation is that as the network depth increases, the performance of the model on both the training set and the test set will become worse. Fig.1 shows a typical example.

ResNet(Residual Network) was first proposed by Kaiming He's team in 2015 to solve the above-mentioned network degradation problem. In addition to solving the problem of network degradation so that we can obtain better performance by using a deeper network, ResNet also has the advantages of fewer parameters and faster training speed than the classic CNN model. This is also the difficulty of medical artificial intelligence-assisted diagnosis. Networks that are not deep enough have insufficient accuracy, and networks that are too deep are not only difficult to train, but also performing worse. Based on this, we used ResNet to implement our cancer auxiliary diagnosis task, and after spending almost 4 hours training the model on GPU, the classification accuracy on the test set reached 97.54%. This is a practical result, as long as the model is equipped with a user interface and more diverse functions such as voice broadcasting, it has the potential to be put into practical use, which is what we aim to achieve in this paper.

2 Related Work

The history of CNN. The earliest CNN model is LeNet[17], proposed by Yann LeCun in 1990, which is designed to process the mnist handwritten digit database. Although the Lenet model is simple, it already has basic structures such as convolutional layers, pooling layers and fully connected layers. In 2012, Krizhevsky et al. proposed Alex-Net[18], and with the help of the high-performance parallel computing power of GPU, it won the first place in the ILSVRC competition, and the accuracy of Alex-Net was 10% higher

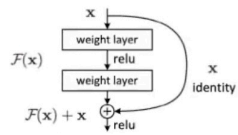

Fig. 2: a basic block

than the second place. The success of Alex-Net is remarkable, and opened the era of CNN. In 2013, Zeiler and Fergus proposed ZFnet[19], which uses a smaller convolution kernel $(7 * 7)$ and a smaller stride on the basis of Alexnet to retain more features and reduce the error rate of the ILSVRC competition to 11.7%(Alex-Net is 15.3%). The champion of the ILSVRC competition in 2014 was the GoogLenet[20] proposed by the google team, which used a multi-scale convolutional inception module to make the network deeper and wider, and reduced the error rate of the ILSVRC competition to 6.7% (the human error rate was 5.1%). VGG(Oxford Visual Geometry Group), which won the first place in the localization task and the second place in the recognition task in the ILSVRC 2014 competition, considers the improvement of the network structure[21]. It uses a stack of small convolution kernels $(3 * 3)$ instead of large convolution kernels to improve accuracy, and reduces the amount of parameters while increasing the depth of the network. After ResNet, different teams have improved ResNet and proposed models such as Wide ResNet[22],ResNeXt[23],Res2Net[24],iResNet[25],ResNeSt[26] and RegNet[27]. In addition to the ResNet series of models, many new network models have emerged in recent years, of which the most influential models are NASNet[28] and ViT[29].

The architecture of ResNet. ResNet consists of a stack of basic blocks as shown in Fig.2. A basic block is to add a shortcut connection to convolutional layers, so that the network can contain both nonlinear and linear components. As shown in Eq.1, ResNet is not to fit the initial distribution $\mathcal{H}(\mathbf{x})$, but to fit the residual function $\mathcal{F}(\mathbf{x})$.

$$\mathcal{F}(\mathbf{x}) = \mathcal{H}(\mathbf{x}) - \mathbf{x} \tag{1}$$

After solving the network degradation problem, we can increase the network depth by stacking more basic blocks to obtain better performance. As shown in Fig.3, it is the architecture of 34-layer ResNet and other networks. In this paper, we use a 101-layer ResNet model to implement the function of auxiliary diagnosis of cancer, in order to obtain a higher diagnostic accuracy.

User interface. The tkinter package is the standard Python interface to the Tk GUI toolkit. In this article, we use the tkinter package to build our user interface.

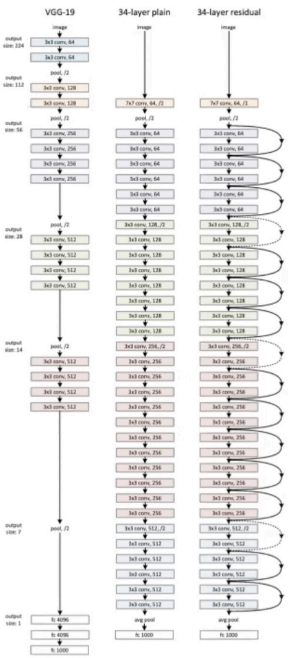

Fig. 3: Example network architectures for ImageNet. **Left:** the VGG-19 model [41] (19.6 billion FLOPs) as a reference. **Middle:** a plain network with 34 parameter layers (3.6 billion FLOPs). **Right:** a residual network with 34 parameter layers (3.6 billion FLOPs). The dotted shortcuts increase dimensions.

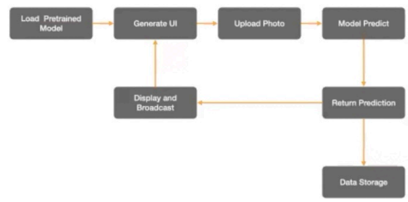

Fig. 4: The Application Architecture

Voice broadcasting.pyttsx3 is a text-to-speech package in Python, and we can realize the task of text-to-speech broadcast through the pptysx3 package, so that we can easily realize communicating of the diagnosis results.

3 Implementation of the application

Architectur of the application. In this application, we need to implement three functions, image classification, user interface and voice broadcasting. Therefore, we designed the application architecture as shown in the Fig.4.

Model Training. When training the model, we used the cancer detection dataset[30] from Kaggle. The dataset consists of a training set of 220k photos and a test set of 57.5k photos. After about 4 hours of training on the GPU, the model achieved a classification accuracy of 97.54% on the test set. The training process is shown in Fig.5.

Implementation of the application. After implementing the three functions separately, we finally completed the application by combining different modules. In the user interface, we use two global variables 'detect' and 'filename' to transfer and update parameters between different modules, so that we can combine the three functions. We have opensourced all the code and trained model on GitHub[31]. The completed model is shown in the Fig.6.

4 Results

In about three weeks, we successfully implemented a cancer AI assisted diagnosis system. The system has a good diagnostic accuracy and can be easily operated by the user. In addition, the system can also make voice broadcasts of the diagnosis results to give users some medical advice. Overall, we have achieved our stated goals about the application.

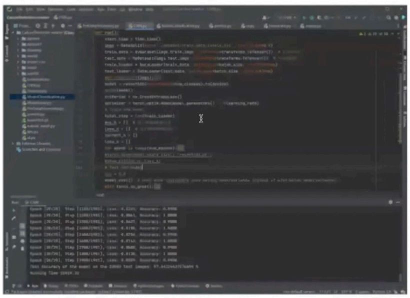

Fig. 5: Training Process

5 Conclusion

Although we have successfully completed the cancer AI assisted diagnosis system and implemented all the planned functions, we still need further improvement in the following aspects.

UI design. Due to the lack of time, we only implemented a very basic user interface, and the application does not have many extended functions. A beautiful and concise user interface can make it easier for users to use the application, and this is what we need to do.

Smarter Voice Interaction. Due to the rush of time, we only implemented a very basic voice broadcast function, and this application does not have the function of an intelligent voice robot. The more intelligent voice interaction function can further facilitate users to use the application, which is also an essential function of a complete set of cancer diagnosis assistant expert system. This is also a function that we need to gradually improve in future research.

More Models. In this application, we only use the 101-layer ResNet model to realize the function of auxiliary diagnosis of cancer, but there are many newer and smaller models that perform no worse than ResNet in image classification tasks. We can try to use more excellent models(such as Vit and NasNet) in the follow-up work to improve the classification accuracy. In addition to the CNN series models, we can also try to use SNN(Spiking Neural Network) to achieve our image classification tasks. As the representative model of the third generation of artificial intelligence, SNN also has great development potential in the traditional artificial intelligence field.

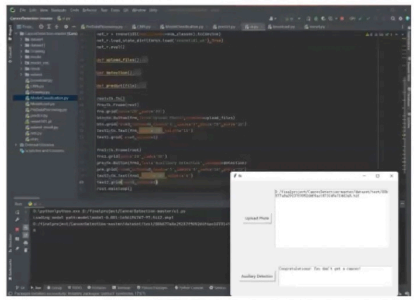

Fig. 6: Implementation of the Application

References

[1] Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25**(2) (2012)

[2] Matthew Zeiler, D., Rob, F.: Visualizing and understanding convolutional neural networks. (2014). ECCV

[3] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229 (2013)

[4] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

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

[6] He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)

[7] Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network

training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448–456 (2015). PMLR

[8] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., *et al.*: Imagenet large scale visual recognition challenge. International journal of computer vision **115**(3), 211–252 (2015)

[9] Girshick, R., Donahue, J., Darrell, T., 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)

[10] Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)

[11] He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence **37**(9), 1904–1916 (2015)

[12] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)

[13] Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015)

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

[15] He, K., Sun, J.: Convolutional neural networks at constrained time cost. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5353–5360 (2015)

[16] Srivastava, R.K., Greff, K., Schmidhuber, J.: Highway networks. arXiv preprint arXiv:1505.00387 (2015)

[17] LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L.: Handwritten digit recognition with a back-propagation network. Advances in neural information processing systems **2** (1989)

[18] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012)

[19] Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: European Conference on Computer Vision, pp. 818–833 (2014). Springer

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

[21] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

[22] Zagoruyko, S., Komodakis, N.: Wide residual networks. arXiv preprint arXiv:1605.07146 (2016)

[23] Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500 (2017)

[24] Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., Torr, P.: Res2net: A new multi-scale backbone architecture. IEEE transactions on pattern analysis and machine intelligence **43**(2), 652–662 (2019)

[25] Duta, I.C., Liu, L., Zhu, F., Shao, L.: Improved residual networks for image and video recognition. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 9415–9422 (2021). IEEE

[26] Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., *et al.*: Resnest: Split-attention networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2736–2746 (2022)

[27] Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., Dollár, P.: Designing network design spaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10428–10436 (2020)

[28] Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V.: Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8697–8710 (2018)

[29] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

[30] (PCam), P.: Histopathologic Cancer Detection Data Sets. https://www.kaggle.com/c/histopathologic-cancer-detection/data (2019)

[31] Wang, B.: Cancer Detection. https://github.com/danzhewuju/CancerDetection (2022)