Recognition of Benign and Malignant Breast Ultrasound Images Based on Deep Transfer Learning

kaiwen wu (✉ 1220632895@qq.com)

Guangdong University of Finance&Economics <https://orcid.org/0000-0001-5289-4024>

Bo Xu

Guangdong University of Business Studies: Guangdong University of Finance & Economics

Ying Wu

First Affiliated Hospital of Jinan University

Research Article

Keywords:

Posted Date: June 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-249760/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Recognition of benign and malignant breast ultrasound images based on deep transfer learning

Kaiwen Wu · Bo Xu · Wu Ying

Received: date / Accepted: date

Abstract Manual recognition of breast ultrasound images is a heavy workload for radiologists and misdiagnosis. Traditional machine learning methods and deep learning methods require huge data sets and a lot of time for training. To solve the above problems, this paper had proposed a deep transfer learning method. the transfer learning models ResNet18 and ResNet50 after pre-training on the ImageNet dataset, and the ResNet18 and ResNet50 models without pre-training. The dataset consists of 131 breast ultrasound images (109 benign and 22 malignant), all of which had been collected, labeled and provided by UDIAT Diagnostic Center. The experimental results had shown that the pre-trained ResNet18 model has the best classification performance on breast ultrasound images. It had achieved an accuracy of 93.9%, an F1-score of 0.94, and an area under the receiver operating characteristic curve (AUC) of 0.944. Compared with ordinary deep learning models, its classification performance had been greatly improved, which had proved the significant advantages of deep transfer learning in the classification of small samples of medical images.

Keywords Deep learning · transfer learning · breast ultrasound images · image classification · breast cancer

F. Kaiwen Wu
School of Information Science,Guangdong University of Finance and Economics, Guangzhou,
Guangdong
E-mail: 1220632895@qq.com

S. Bo Xu
School of Information Science,Guangdong University of Finance and Economics, Guangzhou,
Guangdong
E-mail: xubo807127940@163.com

T. Wu Ying
Ultrasound Department,The First Affiliated Hospital of Jinan University, Guangzhou,
Guangdong
fax :510630

1 Introduction

Breast cancer is considered to be one of the cancers with the highest incidence in women in the world. The incidence and mortality of breast cancer in Chinese women are relatively low in the world. The overall survival rate is the same as that of developing countries, but there is a trend of rapid growth[1]. Early detection of lumps in a timely manner is essential to improve patient survival and cure rates[2]. Both X-ray and CT imaging in the existing medical imaging technology will cause different degrees of damage to human health. Magnetic resonance imaging (MRI) is expensive and complex in imaging and is not suitable for critical and urgent cases. Ultrasound imaging is cheap, convenient, and fast, and the most important thing is that it has no radiation and no harm to patients. In recent years, ultrasound imaging technology has become more mature, and new ultrasound technologies such as ultrasound holography, 3D color Doppler and color Doppler imaging have emerged. In the diagnosis of breast tumors, it can be used to determine the extent of the lesion and its physical properties. It is one of the main imaging methods for the diagnosis of breast lesions, especially breast cancer[3]. However, the traditional manual analysis of ultrasound images is not only time-consuming and labor-intensive, but also the doctor's misdiagnosis rate is high. The emergence of computer-aided diagnosis (CAD) systems effectively solves this problem. It can provide a reliable reference to assist radiologists in determining whether the tumor is benign or malignant[4].

In recent years, Deep Convolutional Neural Networks (DCNN) have become a popular technology for natural image classification. At the same time, in the field of medical imaging, DCNN is also very useful in tasks such as image segmentation, classification, and anomaly detection. However, compared with natural image processing, the application of DCNN technology to medical images has obvious shortcomings: because the labeling of medical images requires a large number of experienced radiologists, In most cases, the collection and labeling of medical images is very difficult. At the same time, patient privacy issues also increase the difficulty of acquiring ultrasound images[5]. These problems are fatal to deep learning that requires massive data support. The application of transfer learning to medical images effectively solves the above limitations. Usually, transfer learning first pre-trains the neural network on the source domain (such as ImageNet, which contains 15 million images and 1000 classified and annotated super large data sets), and then you can choose to freeze a part of the convolutional layer of the model or only Use pre-trained weights to start training from scratch, and finally fine-tune the model to adapt the classifier to medical image classification[6]. Research has found that the pre-trained transfer learning model can effectively improve the accuracy of medical image classification. Therefore, in this article, we plan to use the CNN model from scratch and the fine-tuned CNN model after pre-training to perform benign and malignant breast tumor classification experiments on 131 breast ultrasound images.

2 Related studies

2.1 Transfer learning

Although deep learning has the advantage of automatically extracting expressive features and can meet end-to-end needs, it requires a large amount of labeled data to support it. Medical images involving privacy issues are usually not easy to obtain, and the labeling of medical images requires experienced radiologists, which makes obtaining a large amount of labeled data extremely expensive. At this time, pre-training the neural network on a larger data set and fine-tuning it for our task can achieve higher accuracy on a smaller data set in less time. Although the classification of breast ultrasound images is very different from the classification of natural images, we can still use the general features extracted from the first few layers of the neural network. These general features are the similarities between the source domain and the target domain, which can help us complete the transfer of knowledge[7].

2.2 Residual network

Han et al. [9]used a deep learning framework (GoogleNet network) to classify 9 types of breast lesions and nodules obtained from 10 types of ultrasound imaging. The ultrasound image data set included 7408 cases (4254 cases of benign/3154 cases of malignant), including 17 types of benign. Malignant tumors. The final experimental results have exceeded 0.9 for both the expanded data and the unexpanded data. In another study by Zhang et al. [10], CNN models (InceptionV3, VGG16, ResNet50, and VGG19) were used to train on 5000 breast ultrasound images and tested on 1007 images, of which the InceptionV3 model got the largest The AUC (0.905).

Xiao T et al. [11]built three transfer learning models ResNet50, InceptionV3 and Xception. They fused the features extracted by the above three models, and then put the fused features into CNN3 (three-layer 3*3 convolution and two-layer fully connected layer) to classify malignant tumors and benign tumors. The best result obtained finally has an accuracy of 89.44 and an AUC of 0.93.

Yap et al. [12]used the transfer learning model FCN-AlexNet, patch-based LeNet and U-Net methods on two ultrasound image data sets (A and B) to train and test, and finally the transfer learning model FCN-AlexNet Achieved the highest F-measure (0.92 for data set A and 0.88 for data set B).

Hadad et al.[5]believe that cross-model transfer learning within the domain is better than cross-domain transfer learning. The data sets used by the author are mammography images (282 enhancements to 32064) and breast MR images (123 enhancements) To 19316). They first trained MG-Net from scratch using MR images, and got an accuracy of 0.67 and an AUC of 0.74. Then use the VGG-Net-based fine-tuning CNN (pre-trained on ImageNet) test, and get an accuracy of 0.90 and an AUC of 0.95. Finally, a fine-tuning CNN based on MG-

Net (pre-trained on mammary X-ray images) was tested, and an accuracy of 0.93 and an AUC of 0.97 were obtained.

Rakhlin et al. [13] proposed and compared the performance of three deep learning methods in breast ultrasound lesion detection, namely LeNet, U-Net, and transfer learning FCN-AlexNet. The study used two data sets A/B. In the end, transfer learning FCN-AlexNet obtained a true positive rate (TPF) of 0.99 on data set A and a TPF of 0.93 on data set B, which is the best overall performance.

Kassani et al. [14] proposed to use 5 deep convolutional neural network (DCNN) frameworks and their improved versions to classify hematoxylin and eosin staining (H&E) histological breast cancer images, and compared the differences. The effect of the staining normalization method. The five DCNNs are InceptionV3, InceptionResNetV2, Xception, VGG16, and VGG19. The experimental results show that the network architecture proposed using the pre-trained Xception model has an average classification accuracy of 92.50% due to other DCNN architectures.

Michał Byra et al. [15] studied the influence of ultrasound image reconstruction algorithm on the classification of breast lesions based on transfer learning. In order to minimize the degradation of classification performance caused by the reconstruction algorithm, they proposed a data expansion method and compared the three image reconstruction thresholds, which are 40dB, 50dB, and 60dB. The migration models used at the same time are InceptionV3 and VGG19 pre-trained on the ImageNet dataset. Experimental results show that the method of InceptionV3 model has better classification performance overall. The greater the difference in the threshold between the training set and the test set, the worse the classification performance. The data enhancement technology proposed in the experiment has significantly improved the performance, and achieved the best performance with an AUC value of 0.857 on the InceptionV3 model.

In order to better automate the process of obtaining a region of interest (ROI), Moi Hoon Yap et al. [16] used the image detection deep learning framework Faster-RCNN with Inception-ResNet-v2 as their deep learning network. At the same time, transfer learning and a new three-channel artificial RGB method are used to improve the overall performance. Experiments were conducted on two data sets (A, B). The experimental results show that the three-channel artificial RGB method improves the recall rate on data set A, but this method may not be suitable for images with different resolutions, and further research is needed.

YI WANG et al. [17] proposed using an improved InceptionV3 model to extract features on multiple ultrasound images and compare it with traditional machine learning feature extraction methods (principal component analysis, directional gradient histogram). The experimental results show that the improved InceptionV3 model and the multi-view strategy can achieve the best results (AUC value is 0.9468), indicating that the multi-view strategy can effectively increase the extraction ability of lesion features by simultaneously extracting features from the lateral view and the coronal view.

3 Methodologys

The classification models in this article are all built using the pytorch framework, which is a concise and convenient deep learning framework. The experiment evaluated the performance of four models in the classification of benign and malignant breast tumors on the breast ultrasound image data set. The models are ResNet18, ResNet50 and the transfer learning models ResNet18 and ResNet50 pre-trained on the ImageNet data set.

3.1 Data preprocessing

The data set used in the experiment is 131 breast ultrasound lesion images(109 benign lesions, 22 malignant lesions), of which 70% is used as the training set and 30% is used as the test set. In order to make the trained model more robust and prevent overfitting, We use the data enhancement method that comes with the pytorch framework to perform horizontal flip, random center sampling, and regularization of our images.

3.2 Unpretrained model

Both the ResNet18 and ResNet50 networks have achieved excellent performance in the ILSVRC and COCO competitions [8] We directly use these two models to train from scratch on the breast ultrasound image dataset. For the ResNet18 model, set the learning rate to 0.001 and change its fully connected layer classification number to 2 classifications. For the ResNet50 network, set the learning rate to 0.0256 and change the classification number of its fully connected layer to 2 classifications. The learning rate of the two models is reduced to one-tenth of the original every 7 epochs, and both epochs are set to 100.

3.3 Pre-trained model

As shown in Figure 1, we first pre-train the ResNet18 network on ImageNet, and save the model and the weight of each layer after training. Pytorch provides us with many models and parameters of the network pre-trained on ImageNet, which we can use directly. But in ImageNet training, it is the classification of 1000 kinds of pictures. In order to make the model more suitable for our data set, we freeze the first two layers of the ResNet18 network. The first few layers of the network usually extract the general features of the picture, so the effect of directly using the first few layers as a feature extractor is significantly improved. The subsequent convolutional layer is initialized with the weights obtained by pre-training to train our data set, and finally uses the stochastic gradient descent (SGD) algorithm to update the network weights

and change the 1000 classification output of the fully connected layer to 2 classifications. Both transfer learning models use the SGD optimizer + momentum training model, the learning rate is set to 0.001 (ResNet18 only sets the fully connected layer, ResNet50 sets the non-freezing layer) and the learning rate decreases by one-tenth every 7 epochs, epoch setting is 100.

Fig. 1 The resnet18 network pre-trained on ImageNet is transferred to breast ultrasound images for training. In the figure, each BasicBlock is composed of four 3*3 convolutional layers.

For the ResNet50 network, the processing is basically the same as that of ResNet18. The difference is that ResNet50 is much deeper than ResNet18, and the number of data sets used in the experiment is too small, so if only the first two layers are frozen and then training on the ultrasound image data set will lead to overfitting. After the experiment, we found that all layers except the last Bottleneck and the fully connected layer can be frozen to achieve the best effect, and we need to add a layer of dropout to the final fully connected layer to further prevent overfitting.

4 Experiment

4.1 Data set

The breast ultrasound lesion data set comes from the UDIAT Diagnostic Center and is authorized for use by Dr. Moi Hoon Yap[12]. This data set consists of 131 breast ultrasound images from different women, of which 109 are benign lesions and 22 are malignant lesions. The average image size is 501×440 .

pixels. Each image shows one or more lesions. Malignant tumors include invasive ductal carcinoma, ductal carcinoma in situ, invasive lobular carcinoma, etc. Among benign tumors, there are fibrous tumors and some other benign tumors. Lesions. These lesions are delineated by experienced radiologists for research use. Figure 2 shows some examples of breast ultrasound images in the dataset.

Fig. 2 In the breast ultrasound lesion samples used in this article, (a) and (b) are benign tumors (c) and (d) are malignant tumors

4.2 Evaluation metrics

In this experiment, we use accuracy, precision, recall, F1-score and roc curve (auc value) to evaluate the model's performance in tumor classification on breast ultrasound images. The composition parameters of the evaluation index are shown in Table 1, where P is the predicted result and L is the label.

Table 1 Confusion matrix

	P(Benign)	P(Malignant)
L(Benign)	TP	FN
L(Malignant)	FP	TN

The accuracy refers to the percentage of correct prediction results in the total sample. The expression is:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

The precision refers to the probability that all samples predicted to be benign are actually benign, and the expression is:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

The recall rate refers to the probability of being predicted to be benign in a sample that is actually benign, and the expression is:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

The F1 score considers both the precision rate and the recall rate, so that both reach the highest at the same time and achieve a balance. The expression is:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

In the ROC curve, each point on the abscissa corresponds to the false positive rate (FPR), and each point on the ordinate corresponds to the true rate (TPR). AUC represents the area under the ROC curve and the x and y axes. It is usually used to measure the generalization ability of the model. It is a common indicator to measure the pros and cons of a two-class model.

$$FPR = \frac{FP}{FP + TN} \quad TPR = \frac{TP}{TP + FN} \quad (5)$$

4.3 Result

The experiment was carried out on Windows 10 operating system with 16G memory and NVIDIA RTX2060GPU. The four models were tested for the recognition performance of benign and malignant breast ultrasound images. Each model was tested 30 times, and then the average of the experimental results was taken as the final result. It takes 3 minutes and 5 seconds to train 100 epochs for the pre-trained ResNet50 model, and 2 minutes 10 seconds to train 100 epochs for the pre-trained ResNet18 model.

Table 2 shows the performance of the ResNet18 and ResNet50 networks without pre-training in the classification of malignant/benign tumors on the breast ultrasound image dataset. It can be seen from the table that the unpretrained ResNet model performs poorly on the small sample data set. The ResNet50 model is better than the ResNet18 model in terms of accuracy, precision, recall rate and F1 value, reaching an accuracy of 82.93% and an F1-score of 0.91.

Table 2 Performance comparison of two ResNet models on breast ultrasound image dataset

Model	Accuracy	Precision	Recall	F1-score
ResNet18	77.34%	82.53%	92.21%	0.87
ResNet50	82.93%	82.90%	99.85%	0.91

Table 3 shows the performance of the transfer learning models ResNet18 (Pre-ResNet18) and ResNet50 (Pre-ResNet50) pre-trained on the ImageNet dataset in the classification of malignant/benign tumors on the breast ultrasound image dataset. The best performing numbers are shown in bold in the table. From the table, we can see that the performance of the pre-trained ResNet model is significantly better than that of the model trained from scratch without pre-training. In terms of classification accuracy and recall rate, the effect of Pre-ResNet18 model (93.90% accuracy, 97.54% recall rate) is better than Pre-ResNet50 (91.36% accuracy, 93.79% recall rate). The accuracy of Pre-ResNet50 is the best, reaching 93.67%. Pre-ResNet18 achieved the best results on the breast ultrasound image data set in this experiment, and F1-score reached 0.94. Since the data set used in the experiment is small, the shallower Pre-ResNet18 is less prone to overfitting than Pre-ResNet50, and the training effect is better. In the experiment of Pre-ResNet50 model, if too many layers are frozen, although over-fitting is effectively suppressed, it will also lead to a local optimum with lower accuracy. For the Pre-ResNet18 model, it is only necessary to freeze the convolutional layer from which the general features are extracted, so that the specific features of the breast ultrasound image can be better trained to achieve better results.

Table 3 Performance comparison of two deep transfer learning models on breast ultrasound image dataset

Model	Accuracy	Precision	Recall	F1-score
Pre-ResNet18	93.90%	91.36%	97.54%	0.94
Pre-ResNet50	92.68%	93.67%	93.79%	0.94

Figure 3 shows the ROC curves of the four models on the breast ultrasound image data set. It can be seen that the pre-trained migration learning model is significantly better than the model that is trained from scratch without pre-training. Among the two transfer learning models, the performance of the Pre-ResNet18 model (AUC of 0.944) is better than that of the Pre-ResNet50 model (AUC of 0.904), which also shows that the Pre-ResNet18 model has the most generalization on this data set. The unpretrained ResNet model has poor generalization and is prone to overfitting. The AUC value of the ResNet18 model is 0.658 and the AUC value of the ResNet50 model is 0.662, but it is worth noting that the ResNet18 model is not always better than ResNet50 model. This may be caused by too little training data. This shows that the

transfer learning model has a huge advantage over the general CNN model in small sample training.

Fig. 3 ROC curves of four classification models on breast ultrasound images

5 Conclusions

In this article, we compare the classification performance of 4 different classification models on benign and malignant breast tumors. Among them, we focused on the pre-trained migration learning models ResNet18 and ResNet50, and adjusted their frozen layers and optimization parameters. Among these four models, the pre-trained ResNet18 achieved the best results on our breast ultrasound image data set, with an accuracy of 93.90%, an accuracy of 91.36%, and a recall rate of 97.54%, F1- The score is 0.94. Therefore, when the experimental data set is small and it is difficult to obtain annotations, the transfer learning method is used to make the model learn common features on a large data set first, and then use the pre-trained weights to initialize the model and train our own small sample data set , Can greatly improve the accuracy of the model.

Inspired by Omer Hadad et al.'s research[5] on intra-domain model transfer and transfer between different domains, in future research, we will focus on model transfer in the medical image field. Use related methods of deep network adaptation, such as DDC[18], DAN[19] etc. Adding an adaptive layer on the basis of fine-tuning makes the feature distribution between the target domain and the source domain closer, so as to achieve better classification accuracy.

References

1. ZHENG Y, WU X, ZHANG M. The epidemic and characteristic of female breast cancer in china[J]. *China Oncology*(in Chinese), 2014, 23(8):561-569.
2. MUSTRA M, GRGIC M, RANGAYYAN R M. Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms[J]. *Medical & biological engineering & computing*, 2016, 54(7):1003-1024.
3. WU S, CHEN Y, et al. progress of ultrasonography in diagnosed of breast cancer(in chinese)[D]. [S.l.: s.n.], 2006.
4. ELTER M, SCHULZ-WENDTLAND R, WITTENBERG T. The prediction of breast cancer biopsy outcomes using two cad approaches that both emphasize an intelligible decision process[J]. *Medical physics*, 2007, 34(11):4164-4172.
5. HADAD O, BAKALO R, BEN-ARI R, et al. Classification of breast lesions using cross-modal deep learning[C]//2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). [S.l.]: IEEE, 2017: 109-112.
6. ZHUANG F, QI Z, DUAN K, et al. A comprehensive survey on transfer learning[J]. *Proceedings of the IEEE*, 2020.
7. PAN S J, YANG Q. A survey on transfer learning[J]. *IEEE Transactions on knowledge and data engineering*, 2009, 22(10):1345-1359.
8. HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016: 770-778.
9. HAN S, KANG H K, JEONG J Y, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images[J]. *Physics in Medicine & Biology*, 2017, 62(19):7714.
10. ZHANG H, HAN L, CHEN K, et al. Diagnostic efficiency of the breast ultrasound computer-aided prediction model based on convolutional neural network in breast cancer [J]. *Journal of Digital Imaging*, 2020, 33:1218-1223.
11. XIAO T, LIU L, LI K, et al. Comparison of transferred deep neural networks in ultrasonic breast masses discrimination[J]. *BioMed research international*, 2018, 2018.
12. YAP M H, PONS G, MARTÍ J, et al. Automated breast ultrasound lesions detection using convolutional neural networks[J]. *IEEE journal of biomedical and health informatics*, 2017, 22(4):1218-1226.
13. RAKHLIN A, SHVETS A, IGLOVNIKOV V, et al. Deep convolutional neural networks for breast cancer histology image analysis[C]//*International Conference Image Analysis and Recognition*. [S.l.]: Springer, 2018: 737-744.
14. KASSANI S H, KASSANI P H, WESOŁOWSKI M J, et al. Breast cancer diagnosis with transfer learning and global pooling[J]. *arXiv preprint arXiv:1909.11839*, 2019.
15. BYRA M, SZNAJDER T, KORZINEK D, et al. Impact of ultrasound image reconstruction method on breast lesion classification with deep learning[C]//*Iberian Conference on Pattern Recognition and Image Analysis*. [S.l.]: Springer, 2019: 41-52.
16. YAP M H, GOYAL M, OSMAN F, et al. Breast ultrasound region of interest detection and lesion localisation[J]. *Artificial Intelligence in Medicine*, 2020, 107:101880.
17. WANG Y, CHOI E J, CHOI Y, et al. Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning[J]. *Ultrasound in medicine & biology*, 2020, 46(5):1119-1132.
18. TZENG E, HOFFMAN J, ZHANG N, et al. Deep domain confusion: Maximizing for domain invariance[J]. *arXiv preprint arXiv:1412.3474*, 2014.
19. LONG M, CAO Y, WANG J, et al. Learning transferable features with deep adaptation networks[C]//*International conference on machine learning*. [S.l.]: PMLR, 2015: 97-105.

Figures

Figure 1

The resnet18 network pre-trained on ImageNet is transferred to breast ultrasound images for training. In the figure, each BasicBlock is composed of four 3×3 convolutional layers

(a)

(b)

(c)

(d)

Figure 2

In the breast ultrasound lesion samples used in this article, (a) and (b) are benign tumors (c) and (d) are malignant tumors

Figure 3

ROC curves of four classification models on breast ultrasound images