

Pathological Images Classification based on the Pretrained Convolutional Neural Network

Qifeng Wang *

Department of Automation, Wuhan University of Technology, Wuhan, 430070, China

* Corresponding author's email: 265720@whut.edu.cn

Abstract. Convolutional neural network is more and more used to solve the problem of images classification, but the scale of the dataset in this filed is too small compared with that of natural image recognition, which leads to the slow convergence speed of the neural network, and the instability of the model, which is difficult to be used to predict the results. In this paper, the residual network is applied to the PCam dataset to realize the recognition of tumour cells, and pre-trained through the ImageNet dataset. In the same training epochs, the performance of the pre-train network is much higher than that of the randomly initialized network, which significantly improves the convergence speed of the neural network. In addition, the performance of the pre-train network can converge faster when the amount of data is sufficient, and in the case oof medical image recognition with less data, the performance of the pre-train network is much higher than that of the random initialization network.

Keywords: Pretrained Convolutional Neural Network; Residual Network; PCam.

1. Introduction

Applying artificial intelligence methods to medical image recognition has various benefits. Traditionally, human doctors have made judgments regarding patients based on a variety of data, including symptoms and physical examinations, that can be standardized and processed by Artificial Intelligence (AI). And AI can quickly get results to assist doctors make diagnosis. Some physiological structure images are too complicated for human eyes to recognize their features, but neural networks in deep learning can process a large number of image data at one time, learn a large number of cases, and identify potential features, so as to make better judgment. In addition, there are wide kinds of medical conditions that a human specialist might be good at identifying only one or two types of medical images, while an AI can learn to recognize images of different conditions.

However, there are still some issues with medical image recognition. Data-driven approaches, while achieving good results on different visual tasks, are known to be based on large training sets with detailed annotations. Some natural image datasets are in the millions, such as the MS COCO dataset with over 1million images, and the ImageNet dataset with over 20000 categories and over 14 million images. When solving the problem of medical image classificaiton, the relevant data sets are small. Even if the number of dataset is large enough to accomplish the training target, they need to be labeled because images of different conditions vary greatly. This process requires a great deal of effort. Using a huge dataset of medical images directly to train neural networks is costly and time-consuming. Therefore, other data sets can be used to pre-train the neural network, and then medical images can be utilized to train the pre-train neural network, and to reduce the training time and the accuracy of neural network.

In order to analyse biological images, studies have established some relevant data sets [1-3]. However, compared with natural image recognition, the data amount is limited, and the data sets are too small for stable prediction model and transfer learning. In [2], the authors established a data set for the detection task of organ anatomy. Although it contains multiple organs, the amount of data is far less than the natural image data set. In [4], data from multiple medical datasets are aggregated and synthesized into one 3D dataset containing different diseases, organs, and classifications, but a series of pre-train models still need to be trained to speed up convergence and accuracy, this data set can then be used to solve other, different kinds of medical problems.

Pre-train networks are often used to solve the problem of missing medical images. The pre-train network has many other advantages. Many studies show that applying the pre-train model to medical image recognition can make the model converge quickly and improve the accuracy of the network. In [5], the pre-train model based on ImageNet has a faster convergence rate in the early training period than the randomly initialized model, and only needs 10% training data at least. In [6], the authors had pre-trained the image classification, detection and segmentation problems of radiology, cardiology and gastroenterology. Compared with the convolutional neural network trained from scratch, the pre-train model is more robust. In [7], due to the lack of large amounts of data, the authors used ImageNet data set pre-trained neural network and applied it to renal ultrasound image detection, obtaining better performance. The ImageNet dataset does not consider the weight of medical images, so whether it can be directly transferred to the problem of medical image classification remains to be investigated. These articles do not specifically explain the advance and improvements of pre-train networks on medical image recognition. Therefore, this paper compares ImageNet based pre-train network with zero-based training network and explains the generalization of pre-train network.

2. Method

2.1 Dataset Description and Processing

The data set used in this study comes from the Kaggle competition [8]. It is derived from PCam [9], and the repeated data in the original data set is eliminated. It provides 220,000 RGB case images for classification, and the images are divided into two categories, positive and negative. The positive label means that the 32x32px region in the center of image contains at least one tumour tissue. Tumour cells in the outer area of the image does not affect the label. Figure 1 provides some sample pathological images on this dataset.

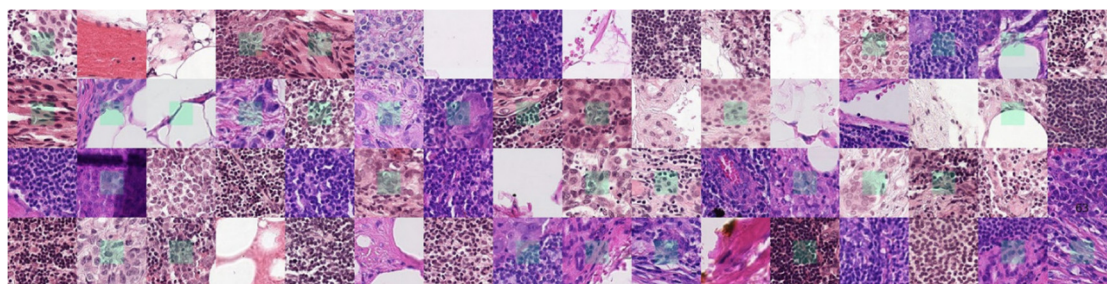


Figure 1. The pathological images

The image was normalized, and the value of each pixel point in each channel was reduced to 0-1. Since there are fewer positive data in the data set, image flipping and rotation are employed to enhance the image and increase the amount of positive data to avoid sample imbalance and strengthen the robustness of the model.

2.2 CNN Model based on Transfer Learning

Convolutional neural network has been frequently used to solve the image recognition problem in recent years. Its excellent performance comes from the convolutional and pooling layers. Basic convolutional neural networks are usually composed of multiple convolutional layers, and each convolutional layer contains multiple convolutional units. In the training process, the back propagation algorithm can be used to continuously iterate the model, so as to optimize the parameters in each convolution unit. The convolution operation can effectively extract various features of the input vector. For example, simple geometric features such as edges, angles and lines can be detected and extracted by low-level convolution layers, while complex features can be iteratively trained from the multi-level networks. Pooling layer is a form of down-sampling. In the pooling layer, the input image is divided into many individual rectangular regions according to the size and mapping

relationship of the pooling layer, and the corresponding values are output. The pooling layer is similar to the human visual system. By adding the pooling layer to the convolutional neural network, the model can have the ability of the biological vision system to simplify and abstract the input image.

Although the convolutional neural networks like AlexNet and VGG have achieved relatively deep neural networks by modifying the convolution kernel and other ways, they have failed to solve the problem of gradient explosion or gradient disappearance. For traditional convolutional network or fully connected network, as the input vector passes through the neural network, some information might be lost along the way, and gradient disappearance or gradient explosion can be caused, which hinders network convergence and makes it impossible to train deep networks [10, 11]. In [12], Residual Neural Network was proposed, after introducing its unique network structure, the training speed and accuracy of the model have been greatly improved. In ResNet there exists a structure called Highway Network. This structure is a channel that directly connects different convolutional layers. In the absence of this network structure, the neural network performs nonlinear transformation on the input vector, and the output of some convolutional layers leads to the problem mentioned above. After using this network, the input vector will be output in a certain proportion to avoid this problem.

In the medical image recognition problem, compared with the data set used for natural image recognition, training a model directly on a small dataset will result in a less stable model. Based on transfer learning method, using the "pre-train and fine-tune" paradigm, ImageNet is used to pre-train the network, which can not only increase the stability of the neural network, but also improve the convergence speed of the model [13, 14]. By pre-training the network with a marked data set with a large amount of data, and then applying the network to the target task domain, the pre-train layer can discover the high-dimensional features of the data. Since the tumour prediction problem only needs to predict two classes, and ImageNet is a thousand classification problem, the pre-train network needs to be fine-tuned to change the final fully connected layer to a binary classification output.

2.3 Implementation Details

All models are trained in GPU, based on the Pytorch framework. The learning rate is 0.001. The batch size is 100. For the optimization part, the Adam algorithm is chosen as the optimizer, and the cross-entropy function is used as the loss function.

Before training the network, 90% of the images are used to train the models and 10% of the images are used as the test set. For the pre-train model, the last layer is replaced with the binary classification output after the pre-training of the residual network is completed.

Considering the small amount of data in many medical datasets, in order to simulate this situation, a sub-dataset containing 10K images is extracted from the experimental dataset, and the performance of the randomly initialized network and the pre-train network are compared on this dataset.

3. Results and Discussion

The precision of the network during the same training period is used as the index to compare the pre-train network and the randomly initialized network. Table 1 shows the precision, the recall and the network accuracy of the pre-train network and the randomly initialized network for positive samples on the test set at 5, 10, 15, and 20 epochs. Figure 2 illustrates the convergence processes of the pre-train network and the randomly initialized network on the training set.

Obviously, the pre-train network converges much faster than the randomly initialized network. At the early stage of training process (epoch=5), the accuracy of the pre-train network is 7% higher than that of the randomly initialized network. However, when the training epochs gradually increase, the performance of the randomly initialized network gradually approaches that of the pre-train network. When epoch=20, the accuracy of the pre-train network is 1.6% higher than that of the randomly initialized network.

Table 1. Accuracy, precision and recall of the pre-train network and the randomly initialized network on the test set at different epochs

		5 epochs	10 epochs	15 epochs	20 epochs
Randomly initialized network	Accuracy	79.51%	85.24%	94.15%	96.22%
	Precision	75.55%	96.62%	97.74%	95.55%
	Recall	97.50%	79.82%	92.74%	98.25%
Pre-train network	Accuracy	85.02%	91.23%	97.31%	97.82%
	Precision	94.46%	96.62%	97.97%	98.35%
	Recall	76.72%	88.53%	93.73%	98.79%

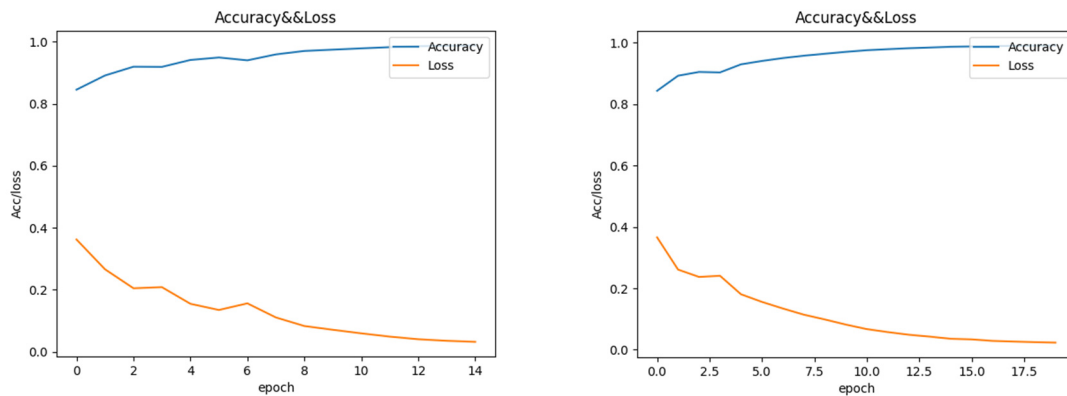


Figure 2. The convergence process of the pre-train network and the randomly initialized network on the training set. LEFT: The pre-train network converges at 15 epochs. RIGHT: The randomly initialized network converges at 20 epochs.

Table 2 provides the precision, the recall for positive sample and the network accuracy on the small sample sub-dataset. When dataset has only 10K images, the pre-train network can still achieve 95% accuracy after several iterations. And randomly initialized networks with the same or more iterations failed to achieve the same result.

Table 2. Accuracy, precision and recall of the pre-train network and the randomly initialized network trained with fewer images on the test set at different epochs

		5 epochs	10 epochs	15 epochs	20 epochs
Randomly initialized network	Accuracy	79.71%	91.93%	92.27%	92.59%
	Precision	88.65%	88.40%	90.34%	94.04%
	Recall	67.85%	98.92%	93.12%	91.91%
Pre-train network	Accuracy	86.62%	93.47%	97.02%	97.13%
	Precision	80.15%	89.83%	96.25%	97.74%
	Recall	96.33%	99.19%	97.47%	97.34%

In order to further study the influence of pre-training on the network, the ResNet with different layers and different VGG networks were selected for the experiment. Table 3 shows the accuracy of different networks at 5, 10, 15 and 20 epochs after random initialization or pre-trained.

For different network structures, compared with random initialization, the pre-trained network can extract features faster and converge as soon as possible. By pre-training the network, the network obtains a good prior probability distribution, thus improving the ability of the network to extract features. However, when the randomly initialized network is trained by multiple epochs, it can also

learn a good prior probability distribution. In this process, the ability of the randomly initialized networks will gradually catch up with the ability of the pre-trained networks.

Table 3. Accuracy of different pre-train networks and the randomly initialized networks at different epochs

	5 epochs	10 epochs	15 epochs	20 epochs
ResNet50 random init	78.79%	89.32%	92.44%	94.45%
ResNet50 pre-trained	84.24%	91.35%	94.54%	97.57%
ResNet152 random init	80.89%	88.21%	95.50%	97.53%
ResNet152 pre-trained	86.00%	93.55%	97.12%	97.55%
VGG16 random init	76.35%	84.69%	90.59%	94.62%
VGG16 pre-trained	86.27%	90.42%	94.06%	95.80%
VGG19 random init	78.24%	84.83%	90.36%	95.56%
VGG19 pre-trained	85.47%	90.92%	94.29%	96.63%

4. Conclusion

In the process of applying ImageNet pre-training network to medical image recognition, due to the lack of relevant background weight, although the deep neural network may not learn the features related to the application situation, the first few layers can still extract some features, and these simple features have some generality.

In general, when the training dataset such as some natural image dataset is large enough to accomplish the training goal, ImageNet pre-training cannot greatly strengthen the stability and the prediction accuracy of the network, the model that are not pre-trained can still achieve the same effect through multiple iterations and training, but through ImageNet pre-training, the network obtains a good prior probability distribution, thus improving the ability of the network to extract features, we can obtain a better initial network, get better results at the beginning of training, and makes the network converge faster than the randomly initialized network.

When the training dataset is a small dataset, such as a medical dataset, the final performance of the network can be better improved through ImageNet pre-training. Although ImageNet pre-training does not have medically relevant weights, it still enables the neural network to learn the basic geometric structure on the image, resulting in a better initialization of the network, which reduces the sample size required for training. Because by using ImageNet pre-training, the neural network can achieve better and faster results with fewer datasets, so in the case of small samples, such as medical image recognition, ImageNet pre-training is still necessary.

Only one tumour dataset was used in the experiment, and no more data sets and other network structures were used. More complex data sets with richer data types should be used in the following research to improve the universality of the conclusion.

References

- [1] Armato S G et al. 2011 The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical physics 38 2 915-31.
- [2] Toro O A et al. 2016 Cloud-Based Evaluation of Anatomical Structure Segmentation and Landmark Detection Algorithms: VISCERAL Anatomy Benchmarks IEEE Transactions on Medical Imaging 35 2459-2475.
- [3] Raudaschl P F et al. 2017 Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015 Medical Physics 44 2020–2036.

- [4] Chen S 2019 Med3D: Transfer Learning for 3D Medical Image Analysis ArXiv, abs/1904.00625.
- [5] He K et al. 2019 Rethinking ImageNet Pre-Training 2019 IEEE/CVF International Conference on Computer Vision (ICCV) 4917-4926.
- [6] Tajbakhsh N et al. 2016 Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Transactions on Medical Imaging 35 1299-1312.
- [7] Ravishankar H et al. 2016 Understanding the Mechanisms of Deep Transfer Learning for Medical Images. LABELS/DLMIA@MICCAI.
- [8] Kaggle 2019 Histopathologic Cancer Detection <https://www.kaggle.com/competitions/histopathologic-cancer-detection/data>.
- [9] Github 2020 Pcam <https://github.com/basveeling/pcam>.
- [10] Krizhevsky A et al. 2012 ImageNet classification with deep convolutional neural networks Communications of the ACM 60 84-90.
- [11] Simonyan K et al. 2015 Very Deep Convolutional Networks for Large-Scale Image Recognition CoRR abs/1409.1556.
- [12] He K et al. 2016 Deep Residual Learning for Image Recognition 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 770-778.
- [13] Qiu Y et al. 2019 Semantic segmentation of intracranial hemorrhages in head CT scans in 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) p 112-115 IEEE.
- [14] Meng J et al. 2022 TL-Med: A two-stage transfer learning recognition model for medical images of COVID-19 Biocybernetics and Biomedical Engineering.