Study for Performance of Un-Pretrained and Pre-trained Models based on CNN

Bingsen Wang *

School of Software and Microelectronics, Peking University, Beijing, 100871, China

* Corresponding author's email: 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, Artificial Intelligence (AI) auxiliary diagnosis systems have attracted more and more attention. In this paper, some commonly used Convolutional Neural Network (CNN) models e.g. MobileNet, VGG and ResNet are trained and compared on the cancer detection dataset, and it is found that the pre-trained models based on the idea of the transfer learning perform better than the newly trained models in terms of training speed and model performance. Thus, it can be seen that the transfer learning method has great potential in the field of cancer diagnosis. This study provides some experimental support and suggestions on how to further improve the property of the transfer learning method in the field of cancer diagnosis. Meantime, the performance of VGG19 can be proved to be better compared to other models (i.e., MobileNet and ResNet).

Keywords: CNN; Un-Pretrained and Pre-trained; VGG19.

1. Introduction

With the rapid development of artificial intelligence, especially deep learning, computers have been employed more and more in the field of medical diagnosis, e.g., cancer diagnosis, and have started to gradually replace the role of medical experts. However, due to the huge amount of data in the medical field, it requires a lot of time to train the model to get excellent performance. To solve this issue, some studies attempting to use the transfer learning method to address the cross-domain application of artificial intelligence in order to obtain a shorter training time, on the premise of ensuring the performance of the model [1-3].

The capability of a system to recognize and apply information and abilities obtained in prior tasks to fresh tasks is a new target of transfer learning [4]. That is, the goal of transfer learning is to apply knowledge gained from one or more source tasks to a target task [5]. In a nutshell, some well-performing pre-trained models trained on standard public datasets can be used and applying these pre-trained models to new target tasks. The goal of domain adjustion is to learn invariant features from the domain so that the learned features are not limited to be only used in the source domain and lead to over-fitting. The performance of the related models is all tested on the Office Dataset and there are many excellent results in Domain Adaptation. Ganin and Lempitsky used the similarity of source and target domain to update the model, so that the designed model can learn target domain's feature without target label [6]. Long et al. proposed a brand-new deep neural network design for domain adaptation, wherein "deep adaptation" is achieved by layer-by-layer adaption of all the layers matching to task-specific characteristics [7].

Although the performance of transfer learning methods on some standard general datasets is excellent [8, 9], there is still a lack of experimental results based on practical application scenarios in general. In the field of medical image recognition, artificial intelligence has shown performance that surpasses that of the human brain [10]. The only factor limiting the large-scale application of artificial intelligence in the medical image recognition field is the high cost of model training caused by massive data. If the transfer learning method has outstanding performance in the medical image recognition field, then artificial intelligence will soon be widely applied in the field of medicine. In this article, some deep convolutional neural network models that have been proven to perform excellent in image classification tasks were selected, e.g., Mobilenet, VGG and ResNet, and evaluate the performance of pre-trained and newly trained models separately on the cancer detection dataset from Kaggle, which consists of a training set of 220k photos and a test set of 57.5k photos. The results of experiments show that on the cancer detection dataset, the pretrained models perform better than the newly trained models.

The contributions of this article are summarized as follows. (1) This study evaluated the performance of the pre-trained models and the newly trained models of different models in the cancer detection dataset, and tried to introduce the transfer learning method into the cancer detection task, hoping to effectively reduce the model training cost. (2) This study introduced transfer learning methods in new target domains and provide data support for the further development of transfer learning methods through experimental evaluation.

2. Methods

2.1 Dataset Preparation and Preprocessing

This paper used the Cancer Detection dataset provided by Kaggle [11]. This dataset is obtained by a simple modification of the PatchCamelyon (PCam) standard dataset (the original PCam dataset contains repeated images because of its probability sampling, but the Kaggle version does not contain copies). The dataset consists of 277, 483 small RGB pathology images and a patch's central 32x32px region has at least one pixel of tumor tissues when the label is affirmative. The label is unaffected by tumor tissue in the patch's external region. This external region is supplied to allow completely convolutional models that do not use zero-padding, to get stable behavior when applied to an entire image. Since the dataset has been pre-processed when it is provided, we did not do any other processing on the dataset, but only converted the label files from csv format to txt format. Figure 1 shows some sample photos.

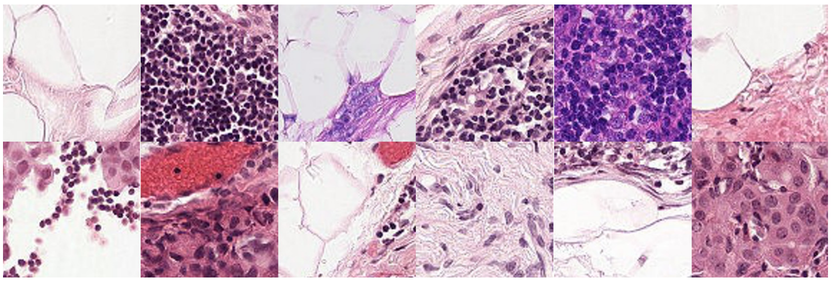

Figure 1. Sample photos of the Cancer Detection dataset

2.2 CNN Models

One of the representative deep learning algorithms is the convolutional neural network (CNN), a type of feedforward neural network including deep network structure and convolution computation. Convolution, pooling, and fully connected layers are the three main types of layers (or building blocks) that make up this mathematical construction [12]. The CNN model's convolutional layer is crucial, and the layer mainly extracts image features through convolution, a special linear mathematical operation. The pooling layer imitates the human visual system to minimize the complexity of the data, and represents the image with higher-level features. This paper selected some representative CNN models such as MobilNet, VGG and ResNet for experiments. These models have their own advantages. MobileNet introduces depthwise separable convolution to make the model very lightweight. VGG proposes the idea of replacing network layers with basic blocks, which makes it possible to reuse these basic blocks when building deep network models. ResNet introduces a residual module, which successfully solves the degradation problem of deep networks. ResNet consists of a stack of basic blocks as shown in Figure 2.

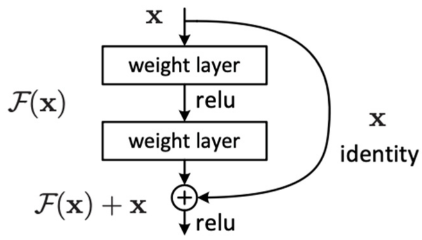

Figure 2. A Basic Block

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 $\mathcal{H}(x)$, but to fit the residual function $\mathcal{F}(\mathrm{x})$.

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

After solving the network degradation problem, we can increase the network depth by stacking more basic blocks to obtain better performance. As shown in Figure 3, it is the architecture of 34-layer ResNet, 34-layer plain network and 19-layer VGG network.

2.3 Implementation Details

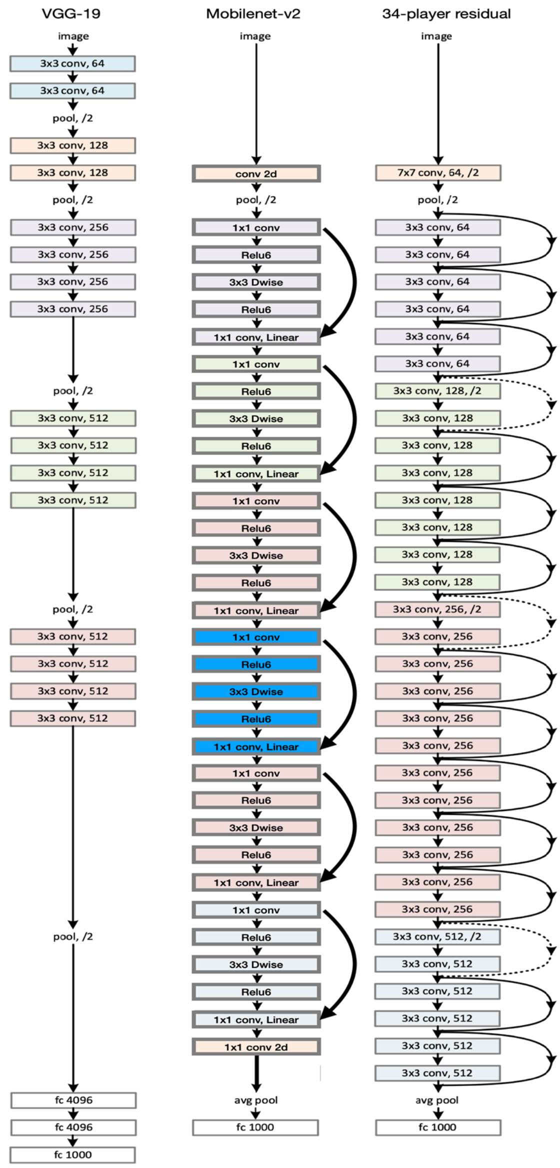

Figure 3. Architecture of different CNN models

In experiments, this paper uses accuracy (as shown in Eq.2) to evaluate the performance of different pre-trained and newly trained models.

$$Accuracy = \frac{TruePositive+TrueNegative}{Total} \tag{2}$$

When loading the pre-trained model, this paper selects different pre-trained models trained on the ImageNet dataset. When training the model, in order to achieve better performance, this paper sets the learning rate to 0.001, the epochs to 20 and the batch size to 100. This article uses the GPU to train the model under the Pytorch framework. In addition, the loss function in this paper is the cross-entropy loss function, and the optimization algorithm is the Adam algorithm.

Besides, since the CNN model trained based on ImageNet is not a two-class model, and what this paper is going to do is a two-class model, we make some adjustments to the output layer of the pre-trained models to enable it to complete the two-class task. Finally, when training the pretrained model, this study uses a two-step fine-tune strategy:(1) The first step freezes all parameters except the fully connected layer, and the optimizer only optimizes the parameters of the fully connected layer. (2) Then unfreeze all parameters and let the optimizer optimize all parameters with a decreasing learning rate.

3. Results and Discussion

As shown in Table 1, the following results through experiments were obtained: (1) Models newly trained on the target dataset perform worse than those pre-trained on the generic dataset. (2) For newly trained models, larger models perform better (this is not shown in pretrained models), but the performance improvement brought by the increase in model size is limited.

Table 1. Performance of Different CNN Models

Model	Performance		
	Training Accuracy	Testing Accuracy	Training Time(s)
Mobilenet(pre1)	0.86	0.8468	3019
Mobilenet(pre2)	1.00	0.9781	8185
Mobilenet(new)	0.99	0.9426	5174
VGG19(pre1)	0.79	0.8316	5870
VGG19(pre2)	1.00	0.9837	20099
VGG19(new)	1.00	0.9683	14672
ResNet101(pre1)	0.87	0.8686	5896
ResNet101(pre2)	1.00	0.9824	20681
ResNet101(new)	1.00	0.9754	15012

Figure 4 shows the variation of each model's training accuracy during training. In the first step of training, the pre-trained model did not show significantly better performance than the newly trained model, and the training accuracy fluctuated after the first epoch. In the second training step, the pre-trained model quickly showed excellent performance, reaching the maximum training accuracy and starting to fluctuate in a small range after the first epoch. Based on this, it can be concluded that pre-trained models can achieve better performance faster than newly trained models.

Since pre-trained models have been trained on a larger and more general dataset, they must far outperform the randomly initialized newly trained models in terms of feature extraction, and also have better initial performance than newly trained models. However, the experimental results did not support this conclusion well. By tracking the model performance of each batch in the first epoch of the training process, this study found that the pretrained model did show significantly better performance than the newly trained model, as shown in Figure 4. Furthermore, since in the first step of training, only the fully connected layers of the pre-trained model are trained, there are few parameters to optimize, so it can converge very quickly to the initial performance of the pre-trained model.

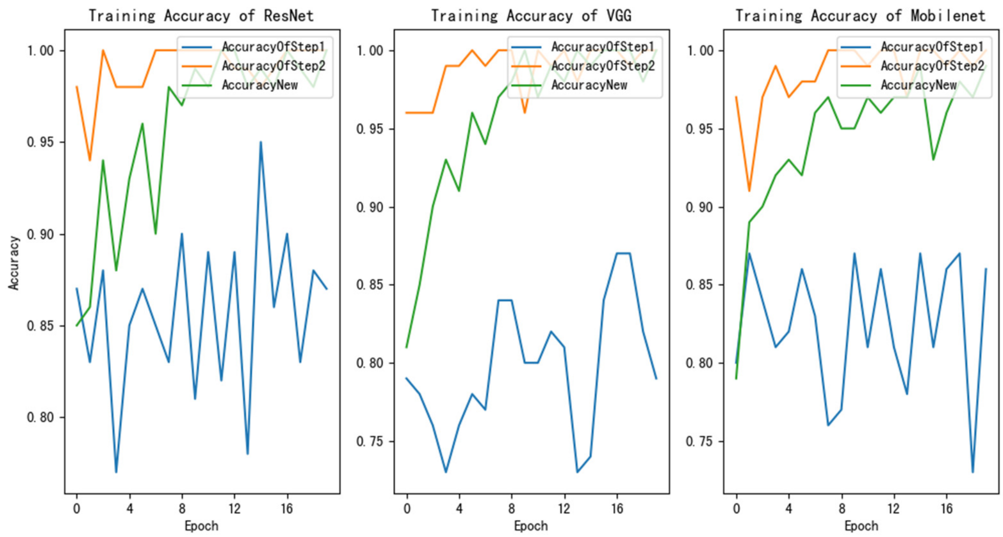

Figure 4. Training Accuracy of CNN models

From the perspective of training speed, the pre-trained model converges to very good performance after 1 and 3 epochs respectively in the two steps, while the newly trained model often has good test accuracy after 10 epochs. This is also likely because the pre-trained model has stronger feature extraction capabilities, and after training with massive data, it already has strong generalization capabilities, and it is already close to the global optimum for image recognition. Therefore, as long as it is simply trained on the target dataset, the model can quickly get excellent performance.

4. Conclusion

In this paper, various CNN models were employed to evaluate the performance in the case of pretraining and unpretraining. This study adopts a simple two-step strategy, and each step is trained for 20 epochs. It is necessary to try fewer epochs and different learning rate variations. Also, it's worth trying the strategy of gradually unfreezing parameters. Finally, it is experimentally verified that the performance of the newly trained model gets better performance as the model size increases, but the reason why the pretrained model in this paper does not follow this pattern is also an interesting question.

References

[1] Long M et al. 2015 Learning transferable features with deep adaptation networks JMLR.org.

[2] Long M, et al. 2013 Transfer feature learning with joint distribution adaptation IEEE International Conference on Computer Vision.

[3] Wang, X et al. 2014 Flexible transfer learning under support and model shift p 1898–1906.

[4] Pan S J et al. A survey on transfer learning 2010 IEEE Transactions on Knowledge and Data Engineering 22(10) 1345–1359.

[5] Pan S J et al. 2010 Domain adaptation via transfer component analysis IEEE transactions on neural networks 22(2) 199–210.

[6] Ganin Y et al. 2016 Domain-adversarial training of neural networks Journal of Machine Learning Research 17(1) 2096–2030.

[7] Long M et al. 2015 Learning transferable features with deep adaptation networks JMLR.org.

[8] Chai Z et al. 2022 Hierarchical shared transfer learning for biomedical named entity recognition BMC bioinformatics 23(1) 1-14.

[9] Qiu Y 2019 Semantic segmentation of intracranial hemorrhages in head CT scans In 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) p112-115 IEEE.

[10] Russakovsky O et al. 2015 Imagenet large scale visual recognition challenge International Journal of Computer Vision 115(3) 211–252.

[11] Kaggle 2019 Histopathologic Cancer Detection Data Sets https://www.kaggle.com/c/histop-athologic-cancer-detection/data.

[12] Yamashita R et al. 2018 Convolutional neural networks: an overview and application in radiology Insights into imaging 9(4) 611–629.