Diabetic retinopathy identification system based on transfer learning

Lingling Li¹, Xueqi Yan¹, Heng Peng¹, Ying Xiu¹, Yiyang Gao¹ and Xin Wang¹, ²*

¹College of Software, Jilin University, Changchun130012, China
²College of Computer Science and Technology, Jilin University, Changchun130012, China
*Corresponding author’s e-mail: w_x@jlu.edu.cn

Abstract. Diabetic retinopathy, a complication of diabetes, is a significant cause of vision loss and blindness. Early detection of diabetic retinopathy can help reduce the risk of blinding. However, automatic diabetic retinopathy identification is a challenging task due to their different morphology during a different stage. Aiming at the problem of low efficiency for most of the existing methods, we developed a diabetic retinopathy recognition system based on transfer learning. The system utilizes transfer learning, which trains a neural network based on the DenseNet201 network model and Messidor Data Set, and can not only train an active network quickly, but also has a reasonable effect on the classification of diabetic retinopathy.

1. Introduction
Diabetic retinopathy (DR) is one of the most critical manifestations of microangiopathy lesions in diabetes. DR stages were classified as non-DR, nonproliferative DR (NPDR), or proliferative DR (PDR)[1]. DR is a systemic disease, which affects up to 80 percent of all patients who have had diabetes for 20 years or more. Despite these intimidating statistics, research indicates that at least 90% of these new cases could be reduced if there were proper and vigilant treatment and monitoring of the eyes[2].

The quality of retina images is essential for the diagnosis and treatment of eye diseases. Blur, distortion, low contrast, among other artifacts, inhibit the viewing of regions of interest[3], and the subtle features in the image that are difficult to distinguish by the naked eye are obstacles to the actual diagnosis of diabetic retinopathy. At the same time, classification accuracy is hugely dependent on the clinical experience of the physician.

The rapid development of deep learning today is expected to change this situation. Deep learning has been frequently used for image recognition and classification, and many developed algorithms have reached expert-level accuracy in DR recognition. In 2016, Gulshan et al.[4] and others used the Inception-V3 network to train on a dataset of 128,175 fundus images, eventually achieved a sensitivity of 97.5% on the three classification tasks of DR. The following year, Gargeya and Leng[5] obtained an AUC score of 97% on 75,137 color fundus images in Kaggle Data Set through 5x cross-validation.

Our work has the following contributions:
- We proposed a neural network model based on transfer learning to classify diabetic retinopathy.
- We utilized data pre-processing to improve the accuracy of our network.
- By comparing three methods commonly used in transfer learning, we find that the most suitable method for diabetic retinopathy grade recognition is to freeze the bottleneck layer and retrain the full-connection layer.
2. Related work

2.1. Assisted diagnosis of fundus images:
Various machine learning algorithms are now used to develop high-performance medical image processing systems such as computer-aided detection (CADe) system [6]. The application of the CAD system can significantly reduce the pressure on doctors to diagnose medical pictures and improve work efficiency. The most important thing is to reduce the influence of personal subjective factors on doctors’ diagnoses and improve the accuracy of clinical diagnosis. Therefore, we have developed this system to assist doctors in diagnosing diabetic retinopathy.

2.2. Transfer learning
Transfer learning has achieved promising results by leveraging knowledge from the source domain to annotate the target domain, which has few or none labels [7]. Therefore, with the limited amount of data obtained, we chose transfer learning technology to develop a Diabetic retinopathy identification system. Figure 1 is a schematic diagram of transfer learning used in this article. The convolution layer is frozen and transferred to a new network. In contrast, the full-connection layer is recreated and randomly initialized retraining, which significantly reduces the training duration and improves the accuracy of the network trained in the Messidor Data Set.

Figure 1. Schematic diagram of transfer learning.

3. Our method

3.1. System Structure
Figure 2 depicts how the system diagnoses fundus images in this paper to obtain a certain level of results. The image is pre-processed first and then placed in a pre-trained DenseNet[8] network. The network includes multiple image convolutions and image pools, then extracts image feature values, and outputs diagnostic results after full-connection layers.
3.2. Data Pre-processing
Before implementing transfer learning, we need to do some pre-processing on the original image.

First, the images in the Messidor Data Set have different sizes, so we achieved a uniform size of 2240 × 1488 for the images;

Then, the Messidor Data Set has only 1200 pics, and the number of pictures at different levels varies greatly, which may cause the original network to extract features inadequately during the transfer learning process and reduce the accuracy of the Retinopathy grade diagnosis. We classified mirrored and rotated the 1200 images according to the level to expand to 4 × 4000 fundus pictures. It overcomes the problems of data-lack and unevenly distributed data categories.

On the other hand, due to the limitation of imaging conditions, the color fundus images are mostly dark and undetailed, we improved the image contrast with a linear transformation to enhance the performance characteristics of the data.

Suppose the input picture is I, the width of a picture is w, the height is of picture h, and the output image is O. The following formula describes the process of a linear transformation, where a is the contrast coefficient, and b is the brightness coefficient:

\[ O(r, c) = a \cdot I(r, c) + b \quad (0 \leq r < H, 0 \leq c < W) \] (1)

3.3. Model selection and training

3.3.1. DenseNet model. DenseNet is a densely connected convolutional neural network. The core of DenseNet is to establish a dense connection between all the front and the back layers, which facilitates the reverse propagation of gradients during training. Figure 3 shows the dense connection mechanism of DenseNet. DenseNet's network structure consists mainly of Dense Block and Transition Layer, where Dense Block is a module that contains many layers, each of which is the same size as the feature map, with dense connections between layers. The Transition module connects two adjacent Dense Blocks and reduces the size of the Feature map by Pooling.
In a traditional convolutional neural network, there are $L$ layers and $L$ connections, but in a DenseNet network, there are $L(L + 1)/2$ connections. The transformation function for the layer $l$ is $H_l$ (usually corresponding to one or two groups of Batch -Normalization, ReLU, and Convolution operations), and the output is $x_l$. The following formula describes the transformation of each layer of DenseNet:

$$x_l = H_l([x_0, x_1, ..., x_{l-1}])$$  \hspace{1cm} (2)

One advantage of DenseNet is that the network is narrower and has fewer parameters, rather than other networks with hundreds of widths. The second advantage of DenseNet is the provincial calculation. This connection makes the transmission of features and gradients more efficient and the network easier to train. Another advantage of DenseNet is its resistance to overfitting and mitigates vanishing-gradient, so the DenseNet model is suitable for applications with relatively scarce training data.

Thus, DenseNet is a network model with superior features, we used the DenseNet network as a network model for transfer learning.

3.3.2. Transfer learning based on DenseNet. We used the Keras framework to load the weights of the pre-trained DenseNet model on the ImageNet Data Set to initialize the convolution layer and retrained the final full-connection layer to identify our image categories and shorten the duration of the training.

We froze layers outside the full-connection layer as a fixed feature extractor, calculating the values after each training set and testing set image passed through the frozen layers of the model - convolution "bottleneck". Since the convolution weight is not updated, we chose to store these values at the time of initial calculation to reduce redundancy and sped up training. Then, the newly initialized full-connection layer retrained the image "bottleneck" as input to achieve the effect of taking advantage of the existing model for new application scenarios.

Using Adam Optimizer with a learning rate of 0.001, we trained a total of 20 epochs with a random gradient descent method for a convolution layer of 40 images per batch. After each training, the test set independent of the images in the training set was used for the Holdout test, passing each image over the network without gradient descent and reverse propagation. The best execution model for analysis is retained to achieve the auxiliary diagnosis of diabetic retinopathy grade.

4. Experiments

4.1. Data Sets

Messidor is a frequently-used fundus image library, which established by the French Ministry of Defense Research in the TECHNO-VISION project funded in 2004. The 1200 eye fundus color digital images were captured using 8 bits per color plane at 1440 × 960, 2240 × 1488, or 2304 × 1536 pixels. It contains 540 normal images and 660 abnormal images. The medical experts have provided retinopathy grade for each image. Table 1 shows the Retinopathy grade and its diagnostic basis.

| Retinopathy grade | μA$^a$ | H$^b$ | NV$^c$ |
|------------------|-------|------|------|
| 0                | 0     | 0    | 0    |
| 1                | (0,5) | 0    | 0    |
| 2                | (5,15)| (0,5)| 0    |
| 3                | [15,+$\infty$) | [5,+$\infty$) | 1    |

$^a$Number of microaneurysms.
$^b$Number of hemorrhages.
$^c$1: Neovascularization. 0: No neovascularization.
4.2. Experiments

Generally, there are four ways to transfer learning:

1. Transfer the network and initialize the parameters randomly.
2. Transfer the network, initialize with the parameters of the pre-trained model, then update all parameters during training.
3. Transfer the network, initialize with the parameters of the pre-trained model, and fine-tune (freeze parameters for some layers, update the parameters of other layers during training).
4. Transfer the network, initialize with the parameters of the pre-trained model, and freeze the parameters of the other layers except for the full-connection layer.

To compare the advantages of the above four methods, we used the ImageNet model parameters trained by DenseNet201 network to carry out the comparison experiments of the above four methods using the Messidor Data with image pre-processing, the results of which are shown in Table 2.

Table 2. Results of the four training methods.

| Training method | Accuracy (%) | Training duration (s/epoch) |
|-----------------|--------------|-----------------------------|
| Method (1)      | 49.07        | 683                         |
| Method (2)      | 64.23        | 613                         |
| Method (3)      | 59.70        | 610                         |
| Method (4)      | 89.22        | 658                         |

The results showed that the method (4) was more accurate than the other three methods.

To verify the impacts of image-enhanced processing on training accuracy, we learned from the data without image-enhanced processing and the data after image-enhanced processing, using the training method (4). The accuracy value is elevated from 82.35% to 89.22%.

We selected five networks for comparison with DenseNet201. We used the dataset after image-enhanced processing for training, and chose method (4) to observe the accuracy and training duration of the six networks. The results are shown in Table 3.

Table 3. Comparison of training results.

| Network          | Accuracy (%) | Training duration(s/epoch) |
|------------------|--------------|-----------------------------|
| VGG19[9]         | 82.33        | 827                         |
| Inception-V3[10] | 78.74        | 650                         |
| MobileNet[11]    | 60.53        | 782                         |
| NasNetLarge[12]  | 72.54        | 659                         |

After a comprehensive analysis of the results obtained from the six types of network training, the DenseNet201 network had the highest accuracy rate, shorter training duration, and the best classification effect compared with other networks. It showed that DenseNet201 was superior as the network model for transfer learning in this system.

5. Conclusion

The application of transfer learning technology to the field of medicine has achieved some results. Based on the DenseNet201 model, this paper found that the transfer learning method of "freezing in addition to the full-connection layer" in recognition of diabetic retinopathy level has a better recognition effect. The experimental results show that transfer learning using the DenseNet201 network has a shorter training duration, and the advantages of high accuracy, better than VGG19, Inception-V3, and other networks. On this basis, we have developed an auxiliary diagnosis system for diabetic retinopathy to classify the level of diabetic retinopathy, which has a good effect on the classification of diabetic retinopathy.

Acknowledgements

This paper was supported by Scientific and Technological Development Program Foundation of Jilin Province, China (No. 201604054YY; No. 20170414006GH)
References

[1] Wenjin, Li. (2019) Application of systematic nursing intervention in patients with diabetic retinopathy[J]. Journal of Clinic Nursing’s Practicality, 4(30): 31.

[2] Li, Z., Lili, Z., Hui, W. (2014) Effects of health education and psychological intervention on quality of life of diabetic retinopathy patients[J]. Journal of Clinical Research, 33(9): 1862-1863.

[3] Alexandre, G. S., Fabio, M., Marina, S. F., et al. (2019) Fundus photography quality assessment based on topological extinction values[J]. International Journal of Computer Applications in Technology, 59(3): 224-235.

[4] Gulshan, V., Peng, L., Coram, M., et al. (2016) Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs[J]. Journal of the American Medical Association, 316(22): 2402-2410.

[5] Gargeya, R., Leng, T. (2017) Automated identification of diabetic retinopathy using deep learning[J]. Ophthalmology, 124: 962–969.

[6] Al-antari, Mugahed A., Al-masni, Mohammed A., et al. (2018) A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification[J]. International Journal of Medical Informatics. 117: 44-54.

[7] Zhiqi, S., Jindong, W., Yiqiang, C., et al. (2017) Balanced Distribution Adaptation for Transfer Learning[J]. IEEE International Conference on Data Mining(ICDM). New Orleans. pp. 1129-1134.

[8] Huang, G., Liu, Z., Maaten, L. V. D., et al. (2017) Densely Connected Convolutional Networks[C]. In: IEEE Conference on Computer Vision and Pattern Recognition. Hawaii. pp. 2261-2269.

[9] Simonyan, K., Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. The 3rd International Conference on Learning Representations (ICLR2015).

[10] Szegedy, C., Vanhoucke, V., Ioffe, S., et al. (2015) Rethinking the Inception Architecture for Computer Vision[C]. In: IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas. pp. 2818-2826.

[11] Debjyoti, S., El-Sharkawy. (2019) Thin MobileNet: An Enhanced MobileNet Architecture[C]. In: 2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference. New York. pp. 0280-0285.

[12] Zoph, B., Vasudevan, V., Shlens, J., et al. (2018) Learning Transferable Architectures for Scalable Image Recognition[C]. In: IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City. pp: 8697-8710.