IMAGE DETECTION OF DENTAL DISEASES BASED ON DEEP TRANSFER LEARNING

Jiakai Zhang
Computer & software school
Hangzhou Dianzi University
Hangzhou, China
290709356@qq.com

Xiaodong Li*
Computer & software school
Hangzhou Dianzi University
Hangzhou, China
* Corresponding author:hzxiaodong22@163.com

Zhigang Gao
Computer & software school
Hangzhou Dianzi University
Hangzhou, China
Gaozhigang@hdu.edu.cn

Jing Chen
Computer & software school
Hangzhou Dianzi University
Hangzhou, China
chenjing@hdu.edu.cn

***Abstract*—Traditional dental disease detection is done by doctors using naked eyes directly, which contains many uncertain factors for misdiagnosis and missed diagnosis. In order to improve the accuracy and efficiency of the detection of dental diseases, a dental disease image detection assistance system based on deep transfer learning is designed, which can autonomously recognize the photos obtained from the camera that assists the doctor in the detection. Performing transfer training on the trained model on the tooth data set, retain all pretrained convolutional layer parameters, and fine-tune the model to be more suitable for tooth image recognition. At the same time, AlexNet, GoogLeNet, and VGG models will be used for traditional deep learning training and the results obtained will be compared and analyzed with the results obtained by deep transfer learning in terms of accuracy and timeliness.**

Keywords-dental disease; convolutional neural network; deep transfer learning

I. INTRODUCTION

Currently, dental diseases are mainly diagnosed by manual observation methods, which have disadvantages such as lack of objectivity and accuracy and low efficiency, while automatic classification of medical images by using machine vision-based image processing techniques can eliminate these disadvantages[1]. In recent years, many researchers at home and abroad have conducted a lot of research on the application of machine vision-based image processing techniques for the detection of diseases, and have made great progress[2-6]. Defect feature extraction is a key step in the detection of dental diseases using image processing techniques, and it directly affects the accuracy of disease diagnosis.

Convolutional neural network (CNN) learns features directly from image data extraction and has a strong generalization ability to features.CNN has been successfully applied to research areas such as handwritten digit recognition, face recognition[8], and has made some progress in dental disease classification[7]. Although convolutional neural networks have achieved better results in the above studies, a large amount of data is needed to obtain convolutional neural network models with better classification and recognition performance. Especially for complex background images, convolutional neural networks need a large number of training samples to learn the ability to extract image features[9-11]. In contrast, transfer learning uses a convolutional neural network architecture that has been trained on large data sets to obtain better feature extraction capabilities, which can reduce the number of samples required to train the network model[13].

In this paper, dental images are used as training samples in a complex background with large multi-angle lighting variations, and the training of migration learning is performed based on convolutional networks AlexNet[20], GoogLeNet[21], VGG[12], and ResNet models, using migration learning methods to train models with higher recognition accuracy using smaller training samples. Also, in order to explore the training results of the above mentioned traditional convolutional neural networks, this paper compares the results obtained from the initialized model and the migration learning model of the above mentioned models. After conducting experiments, it is found that the accuracy and convergence time of using migration learning to identify dental diseases are better than the traditional convolutional neural network model mentioned above.

II. NETWORK STRUCTURE TRAINING FOR TRANSFER LEARNING

A. Deep Transfer Learning

Deep transfer learning is a deep learning algorithm. It re-uses the knowledge learned from solving a problem and applies it to another different but related problem to solve it. This requires the use of the model in a domain (source domain), then transfer to another domain (target domain) to obtain a better learning effect [13-15].

Here are some concepts about transfer learning:

Domain: a domain consists of a feature space and a marginal probability distribution on the feature space, where.represents the distribution of.

978-1-6654-2186-7/21/$31.00 ©2021 IEEE
DOI 10.1109/ICAICE54393.2021.00151

Task: After a given domain, a task consists of a label space and a conditional probability distribution, which is usually learned from the training data composed of feature-label pairs.

Source domain, target domain: In the concept of transfer learning, the source domain is the knowledge that has been learned, and the target domain is the new knowledge that we have to learn through the previous knowledge.

Negative transfer [16]: refers to the knowledge acquired from the source domain, which does not have a positive effect on the subsequent learning in the target domain, but has a negative effect. There are two situations where this happens:

The first situation is that the above two domains have very low relevance and almost no similarities. Therefore, the previous learning is not helpful to the subsequent learning and cannot be used for transfer learning.

Another situation is that there is no problem with the correlation between the above two domains, but the transfer learning itself is not perfect, and there is no component suitable for transfer, which leads to a negative effect of transfer learning.

The simple use of transfer learning is to fine-tune the network with a pre-trained model, while deep network training requires a large number of training samples for obtaining a better model. ImageNet[17] is the largest image recognition database at present, containing 1000 categories, covering a wide range, and rich in sample diversity. Therefore, parameters are obtained from the model trained by ImageNet, and they are adjusted according to the current state of the new data set to solve the problem of few-shot learning and the long training time. It can train a network with good performance.

In recent years, the research of transfer learning has received more and more attention. Under the situation that deep learning has been widely used, the shortcomings of deep learning have also become more and more exposed, and transfer learning can play a good role in few-shot learning by generalizing the knowledge learned in the past[18]. This has greatly helped many current research situations. Researchers such as Yosink[19] conducted extensive and in-depth research on the portability of the features learned from ImageNet. They conducted research by fine-tuning different data sets and found that if the selected basic task is farther from the target task, then Portability will be lower.

The drawbacks of deep learning are also becoming more and more evident, while transfer learning can serve well to generalize previously learned knowledge in the absence of training data. This has been very helpful in many research situations nowadays.Yosink et al.[19] researchers conducted an extensive and in-depth study on the portability of features learned from ImageNet, they did it by fine-tuning different datasets and found that if the basic task selected is further away from the target task, then the portability will be lower.

B. *AlexNet Model*

AlexNet has an 8-layer structure, which consists of 5 convolutional layers and 3 fully connected layers[20]. AlexNet uses the ReLU function to alleviate the previous gradient dispersion problem and implements a local response normalized LRN, which applies Dropout to the front randomly deactivated neurons in the fully connected layer to reduce the overfitting problem.

C. *VGG Model*

AlexNet has an 8-layer structure, which consists of 5 convolutional layers and 3 fully connected layers[20]. AlexNet uses the ReLU function to alleviate the previous gradient dispersion problem and implements a local response normalized LRN, which applies Dropout to the front randomly deactivated neurons in the fully connected layer to reduce the overfitting problem.

D. *GoogLeNet Model*

The VGG network contains 16 layers. The structure of the network is very consistent, using small 3×3 convolutional kernels and small 2×2 pooling kernels from beginning to end. By stacking small convolutional kernels instead of large convolutional kernels, more activation functions are used, and more features are obtained, resulting in better discrimination ability. In addition to the above, if we only talk about the role of convolution itself, the 3×3 convolutional layer can find more feature changes than the 7×7 convolutional layer; the 3×3 convolutional layer, the middlemost grid on the surface of the nine-box grid is a perceptual field center, and the center point can capture the feature changes that occur around[20]. The three 3×3 stacking approximates a 7×7. This structure deepens the number of network layers and also uses the ReLU function twice more, which expands the network capacity and can distinguish different classes more clearly.

E. *ResNet Model*

GoogLeNet proposes a new network structure, Inception structure, to solve the phenomena of overfitting, gradient explosion and gradient disappearance brought by the deepening of network layers[20]. This structure can fuse the features of different scales together. In addition to this structure, GoogLeNet also uses 1×1 convolutional kernel to reduce the dimensionality of the whole structure data volume, because the use of fully connected layers is abandoned, so Google adds two auxiliary classifiers to help training, from the previous Maxpool to use the average pooling layer, which can greatly reduce the model parameters. The model is as follows.

F. *Network Classification*

ResNet network has a very deep network layers, breaking the above convolutional neural network model for training will appear degradation phenomenon with the increase of depth[20]. This degradation is mainly manifested by the fact that as the number of layers increases, the training set loss gradually decreases and then tends to saturate, and when you increase the depth of the network, the same phenomenon as overfitting occurs, and the training set loss increases instead. ResNet's deep network brings very good classification performance, and the main reason that enables it to reach the deep network is that it uses a structure called the residual network.

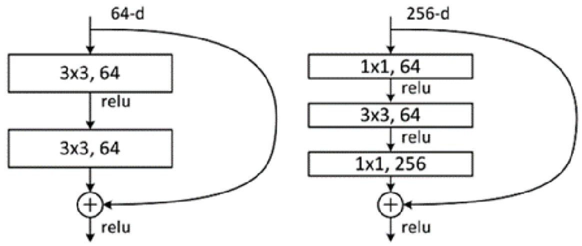


Figure 1. Deep residual network model

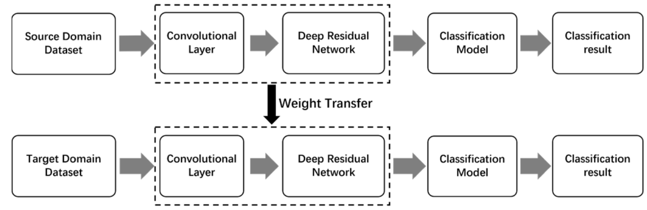


Figure 2. The transfer learning framework based on ResNet.

The above figure is the model of deep residual network, such a structure can map the information of the shallow network to the deep layer to solve the problem of gradient disappearance caused by the deep network. The residual structure on the left is for networks with fewer layers, such as ResNet18 and ResNet34 networks. The right side is for networks with more network layers, such as ResNet101, ResNet152, etc. In this experiment, given that tooth recognition is only binary classification for cavities and non-cavities, the ResNet34 network with fewer layers, which is the network on the left in the figure above, is finally chosen in this paper. The main branch of this residual structure is composed of two 3×3 convolutional layers, while the connection line on the right side of the residual structure is the shortcut branch also known as the shortcut branch. In order to allow the output matrix on the main branch to be summed with the output matrix on our shortcut branch, it is necessary to ensure that these two output feature matrices have the same shape.

G. *Network Classification*

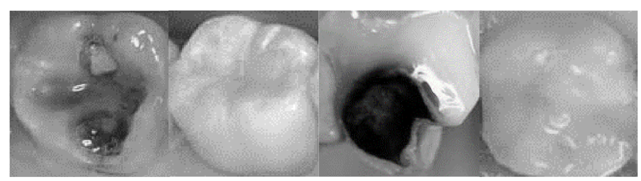
The advantages of each model have been listed above, and the next step is to use them. In terms of models, the data structure of the above models is now publicly available, so only the training aspect needs to be considered. Since we do not have a lot of dental disease samples, direct training may cause overfitting and the trained model may not have a good classification effect, which is not suitable for our purpose. The research shows that convolutional neural networks have the ability to generalize and can migrate features learned from training on a database of a domain to a database of a similar domain, that is, migration learning. The training method is composed of two parts, the first part is pre-training, in order to make this part of the model accuracy is high using iterative network training and optimization of the parameters in the model to achieve, and then convergence model output pre-training model, overall pre-training steps are more similar to random initialization. The second part is migration training, in which the parameters of the trained model are migrated, and then the network is fine-tuned to optimize the parameters by our given data set to better suit our classification purpose. The migration learning process is shown in the following Figure 2 (using ResNet34 as an example).

III. EXPERIMENT

A. *Experimental Settings*

The experimental environment of this paper: Intel i9-9900k processor, 32G RAM, NVIDIA RTX2080Ti GP, Ubuntu OS 16.04, Python 3.7, and PyTorch platform is used for training.

The dataset used for the training of the classification experiments in this paper are 1000 images of single teeth, which are divided into two attribute annotations of decayed and non-decayed teeth, 800 images each, with one attribute annotation for each image. In addition to this, 200 images of teeth were used to test the accuracy of the trained model. The difficulty of classifying this datasets is that the tooth under different light and different angle will have an impact on the dental health condition judgment. The dental images are shown in Figure 3 below.



(a) (b) (d) (e)

Figure 3. Dental datasets, (a)(c)decay teeth pictures, (b)(d)non-cavity teeth pictures.

In view of the small number of images in this experimental datasets, the ten-fold cross-validation method was used during the experiment. The images of both decayed and healthy teeth in the dental dataset were split into ten subsamples, each containing 80 images, 9/10 as training samples and 1/10 as validation samples. This approach can fully utilize the training samples and obtain more information.

In order to implement migration learning, the network model is trained on the selected source domain dataset in advance. In this paper, the source domain dataset selected is ImageNet database, as the largest image recognition database at present, ImageNet covers rich samples with category representation information, which makes the above network models can be sufficiently supervised training. Meanwhile, the parameters of AlexNet, VGG, GoogLeNet, and ResNet34 model weights trained based on the ImageNet database have been publicly available on the Internet, which makes the otherwise very complicated work simple, saves the computer thousands of iterations, and provides help to our ordinary computers for subsequent training.

Table 1. The comparison of evaluation indicators of various model experiments

Model Name	Correct number of decayed teeth	Correct number of non-decayed teeth	Total number of correct predictions	Correct Rate（%）	Number of iterations	Convergence time（s）
AlexNet	84	57	141	70.5%	100	800
VGG	93	78	171	85.5%	110	880
GoogLeNet	92	69	161	80.5%	110	1150
ResNet34	97	78	175	87.5%	120	770
ResNet34- T	100	94	194	**97%**	55	130
GoogLeNet-T	100	90	190	**95%**	60	250
VGG-T	100	94	194	**97%**	60	240
AlexNet-T	100	87	191	**93.5%**	60	200

B. Experimental Results and Analysiss

In Figure 4 shows that the models after transfer learning have a very large increase in the convergence speed of the models, which converge in about 60 iterations, and the accuracy of the models has improved significantly compared with the original models, with all models reaching an accuracy rate of over 95%.

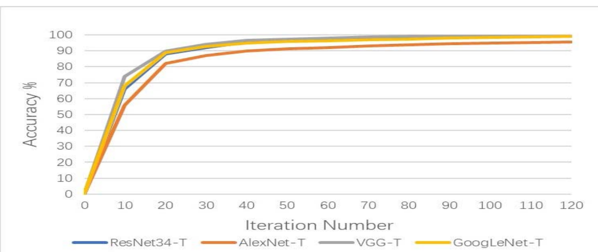


Figure 4. Model training curve using transfer learning.

In Table 1, the accuracy results of the original models of AlexNet, VGG, GoogLeNet, ResNet34 and their transfer learning models are shown after the training is completed and the models are tested using 200 test samples. The table can be seen intuitively and once again verifies that migratory learning has improved significantly in both accuracy and convergence time for this recognition of dental disease images.

IV. Model Test Results

The analysis of the correct number of decayed teeth and non-decayed teeth shows that the accuracy of both traditional convolutional neural networks and transfer learning networks in determining non-decayed teeth is lower than the accuracy of decayed teeth, and some models differ by a very large amount. This may be due to the fact that there are some negative samples to mislead the judgment e.g., shadows of teeth, original depressions of teeth, et al. From the data, the transfer learning is also less disturbed than the traditional convolutional neural network in the face of these distracting information.

Overall, the network model using deep transfer learning is far superior to their original convolutional neural network model in this dental disease recognition experiment in terms of accuracy, convergence efficiency and interference resistance, and also shows that transfer learning can perform quite well in a small sample training set.

V. Conclusions

In this paper, deep transfer learning based dental disease recognition model is used to evaluate their validity for detection tasks in dental image. Experiments show that the accuracy, convergence efficiency and anti-interference ability of the model after deep transfer learning are very much improved. In this paper, we have used this method to identify and classify dental disease images with a high degree of precision, which can reduce the possibility of missing and misdiagnosis caused by the naked eye in the manual recognition of dental diseases. We further assessed the best transfer models by using visual explanations with different activation mapping in future.

Acknowledgment

X.D. is supported by Zhejiang Provincial Natural Science Foundation of China (Grant No. LY19F0200).

References

[1] M. Frid-Adar, I. Diamant, E. Klang , et al., “ GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification”, Neurocomputing, 2018, 321(DEC.10):321-331.

[2] L. Cai , J. Gao , D. Zhao, “ A review of the application of deep learning in medical image classification and segmentation”, Annals of Translational Medicine, 2020, 8(11).

[3] X. Yao, J. Han, D. Zhang, et al., “ Revisiting Co-Saliency Detection: A Novel Approach Based on Two-Stage Multi-View Spectral Rotation Co-clustering”, IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2017, 26(7):3196-3209.

[4] J. Y. Gao, X. S. Yang, T. Z. Zhang, et al., “ Robust Visual Tracking Method via Deep Learning”, Chinese Journal of Computers,2016,39(07):1419-1434.

[5] S. Xie, X. Zheng, Y. Chen, et al., “ Artifact Removal using Improved GoogLeNet for Sparse-view CT Reconstruction”, Scientific Reports, 2018, 8(1):6700.

[6] P. Zhou, G. Cheng, Z. Liu, et al., “ Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping”, Multidimensional Systems and Signal Processing, 2016, 27(4):925-944.

[7] D. F. Wang, C. B. Chen, T. L. Ma,et al., “Dental disease recognition system based on convolutional neural network”, Foreign Electronic Measurement Technology, 2019,38(6):93-97.

[8] W. Yang , L. Jin, D. Tao, et al. ," DropSample: A New Training Method to Enhance Deep Convolutional Neural Networks for Large-Scale Unconstrained Handwritten Chinese Character Recognition", Pattern Recognition, 2016, 58(4):190-203.

[9] J. S. Wang, Y. Chen, Z. Q. Zeng, et al.,"Extraction of litchi fruit pericarp defect based on a fully convolutional neural network", Journal of South China Agricultural University,2018,39(06):110-116.

[10] E. Mohebi, A. Bagirov, "A convolutional recursive modified Self Organizing Map for handwritten digits recognition", Neural Networks, 2014, 60:104-118.

[11] L. Santos, L. Castro, N. Rodriguez-Fernandez, et al., " Artificial Neural Networks and Deep Learning in the Visual Arts: a review", Neural Computing and Applications, 2021, 33(1):1-37. .

[12] J. Gu , Z. Wang, J. Kuen, et al.," Recent Advances in Convolutional Neural Networks", Pattern Recognition, 2015.

[13] Y. GUO, H. SHI, R. A. KUMA , et al. SpotTune: transfer learning through adaptive fine-tuning,"Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", 2019: 4805-4814.

[14] D. Han, Q. Liu , W. Fan," A new image classification method using CNN transfer learning and webdata augmentation", Expert Systems with Applications, 2018, 95(2): 43-56.

[15] Y. Wen, K. Zhang, Z. Li, et al., " A Discriminative Feature Learning Approach for Deep Face Recognition", European Conference on Computer Vision. Springer, Cham, 2016.

[16] V. M. Patel , R. Gopalan , R. Li , et al. ,"Visual domain adaptation: a survey of recent advances", IEEE Signal Processing Magazine, 2015: 32(3): 53-69.

[17] A. Krizhevsky, I. Sutskever, G. E. Hinton," ImageNet classification with deep convolutional neural networks", Advances in Neural Information Processing Systems, 2012: 1097-1105.

[18] H. C. Shin , H. R. Roth, M. Gao, et al.," Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning", IEEE Transactions on Medical Imaging, 2016, 35(5):1285-1298.

[19] J. Yosinski , J. Clune , Y. Bengio , et al.," How transferable are features in deep neural networks?", Advances in neural information processing systems, 2014, 27.

[20] A. Khan, A. Sohail, U. Zahoora, et al.,"A Survey of the Recent Architectures of Deep Convolutional Neural Networks", Artificial Intelligence Review, 2019(1-87).

[21] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview", Neural Network, 2015, 61:85-117.