Classification of Alzheimer's disease in MobileNet

Xiaoling Lu, Haifeng Wu * and Yu Zeng
School of Electrical and Information Technology, Yunnan Minzu University, Kunming, Yunnan, 650000, China
*Corresponding author’s e-mail: whf5469@gmail.com

Abstract. As the aging of Chinese society becomes more and more serious, the number of elderly people has increased dramatically. At the same time, the number of patients with Alzheimer's disease (AD) has increased. At present, the main diagnostic method for Alzheimer's disease relies on experienced radiologists to analyse brain structural nuclear magnetic resonance (MRI) images to judge the condition, but this method is time-consuming and labor-intensive, and there is a certain subjectivity. This may cause misdiagnosis. By classifying the MRI images of patients with Alzheimer's disease and healthy controls (NC), for image classification, the Convolutional Neural Network (CNN) in deep learning has outstanding performance and accurate classification. The VGG 16 network model and the MobileNet network model of the convolutional neural network are compared, using deep learning and transfer learning. We can find that the MobileNet network model is superior to the VGG 16 network model in classification accuracy.

1. Introduction
Alzheimer's disease (AD) is one of the most common forms of dementias and a kind of progressive and neurodegenerative disease. The disease cannot be cured by current medical methods [1] and its cause has not been known so far. If doctors can correctly diagnose the Alzheimer's disease, patients can take right medication to slow down the condition. At present, the main diagnostic method for Alzheimer's disease is to rely on experienced radiologists to analyse the brain structure nuclear magnetic resonance image (MRI) to judge the condition. However, such a method is time-consuming, laborious and subjective, and may lead to misdiagnosis. Present studies have shown that compared with the normal control (NC) group, AD patients have a reduction in gray matter volume in the temporal lobe, hippocampal formation, insula and left thalamus [2]. MRI is a biomagnetic spin imaging technology, which obtains images with clarity, high resolution, good contrast and large amount of information. Besides, MRI images can clearly distinguish between gray matter and white matter, and can better show brain structure and reflect some subtle change. The most important thing is that there is no harmful radiation to the human body and thus has no adverse effects.

AD can be diagnosed by means of neuropsychological testing, neuroimaging, EEG analysis, and cerebrospinal fluid examination. With the development of computer technology and imaging technology, many classification methods of machine learning are widely used in medical image analysis of AD. In 2014, Falahati et al. used multivariate data analysis and machine learning to process MRI images of AD [3]. In 2016, Salvatore et al. performed early diagnosis of AD through MRI brain image and support vector machine [4]. In 2017, YANG C H proposed an AD classification recognition algorithm based on principal component analysis and linear discriminant analysis fusion, in order to classify MRI images of AD [5]. In 2018, ZHOU WEN et al. used the kernel principal component analysis method to extract the features of the gray matter image, and combined the Adaboost algorithm to classify AD [6]. In 2019,
Krishnakumar Vaithinathan et al. used a new T1 weighted MRI texture extraction technique for the classification of AD [7]. It is feasible to extract the eigenvalues of MRI images of AD and NC by using traditional machine learning methods and classify MRI images. However, the accuracy of classification depends on the extracted features. It is difficult that the features are extracted manually. If the extracted features are not very correct, the accuracy of the classification will not be very high.

Therefore, the method we use is deep learning, which can automatically extract features without artificial errors, which can improve the accuracy of classification. Two network models of convolutional neural networks are used to classify the MRI images of AD and NC, which are VGG 16 network model and MobileNet network model. The VGG 16 network model is a 16-layer network built by the Visual Geometry Group, and won the second place in the image classification of the 2014 Large Scale Visual Recognition Challenge (ILSVRC). On June 15, 2017, Google opened up the code for the MobileNet network model. The MobileNet network model uses a deep separable convolution to replace the traditional convolution, which reduces the redundancy of the convolution kernel and reduces complexity. So far, no one has used the MobileNet network model to classify MRI images of AD and NC.

However, deep learning is training from scratch, which has three limitations:
1) Since the training data is trained under the CPU, it will lead to excessive training time;
2) Deep learning requires a lot of training data, but the training data we need comes from the expensive medical imaging field, and the training data set we have is not enough to support.
3) Since deep learning is a deep neural network, its complexity is very high compared to machine learning.

Therefore, transfer learning is adopted, which is one of the research hotspots in the field of machine learning, and has been widely used in data mining and computer vision. Transfer learning is to discover and learn useful knowledge from relevant learning areas to improve the performance of learners in the target learning field. Since deep learning requires a large-scale training data set, transfer learning can be employed in the case of only small training data sets. Transfer learning uses the pre-trained weights and MRI datasets to initialize the network layer, and then retrains the new fully connected layer with training data, simply fine-tuning the final fully connected layer of the convolutional neural network. There are already some pre-training weights in Keras, which gives us a very strong pre-training weight, and using these weights as our initial values can increase the training speed.

The network structure of the MobileNet network model is simpler than the VGG 16 network model, and the complexity is relatively small, and the training speed is relatively fast. The final experimental results show that the accuracy of classification of MRI images of AD and NC by transfer learning combined with VGG 16 network model and MobileNet network model is 92% and 94%. It can also indicate that the accuracy of classification of MobileNet network model is higher than that of VGG 16 network model.

2. Related work

Using the above machine learning method to extract the features of MRI images of AD and NC, and classify MRI images. these methods have two shortcomings:

(1) The traditional machine learning method requires manual extraction of features. If the extracted features are not very correct, the accuracy of the classification will not be high.

(2) The accuracy of classification of MRI images of AD and NC by machine learning is not high compared to deep learning.

Therefore, this paper uses the two network structures of deep learning to extract the MRI image features of AD and NC, and classify the MRI images. Due to the high cost of medical data, the data set does not support the traditional deep learning of training from scratch, so the transfer learning is adopted.

As shown in Figure 1, in transfer learning, the existing knowledge is called the source domain, the new knowledge to be learned is called the target domain, and the domain is composed of data features and feature distribution. Transfer learning studies how to transfer knowledge from the source domain to the target domain. In the field of machine learning, transfer learning studies how existing models can be applied to new, different but related areas. Traditional machine learning is not flexible enough and the
results are not good enough when dealing with tasks such as data distribution, dimensions, and output changes of the model. Transfer learning reduces these assumptions. Under the conditions of data distribution, feature dimension and model output change, the knowledge in the source domain is used to better model the target domain. In addition, in the absence of calibration data, transfer learning can make good use of the calibration data in the relevant field to complete the calibration of the data. Formal definition of transfer learning: Given a source domain \( s_D \), a learning task \( s_T \) on the source domain, a target domain \( t_D \), and a learning task \( t_T \) on the target domain, we use \( s_D \) and \( s_T \) to learn the prediction function \( f() \) on the target domain. The task is composed of the function and the learning result.

Transfer learning has the following four advantages:
1) Reusing existing knowledge domain data, a lot of work has not been completely discarded;
2) There is no need to spend a huge price to re-collect and calibrate large new data sets;
3) Good results can also be obtained for small data sets;
4) For new areas that are emerging rapidly, they can be quickly transferred and applied, reflecting the timeliness advantage.

3. Problem description
The use of deep learning methods can improve the accuracy of MRI image classification for AD and NC. But there is a big problem, deep learning itself requires a very large data set. The symptoms of AD patients are also diverse, and the stages of AD are different, which leads to subtle differences in MRI images. Therefore, in order to make the experimental results more convincing, it is necessary to make the data set include all the symptoms and different stages. Most importantly, MRI images of AD and NC are not so easy to obtain, and we don't have equipment to get data sets for AD and NC. Our data comes from the OASIS website, one of the two largest Alzheimer's databases in the world. However, these downloaded data sets do not fully support the traditional deep learning method of training from scratch. In short, it is ok to start training from scratch, but the accuracy of its testing will be very low. If you want high accuracy, you need a large training data set. Therefore, we use the transfer learning method to obtain a higher accuracy rate when training a smaller training data set without obtaining a large training data set.

4. Method

4.1. Complexity analysis
As shown in Figure 2, the difference between the traditional CNN of the MobileNet network model and the VGG 16 network model is mainly that the traditional CNN is a 3×3 convolution layer in front of batch normalization and ReLU. But the convolution of the MobileNet network is separable convolution. Separable convolution is to decompose the standard convolution into a deep convolution and a 1×1 point convolution, placed before batch normalization and ReLU.

The calculation cost of the standard convolution is

\[
D_K \times D_K \times M \times N \times D_F \times D_F
\]
\( D_F \) is the spatial width and height of the input features, \( M \) is the number of input channels, \( D_K \) is the spatial width and height of the output features, and \( N \) is the number of output channels. The calculation cost depends on the number of input channels \( M \), the number of output channels \( N \), the size of the convolution kernel \( K \) \((D_K \times D_K)\), and the size of the feature \( F \) \((D_F \times D_F)\). First it uses a deeply separable convolution to break the interaction between the number of output channels and the size of the kernel. Standard convolution has the effect of filtering features based on convolution kernels and feature combinations to produce new representations. The filtering and combining steps can be integrated into two steps by using a deeply separable convolution to greatly reduce computational cost.

The MobileNet model uses a deeply separable convolution which integrates the standard convolution into a deep convolution and a point convolution. For MobileNet, deep convolution applies a single filter to each input channel, and then point convolution uses a \(1 \times 1\) convolution to combine the deep convolution of the output. A standard convolution simultaneously filters and combines the inputs to get a new set of outputs. The deep separable convolution divides it into two layers, one for filtering and the other for combination. This decomposition has the effect of greatly reducing the amount of calculation and the size of the model.

Deep convolution is very effective compared to standard convolution. But it just filters the input channels and doesn't combine them to create new features. Therefore, an additional layer is needed to calculate the linear combination of the output of the deep convolution via the \(1 \times 1\) convolution to produce these new features. The combination of deep convolution and \(1 \times 1\) convolution is called deep separable convolution. The computational cost of deep separable convolution is

\[
D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F
\]

(2)

The calculated cost ratio of the depth separable convolution to the standard convolution is

\[
\frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F}
\]

(3)

The MobileNet network model uses a depth-separable convolution to reduce the computational complexity by 8-9 times compared to traditional standard convolution, reducing complexity and simplifying the model.

At the same time, the MobileNet network model adds two hyperparameters: the width coefficient and the resolution coefficient. While the underlying MobileNet model is small, in many cases, a particular application may require the model to be smaller and faster. In order to construct these smaller computational models, a very simple parameter \( \alpha \), called the width coefficient, is introduced. The effect of the width coefficient \( \alpha \) is to evenly thin the network at each layer. Simply put, it is the ratio of the number of convolution kernels to be used by each module in the new network compared to the standard MobileNet. The computational cost of a depth separable convolution with a width coefficient of is

\[
D_K \times D_K \times \alpha M \times D_F \times D_F + \alpha M \times \alpha N \times D_F \times D_F
\]

(4)

| Width Multiplier | ImageNet Accuracy | Million Mult-Adds | Million Parameters |
|------------------|-------------------|-------------------|-------------------|
| 1.0 MobileNet-224| 70.6 %            | 569               | 4.2               |
| 0.75 MobileNet-224| 68.4 %          | 325               | 2.6               |
| 0.5 MobileNet-224| 63.7 %            | 149               | 1.3               |
| 0.25 MobileNet-224| 56.6 %            | 41                | 0.5               |

\( \alpha \in (0,1) \), and \( \alpha \) is typically set to 1, 0.75, 0.5 and 0.25. When \( \alpha = 1 \), it is the standard MobileNet. When \( \alpha < 1 \), it is a simplified MobileNet, and the width coefficient has the effect of reducing the calculation cost and the number of secondary parameters. The width coefficient can be applied to any model structure, defining a new, smaller model with reasonable precision, latency and size trade-offs, which is used to define a new simplified structure. As shown in Table 1, we can see that when \( \alpha \) is used to reduce the network parameters, the accuracy on ImageNet decreases with the decrease of \( \alpha \).
The second hyperparameter used to reduce the computational cost of the neural network is the resolution coefficient $\beta$. People apply it to the input image, and then the internal representation of each layer is subsequently reduced by the same coefficient.

The computational cost of the network core layer can now be expressed as

$$D_K \times D_K \times aM \times \beta D_F \times \beta D_F + aM \times aN \times \beta D_F \times \beta D_F$$

(5)

$\beta \in (0, 1)$, and usually the input resolution of the network is 224, 192, 160 or 128. When $\beta = 1$, it is the standard MobileNet, when $\beta < 1$, it is the simplified calculation of MobileNet. Reduce the cost of calculation by $\beta$. As shown in Table 2, when a different coefficient $\beta$ is applied to the standard MobileNet, the accuracy decreases with the decrease of coefficient $\beta$.

| Resolution | ImageNet Accuracy | Million Mult-Adds | Million Parameters |
|------------|-------------------|-------------------|--------------------|
| 1.0 MobileNet-224 | 70.6 % | 569 | 4.2 |
| 1.0 MobileNet-192 | 69.1 % | 418 | 4.2 |
| 1.0 MobileNet-160 | 67.2 % | 290 | 4.2 |
| 1.0 MobileNet-128 | 64.4 % | 186 | 4.2 |

4.2. Transfer learning network architecture

As shown in Figure 3, the MRI images are first preprocessed, and then the bottleneck features are extracted by using the convolution layer part, and then the newly designed fully connected layer is connected to form a new network structure. We use transfer learning to load pre-training weights on new networks, and finally train networks and classify the MRI images.

The fully connected layer is designed according to the top layer of the MobileNet network and the VGG 16 network. It is a linear stack of multiple neural network layers. As shown in Figure 4, a flat layer is created firstly, which converts the image created in the previous step into a one-dimensional vector. MobileNet's fully connected layer is different from VGG 16 in this layer. It needs to establish a global average pooling layer to average the data. Next, we create a hidden layer with a total of 256 neurons and use the ReLU activation function. Then, we add Dropout to avoid overfitting. Finally, the output layer is built with only 1 neuron, and use the sigmoid activation function to convert to represent the probability.
5. Experiment

5.1. Experimental data
Our MRI data is all from the OASIS website, one of the two databases of Alzheimer’s disease worldwide. To test the generalization ability of transfer learning, a three-dimensional brain image was cut into 176 two-dimensional MRI images. Then, based on the image entropy, we selected 32 MRI images with the highest image entropy. The experiment only used these limited numbers of images to train the network. This data set contains a total of 5120 training samples and 1280 validated samples. The format of these MRI image samples is jpg and the size is 176×208. In order to be compatible with the pre-training models of VGG16 and MobileNet, the image size of VGG16 is set to 150×150, and the image size of MobileNet is 160×160. The experiment is mainly based on the Linux system Ubuntu using Keras deep learning framework written in Python with Tensorflow as the back end.

Figure 5. MRI image: AD (left), NC (right).

5.2. Analysis of results
In this section, we provide experimental results of the VGG 16 network model and the MobileNet network model. Our goal is to analyze MRI data by convolutional neural network to distinguish AD patients from NC.

Figure 6. (a) Accuracy of VGG 16 and MobileNet on the training set; (b) Accuracy of VGG 16 and MobileNet on the validation set; (c) Loss of VGG 16 and MobileNet on the training set; (d) Loss of VGG 16 and MobileNet on the validation set.
Figure 6 shows the accuracy and loss of the VGG 16 network model and the MobileNet network model on the training set and the validation set. As the number of iterations increases, the accuracy of the model gradually converges. When the number of iterations reaches 200, the accuracy of VGG 16 converges to 93% and the accuracy of MobileNet converges to 98% on the training set. On the validation set, VGG 16 converges to 92% and MobileNet converges to 94%. It is clear that the MobileNet network is more accurate than the VGG 16 network. As the number of iterations increases, the loss of the model gradually converges.

6. Conclusion
This paper uses deep learning and migration learning to verify the performance of the VGG 16 network model and the MobileNet network model for MRI image classification. The experimental results can be concluded as follows: First, the accuracy of the MobileNet network model is higher than that of the VGG 16 network model in terms of classification accuracy. Second, the complexity of the MobileNet network model is lower than that of the VGG 16 network model. In summary, the MobileNet network model is superior to the VGG 16 network model in classifying MRI images of AD and NC.

However, the work in this paper is only a two-category of MRI images of AD and NC. If mild cognitive impairment (MCI) can be diagnosed, treatment can be performed earlier. In the following work, we will further classify AD, MCI and NC. Then, further improvements are made in the accuracy of the classification. If the accuracy of MRI image classification of mild cognitive impairment is high, it can be relieved by drugs or treatment in advance.

Acknowledgments
This work was supported in part by the National Natural Science Foundation of China under Grant No. 61762093. The 17th batches of Young and Middle-aged Leaders in Academic and Technical Reserved Talents Project of Yunnan Province under Grant No. 2014HB019. The Program for Innovative Research Team (in Science and Technology) in University of Yunnan Province. The Key Applied and Basic Research Foundation of Yunnan Province under Grant No. 2018FA036. The Science Research Fund Graduate Program in Education Department of Yunnan Province under Grant No. 2018Y106. The graduate student innovation fund research project in Yunnan Minzu University under Grant No. 2018YJCXS175.

References
[1] Gerardin, E., Chételat, G., Chupin, M., et al. (2009) Multidimensional classification of hippocampal shape features discriminates Alzheimer’s disease and mild cognitive impairment from normal aging. Neuroimage, 47(4): 1476-1486.
[2] Yang, J., Pan, P. L., Song, W., et al. (2012) Voxelwise meta-analysis of gray matter anomalies in alzheimer’s disease and mild cognitive impairment using anatomic likelihood estimation. Journal of the Neurological Sciences, 316(1-2): 21-29.
[3] Falahati, F., Westman, E., Simmons, A., (2014) Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. J Alzheimers Dis, 41(3), 685–708.
[4] Salvatore, C., Battista, P., Castiglioni, I., (2016) Frontiers for the early diagnosis of AD by means of MRI brain imaging and support vector machines. Curr Alzheimer Res, 13(5): 509.
[5] Yang, C.H., Yu, C.J., (2017) Alzheimer's disease classification algorithm based on fusion principal component analysis and linear discriminant analysis. J Xiamen Univ NatSc i, 56(2): 226–230.
[6] Zhou, W., Wang, Y., Xiao, H.B., Cao, L.H., (2018) Auxiliary diagnosis of Alzheimer’s disease based on KPCA algorithm. Chinese Journal of Medical Physics, 35(4): 404–409.
[7] Vaithinathan, K., (2019) A Novel Texture Extraction Technique with T1 Weighted MRI for the Classification of Alzheimer’s Disease. Journal of Neuroscience Methods, 318: 84–99.