Medical & Biological Engineering & Computing
https://doi.org/10.1007/s11517-019-02111-w

ORIGINAL ARTICLE

A deep convolutional neural network architecture for interstitial lung disease pattern classification

Sheng Huang[1] · Feifei Lee[1] · Ran Miao[1] · Qin Si[1] · Chaowen Lu[1] · Qiu Chen[2]

Received: 28 June 2019 / Accepted: 21 December 2019

Abstract
Interstitial lung disease (ILD) refers to a group of various abnormal inflammations of lung tissues and early diagnosis of these disease patterns is crucial for the treatment. Yet it is difficult to make an accurate diagnosis due to the similarity among the clinical manifestations of these diseases. In order to assist the radiologists, computer-aided diagnosis systems have been developed. Besides, the potential of deep convolutional neural networks (CNNs) is also expected to exert on the medical image analysis in recent years. In this paper, we design a new deep convolutional neural network (CNN) architecture to achieve the classification task of ILD patterns. Furthermore, we also propose a novel two-stage transfer learning (TSTL) method to deal with the problem of the lack of training data, which leverages the knowledge learned from sufficient textural source data and auxiliary unlabeled lung CT data to the target domain. We adopt the unsupervised manner to learn the unlabeled data, by which the objective function composed of the prediction confidence and mutual information are optimized. The experimental results show that our proposed CNN architecture achieves desirable performance and outperforms most of the state-of-the-art ones. The comparative analysis demonstrates the promising feasibility and advantages of the proposed two-stage transfer learning strategy as well as the potential of the knowledge learning from lung CT data.

Keywords Interstitial lung diseases (ILDs) · Convolutional neural networks (CNNs) · Deep convolutional autoencoder · Transfer learning

1 Introduction

Interstitial lung disease (ILD) includes a group of more than 200 chronic inflammation of lung tissues, which can severely affect the pulmonary interstitium, and may even impair the breathing ability of the patient [32]. Therefore, early diagnosis of these diseases is essential for making treatment decisions. Although ILDs are histologically heterogeneous, most of them have similar clinical pathological characteristics. There appears low inter-class distinctions as well as high intra-class variance among some patterns, and even exists different combinations of pathological patterns on HRCT images [9]. The complexity of diagnosis is challenging even for many experienced experts in this field and may leads to as high as 50% ambiguity in the radiological assessment [3, 31]. It is also a time-consuming and laborious work for radiologists to scrutinize a large amount of cases. In order to assist the radiologists, analysis methods that combine digital image processing and pattern recognition techniques, incorporated into computer-aided diagnosis (CAD) system, have been studied extensively [38]. Effective characterization and identification algorithm for different tissues are the most important parts of a CAD system. Consequently, a great variety of conventional image descriptors and classifiers have been elaborately used for analysis.

More recently, deep learning methods especially convolutional neural networks (CNNs) have been shown to be extremely effective for many computer vision tasks. CNNs

Sheng Huang and Feifei Lee contributed equally to this work.

✉ Feifei Lee
feifeilee@ieee.org

✉ Qiu Chen
q.chen@ieee.org

[1] Department of Control Science and Engineering, University of Shanghai for Science and Technology, Shanghai, China

[2] Graduate School of Engineering, Kogakuin University, Shinjuku, Japan

Published online: 22 January 2020

are able to automatically learn distinctive features and show better robustness, which also enable them to achieve impressive performance in medical image analysis, such as detection [40], segmentation [5] and classification [25]. Studies [1, 21] indicate that deep learning techniques have rapidly attracted much attention in the field of medical application, and the potential of CNN-based method in this field is constantly being explored. Generally, deeper networks, such as AlexNet [17],VGG [30] and GoogLeNet [34], can yield more competitive results. Accordingly, deeper architecture contains more parameters and requires massive amounts of labeled data for training.

Unfortunately, collecting a sufficient number of high-quality annotated training samples in the medical field is not only tedious and time consuming, but also demanding of expertise and proficient skills. Hence it is hard to create a large dataset for training a complicated CNN without overfitting. Transfer learning [24] is a common way to tackle this kind of "data starvation" problem, which can compensate for the lack of data in a target domain by inheriting or preserving the knowledge learned from a data-rich source domain. Some studies [21, 35] indicate that employing the pre-trained CNNs as feature extractors or fine-tuning of pre-trained CNNs can achieve better performance for a variety of medical image analysis tasks.

Nevertheless, the effectiveness of transfer learning might be affected by the similarities between the source and target domains [4]. In most transfer learning approaches for image recognition tasks, ImageNet [8] is often employed as the source dataset due to its abundant categories and significant quantity of images. However, there exists certain distinct characteristics between nature images and medical imagery data, which may limits the fine-tuning process to accomplish effective adaptation for the target task. Some studies like [22, 36, 43] try to minimize the distance among domains when perform the transfer learning. However, these methods are still limited if the gap between the source and target domain is too large. Christodoulidis et al. [7] resort to transfer knowledge from multiple related source domain datasets, but their method using ensemble and model compression is complicated.

In this paper, we design a new deep CNN model exclusively for the classification task of ILD patterns and build a deep convolutional autoencoder (DCAE). In order to improve the performance of the network, we propose a novel two-stage transfer leaning (TSTL) approach that transfers knowledge from a general texture source dataset and an intermediate domain dataset, which general framework is shown in Fig. 1. In our approach, we apply a large number of unlabeled lung pattern patches as the intermediate domain data, which are easily extracted from the unannotated areas of the lung CT slices. These data possess unique and highly similar texture feature as target ILD patterns. The capability of network will be strengthened if rich representations can be learned from these data. Hence, instead of directly transferring to the target domain, the pre-trained CNN or DCAE in source domain is firstly fine-tuned using intermediate domain dataset. The contributions of this paper are concluded as follows:

1. A new customized CNN architecture for ILD pattern classification is designed in our work. Experimental results show that this lightweight CNN architecture outperforms most of other CNNs in this classification task. Furthermore, a deep convolutional autoencoder is built to perform the proposed transfer learning strategy.
2. To our knowledge, the unlabeled lung pattern patches extracted from the unannotated area of the lung field are investigated for the first time.
3. For the first time, the unlabeled data are adopted as the intermediate domain in the two-stage transfer learning approach. Furthermore we propose a new method to train the network using auxiliary unlabeled data in unsupervised manner.
4. We demonstrate that the proposed method outperforms the state-of-the-art ones in ILD pattern classification with scarce data, which is the bottleneck of this issue.

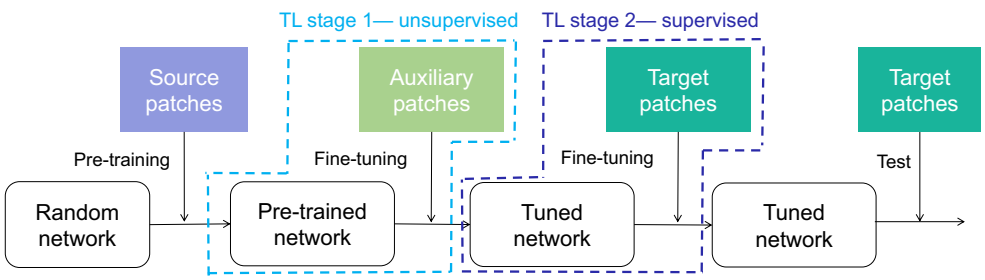


Fig. 1 The general framework of the proposed two-stage transfer learning method. The network is sequentially trained on source dataset, auxiliary intermediate dataset and target dataset. Here the "auxiliary patches" are unannotated lung patches to train the specific lung texture patterns and that "target patches" are annotated by ILD typologies or as healthy lung tissue

2 Related work

2.1 Pattern classification with CNNs

Ever since the detection of interstitial lung disease(ILD) is concerned, a number of studies and great progress have been made based on the digital image processing and pattern recognition techniques. The typical procedure can be described as extracting the feature of the input data and then feeding to a classifier for categorization. Recently deep learning techniques especially CNN has been proved very successful in computer vision field. Compared with traditional handcrafted description methods, CNNs are capable of learning more intrinsic and informative features effectively from training data. And CNNs can be generalized to solve different kinds of problems using similar designs, which makes it popular in many applications.

For the issue of ILD pattern classification, the patch-wise classification method has been demonstrated by a neural network which composes of one convolutional layer and three fully connected layers [20]. Similarly, Tarando et al. [37] adopt two cascaded layers of convolutional filters and two fully connected layers. However, these two shallow architectures do not have enough discriminative power. Gao et al. [10] investigate the slice-wise ILD pattern classification problem using an architecture modified based on the AlexNet [17]. This network is deeper and able to produce higher accuracy. Since some fairly deep networks like VGGNet [30] and GoogLeNet [34] exhibit impressive performance in many computer vision applications, the discriminative power of them is also investigated for lung tissue classification in [29]. Although the architecture of these networks are not optimal for solving the lung pattern classification problem, relatively good results can be achieved by fine-tuning their pre-trained models. The more advanced architecture of DenseNet [15] is also investigated for the classification task ILD patterns by [11] where authors propose the small kernel DenseNet (SK-DenseNet). The architecture of proposed SK-DenseNet contains two dense blocks and the size of all convolution kernels is set as 2×2 to extract high level and small pathological features of ILD pattern.

To achieve the best trade-off between discriminative power and avoid over-fitting, some networks suitable for ILD pattern textural feature extraction are investigated. In [2], authors propose a customized CNN architecture to classify HRCT image patches of ILD which outperforms shallower networks and some deeper networks. O'Neil et al. [23] modify two network base on [2, 20] and prove the well-designed deep learning architecture with sufficient capacity can outperform traditional model-based approaches. However, these networks present limited performance since they are shallow and difficult to train to a certain degree. In this paper, we propose a new CNN architecture to address the ILD pattern classification problem.

2.2 Transfer learning

Transfer learning is generally defined as a learning system that applies knowledge learned from one task, to another related task which shares similar characteristics. The research of transfer learning is inspired by the ability of human that utilizing previous learned experience to find effective solutions to solve new problems. A survey on transfer learning [24] indicates that a lot of attention has been paid on developing varieties of methods to transfer knowledge across domains.

Recently, deep CNNs have shown remarkable abilities in many different tasks and the knowledge learned from training samples is incorporated in the weights. Some researches [28, 42] expect to reutilize network's feature extraction capability to get "off-the-shelf" features which can be used for classification. Other researches manage to realize transfer learning by weight transferring ways that freeze or fine-tune the parameters of the network. Moreover, the specificity of different layers weights has been analyzed in [41]. Generally, the earlier layers near the input are generic while the last layers of the network are more task specific. Hence, in the transfer learning process, the parameters of the first layers can be fixed, the latter layers can be fine-tuned, and for the non-transferred weights of the network are usually randomly initialized.

Transfer learning can not only cover the shortage of training data but also promote the result of target tasks using the experience gained from the source domain. The merit of transfer learning also make it a popular technique in the medical field where the medical images for training are usually scarce. Many scholars try to employ well-known CNNs pre-trained on ImageNet for medical image identification and classification tasks with ultrasound, CT and X-ray imaging [6, 39, 40]. Although these studies show the potential of knowledge transfer from natural image to the medical imaging domain, the factors that influence the transferability of knowledge, i.e., domain dissimilarity, still exist. Recently, some studies have explored different methods to mitigate this effect. Lu et al. [22] propose a novel transfer learning approach to learn useful knowledge from source data by adding a constraint between source and target classifier predictions. A common method that refers to "multi-stage transfer learning" is proposed in [27, 33] where authors utilize a intermediate domain to bridge source and target domain. In [7], the authors try to exploit relevant knowledge from six texture benchmark databases and transfer to the target classification task of ILD patterns.

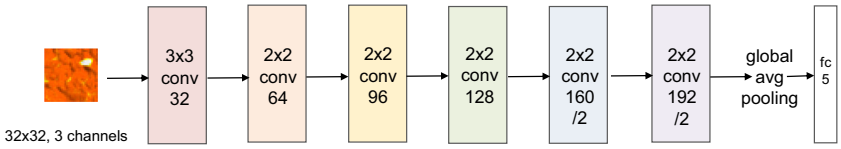


Fig. 2 The architecture of the proposed CNN for lung pattern classification

These studies inspire us to make use of the similar source dataset for training and encourage the network to learn a characteristic similar to the target domain before fine-tuning.

3 Methods

3.1 Convolutional neural network architectures

Based on the network design principles analyzed in [2], we design a new network for the classification of ILD patterns. The architecture consists of six convolutional layers, followed by a fully connected layer. A simplified illustration of the network architectures is shown in Fig. 2. The input of the network is 32 × 32 pixels which equals to the extracted image patch. To facilitate comparison with other networks and [10] has demonstrated that using different CT attenuation channels improves classification results over the usage of a single CT windowing channel, so the input channel is set as 3. Simonyan et al. [30] prove that the small kernel can improve the performance of ConvNet. Therefore the size of the kernels in each layer are set to 2×2 except the first convolutional layer which mainly learns low-level feature. Each convolutional layer contains increasing number of kernels from 32 up to 192. The kernel strides of the first four layers and the last two layers are set to 1 and 2, respectively. Consequently, the output of the last convolutional layer is 8 and thus reducing the computation in the followed global average pooling operation. The last fully connected layer serves as a 5-way softmax classifier which computes the probabilistic interpretation of each class. Different from [2], the extra dense layers are canceled, instead, batch normalization are added right after each convolutional layer to accelerate convergence. Besides, we use ReLu instead of LeakyReLu as the activation function.

Based on the proposed network we build a deep convolutional autoencoder (DCAE) by cascading its mirror deconvolutional counterpart after last convolutional layer, which architecture is shown in Fig. 3. Generally, the framework of an autoencoder [13] includes encoding and decoding processes. In our paper, the proposed CNN serves as encoder for learning the hidden representation and the deconvolutional counterpart as decoder for attempting to reconstruct the input from the hidden representations.

For training, the Adam optimizer [16] is used to minimize the categorical cross entropy, and the learning rate starts from 0.0008 and decays 10% every 20 epochs. The weight updates are performed in mini-batches with the size of 128 and the training ends when there are no significant improvements of the performance on the validation or test set.

3.2 Two-stage transfer learning

The common transfer learning approach is single-stage transfer, i.e., pre-training on source domain and then directly fine-tuning on the target dataset. Beyond performing the conventional transfer learning method, we try to explore the effectiveness of multi-stage transfer on the classification task of ILD patterns. As mentioned before, the intermediate auxiliary data have the potential for transfer learning. Hence, instead of transferring directly to the target domain, the pre-trained model is first fine-tuned in the intermediate domain. The whole framework of the proposed two-stage transfer learning (TSTL) is shown in Fig. 1.

The key element of the two-stage transfer method is learning knowledge from the intermediate domain dataset. Since the intermediate auxiliary dataset has no label information, we consider training the network in unsupervised manner. In this study, we investigate the performance of two-stage transfer learning using the proposed CNN and DCAE, respectively.

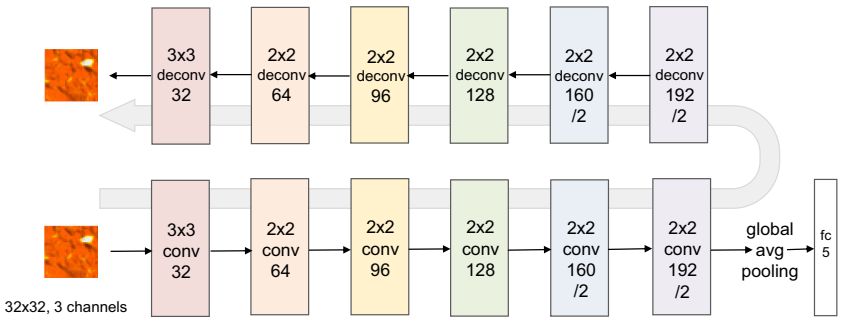


Fig. 3 The simplified illustration of the deep convolutional autoencoder, this network consists of encoding process, decoding process and categorizing process

3.2.1 Pre-training on source domain

For the supervised pre-training of CNN, the optimization objective is categorical cross-entropy loss function computed by softmax classifier:

$$\text{Loss}_{\text{cls}} = -\sum_{i=1}^{c} 1\{y = i\} \log \frac{e^{z_i}}{\sum_{l=1}^{c} e^{z_l}} \tag{1}$$

where c denotes the number of classes, z_i denotes the output of each unit in the last layer and y corresponds to the label information of each input instance.

As for DCAE, the encoding process $f_e(\cdot)$ can map input x to a latent representation h. Then h is mapped back to a reconstruction $\hat{x}$ through decoding function. The reconstruction loss equals to the mean squared error between the pixel values of the original image and the reconstructed image:

$$\text{Loss}_{\text{recon}} = \|\hat{x}_i - x_i\|_2^2 \tag{2}$$

For the supervised pre-training of DCAE, the feature extraction is realized by minimizing the cross-entropy loss and reconstruction error simultaneously:

$$J = \text{Loss}_{\text{cls}} + \text{Loss}_{\text{recon}} \tag{3}$$

During the pre-training progress, we expect all kernels are well-trained for extracting visual features, such as edge structures, from input images of source data.

3.2.2 Unsupervised training on intermediate domain

Since the intermediate domain has no label information for supervised training, we introduce the entropy function as the confidence on the predictions:

$$\text{Conf}_{\text{predi}} = \sum_{k=1}^{|Y|} p_i^k \log \frac{1}{p_i^k} \tag{4}$$

Here, we assume the prediction of instance x_i by the output of classifier is p_i and $|Y|$ denotes the output nodes of classifier. We consider the predictions as the side information, and use the confidence on the predictions to guide unsupervised learning, so we expect this value can be maximized.

Additionally, we also incorporate a novel deep representation learning method called Deep InfoMax (DIM), proposed by Devon et al. [14], into the objective function. DIM can simultaneously estimate and maximize the mutual information (MI) between input data and learned high-level representations, thus empowering the network with the ability to extract informative features from training data. The mutual information, $\hat{I}(x; f_e(x))$, can be estimated by a Jensen-Shannon-based MI estimator:

$$\begin{aligned}\hat{I}(x; f_e(x)) = {} & \mathbb{E}_{\mathbb{P}}[-sp(-T(x; f_e(x)))] \\ & -\mathbb{E}_{\mathbb{P}\times\tilde{\mathbb{P}}}[sp(-T(x'; f_e(x)))]\end{aligned} \tag{5}$$

Where x denotes the input data, and x' is a fake input sampled from $\tilde{\mathbb{P}} = \mathbb{P}$. The whole encoder $f_e(x)$ generates the high-level feature vector. T denotes a discriminator that computes the score of the concatenated feature map / feature vector pair. And $sp(z) = \log(1 + e^z)$ refers to the softplus function.

In our approach, we implement the training method follows as [14], but only MI maximizations of the global and local features are added into the objective function. This function is concluded as:

$$\begin{aligned}\text{MI} &= \alpha \text{MI}_{\text{Global}} + \beta \text{MI}_{\text{Local}} \\ &= \alpha \hat{I}(f_{e0}(x); f_e(x)) + \frac{\beta}{M^2} \sum_{i=1}^{M^2} \hat{I}(f_{e0}^i(x); f_e(x))\end{aligned} \tag{6}$$

where $f_{e0}(\cdot)$ is a small part of the encoder that encodes the input to low-level feature maps. And M equals to the size of the low-level feature map. The trade-off hyperparameters α and β are used to control the effect of each term and defaulted as $\alpha = 0.5$ and $\beta = 1$.

In this study, the $f_{e0}(\cdot)$ is the first three convolutional layers of the proposed network. The low-level feature maps come from the training data or another unrelated image. For MI maximization of global features, the low-level feature maps are first passed through two 2×2 convolutional layers to reduce the dimension. These two convolutional layers have 64 and 32 kernels, respectively. The output feature maps will be flattened and concatenated with the high-level feature vector then passed to the discriminator and computed by (5). For MI maximization of local features, before concatenating with low-level feature map,the high-level feature vector is first replicated to the size of $M \times M$, thus the local feature discriminator can score the feature pair at every location. Note that the global feature discriminator consists of two 512-unit fully connected layers and 1-unit output layer. The local feature discriminator is a 1×1 convolutional network with two 512-unit hidden layers. The simplified illustration of the training framework is depicted in Fig. 4.

When the unsupervised training process is performed by CNN, the objective includes these two surrogate functions:

$$J_{\text{ft}} = -\text{Conf}_{\text{predi}} - \text{MI} \tag{7}$$

Note that the training is conducted based on the pre-trained CNN; hence, it is actually a fine-tuning process. We fine-tune all convolutional layers but the fully connected layer is randomly initialized and altered to 5-unit.

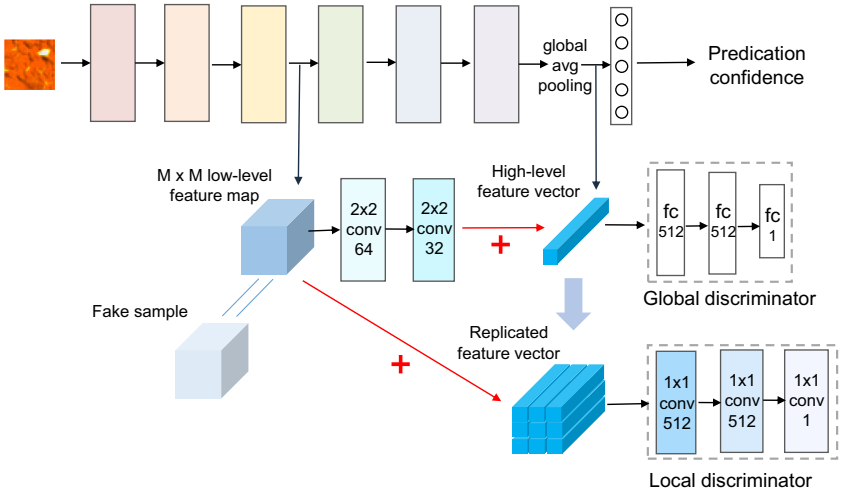


Fig. 4 The overall framework of intermediate unsupervised training process performed by proposed CNN

When we fine-tune the pre-trained DCAE on the intermediate dataset, the reconstruction error is also included in the optimization objective function:

$$J_{\text{ft}} = \text{Loss}_{\text{recon}} - \text{Conf}_{\text{predi}} - \text{MI} \tag{8}$$

It is noteworthy that the DCAE has learned how to extract the feature of input source data supervised by the label information in the pre-training process. So it is feasible to fine-tune in unsupervised manner by minimizing the reconstruction error and need not adopt the layer-wise fashion.

Although the intermediate auxiliary dataset has no label information, we assume that the learned knowledge by supervised pre-training can guide the unsupervised learning process to discover suitable features. After fine-tuning on the intermediate dataset, the networks gain the ability to excavate the textural features of lung tissue and thus improving the transferability of weights to the target task.

3.2.3 Fine-tuning on the target domain

The TSTL method is completed when the pre-trained model is fine-tuned on the target domain dataset. Since the target training data contain label information, the fine-tuning process is similar to the pre-training on the source dataset. CNN is fine-tuned by minimizing the cross-entropy loss function, i.e., (1), and DCAE is fine-tuned by simultaneously minimizing the cross-entropy loss and reconstruction loss, i.e., (3). Note that the node of last fully connected layer has been altered to 5 in the unsupervised training process, so this layer needs not to be freshly trained when fine-tuning.

4 Results

4.1 Experimental setup

4.1.1 Databases

Source domain dataset A publicly available texture classification benchmark Kylberg Texture Dataset (KTD) [18] is employed to serve as source domain. This dataset includes 28 texture categories and each category corresponds to a specific texture (e.g., cushion, canvas, wall, stone, etc.). Considering the limited number of samples, we make the artificially data augmentation, such as flip and rotation. Finally, we got total 17,920 image patches used for training. Note that all instances are resized to 32 × 32 thus matching to the input size of latter proposed network.

Target domain dataset We utilize the publicly available HUG database [9], which is collected at the University Hospitals of Geneva from 128 patients undergoing high-resolution thorax CT. One hundred eight annotated HRCT image series are applied and each case has an average of 25 slices with 512 × 512 pixels. The slice thickness is 1–2 mm, and the average pixel spacing is 0.68 mm. This database is also annotated manually by experienced radiologists using polygon to delineate the pathologic patterns. Seventeen different lung patterns are provided and 1946 ROIs are delineated, along with clinical parameters from patients with histologically proven diagnoses of ILDs. Among all patterns, five most relevant ILD patterns are study in this paper, namely healthy (NM), ground glass (GG), emphysema (EM), micronodules (MN), and fibrosis (FB).

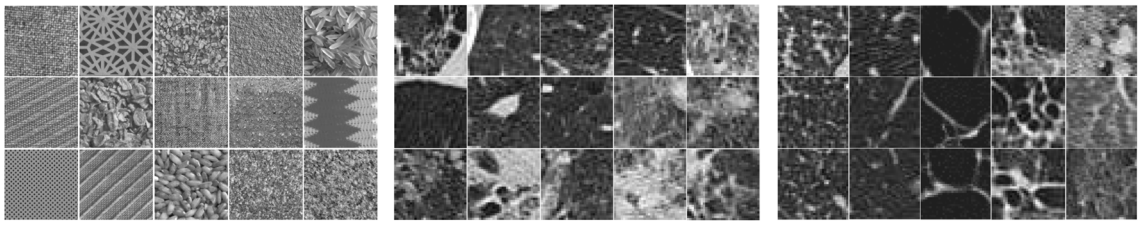


Fig. 5 Typical samples from each dataset. **a** Source data. **b** Intermediate data. **c** Target data, each column from left to right shows different ILD patterns: MN, NM, EM, FB, GG. All samples are the same size of 32 × 32

In order to normalize the scans with different pixel spacing, all axial slices are rescaled to a specific value (i.e., 1 mm). And in order to fit the proposed network we also generate 3 channels by considering 3 different HU windows according to [10]. After these serial preprocessing steps are applied, 30% overlap (for FB, GG, and EM) or non-overlapping (for NM and MN) patches of 32 × 32 mm are extracted from each ROI. Only image patches with at least 80% of its pixels falling inside of the annotated polygonal regions are used and we select approximately 1020 patches for each class.

Intermediate domain dataset We notice the annotated area of each pathologic pattern fills only a small part of whole lung field and there are large areas with similar texture as investigated ILD patterns. Figure 6 shows an example. The data extracted from these areas can be used to realize better transfer learning. For the auxiliary intermediate domain data, we extract non-overlapping patches of 32 × 32 mm from non-labeled areas in different slices. Finally, we get total 18,706 patches. Figure 5 shows some samples of these three datasets. We find that many patches in intermediate domain dataset share highly similarity with target domain which motivates us to utilize these data for transfer learning.

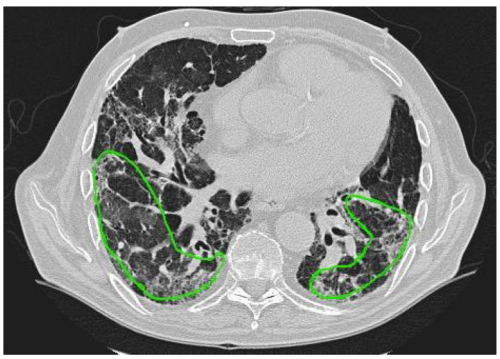


Fig. 6 Example of extracting image patches from a CT slice. The green polygon denotes the annotated pathological area, where the target domain data are extracted. Other areas in the lung field, from which the intermediate domain data are extracted, show particular textural feature

Dataset split scheme In order to fully evaluate the proposed method, we conduct the experiments on different target datasets which formed by 3 different split schemes. The differences include training sample quantity and selection manner. The details of the resulting datasets are summarized in Table 1. To be more specific, for scheme A, We divide all patches into five folds with disjoint patient subsets and conduct the fivefold validation. For scheme B, we adopt special case selection manner, i.e., some patches extracted from a same case or the same polygon ROIs are all used for training. Thus, we can investigate the generalization of the network. For scheme C, 750 patches were randomly selected for the test set and validation set, respectively. The remained patches are under data augmentation by label-preserving transformations, such as flip and rotation.

4.1.2 Evaluation

We use average F_1-score as the evaluation metric, which is a harmonic mean between recall and precision and calculated as follows:

$$F_{\text{avg}} = \frac{1}{5}\sum_{c=1}^{5} 2 \times \left(\frac{\text{recall}_c \times \text{precision}_c}{\text{recall}_c + \text{precision}_c}\right) \tag{9}$$

where recall_c denotes the fraction of samples correctively classified as c over the total test number of instances of this class, and precision is the percentage of the instances correctly classified as over the all number classified as c.

The designed network is somehow unstable which may result in the deviations of the results. In order to minimize this effect, we repeat each experiment at least ten times, then calculate the mean values.

Table 1 Details of three datasets collected by different schemes

Scheme	Training sample	Test sample	Split manner
A	4132	1031	Fivefold
B	2560	2479	Case selection
C	11,289	750	Random

Table 2 Classification performance comparison for test data

Network	Method	A	B	C
CNN	Without transfer	0.9674	0.9532	0.9663
	Single-stage transfer	0.9714	0.9598	0.9717
	Two-stage transfer	0.9791	0.9641	0.9768
DCAE	Without transfer	0.9691	0.9542	0.9677
	Single-stage transfer	0.9748	0.9582	0.9712
	Two-stage transfer	0.9810	0.9654	0.9786

4.1.3 Implementation details

All CNN training are implemented with the pytorch [26] deep learning framework and coded in Python. All experiments are performed on a Linux machine with CPU Intel Xeon(R) E5-2630 v3 @ 2.40 GHz, GPU NVIDIA Tesla K40C, and 12 GB of RAM.

4.2 Classification performances

To verify the advantage of the proposed two-stage transfer learning strategy, we conduct two groups of contrast experiments that performed by the proposed CNN and DCAE, respectively. Each group compares the proposed method with two baseline methods. The first one trains the proposed model with the target training data from scratch without any knowledge transfer and then tests it directly on the test set. The second one is a single-stage transfer learning and used for comparison with both training from scratch and two-stage transfer.

The experimental results are presented in Table 2. The classification performances are all reported with the average F_1-score. The second column includes three different training method for CNN or DCAE. And the columns indicated by A, B, and C list the values produced by the networks performance on three different datasets we described before. It can be noticed that the performance of single-stage is better than the baseline on three datasets at various degrees. This experiment proves that transfer learning is effective at avoiding over-fitting and improving the performance of the network. Comparing to the single-stage transfer leaning, two-stage transfer clearly shows further improvement especially performed by DCAE. This indicates that our unsupervised training method can learn useful knowledge from intermediate auxiliary data and thus promoting the transfer learning. This series of experiments also demonstrate that the data extracted from the unannotated areas of the lung field are useful for ILD pattern classification.

Table 3 Confusion matrix of DCAE learning from scratch

Ground	Prediction				
truth	EM	FB	GG	NM	MN
EM	0.98	0.00	0.00	0.02	0.00
FB	0.00	0.98	0.00	0.01	0.01
GG	0.00	0.00	0.98	0.02	0.00
NM	0.01	0.00	0.00	0.96	0.03
MN	0.00	0.01	0.02	0.06	0.91

Table 4 Confusion matrix of TSTL performed by DCAE

Ground	Prediction				
truth	EM	FB	GG	NM	MN
EM	0.98	0.00	0.00	0.02	0.00
FB	0.00	0.99	0.00	0.01	0.00
GG	0.00	0.01	0.99	0.00	0.00
NM	0.01	0.00	0.01	0.97	0.01
MN	0.00	0.01	0.00	0.02	0.97

To better present the classification results and highlight the effectiveness of the proposed TSTL method, we provide the confusion matrix of DCAE learning from scratch Table 3) and two-stage transfer performed by DCAE Table 4) on dataset C. The confusion matrix indicates the classification accuracy of each considered category. By comparison, the matrix of two-stage transfer is more balanced than the one learning from scratch. Meanwhile, it is worth noting that the confusion is basically between healthy (NM) and micronodules (MN). It is expected since between two of them have similar texture feature which make it difficult for classifier to distinguish. Figure 7 shows some misclassified samples produced by the DCAE under the two-stage transfer strategy. Based on the classification performance of the network, we can further evaluate the result associated with the location of each image patch. Figure 8 illustrates a typical example of predicted output

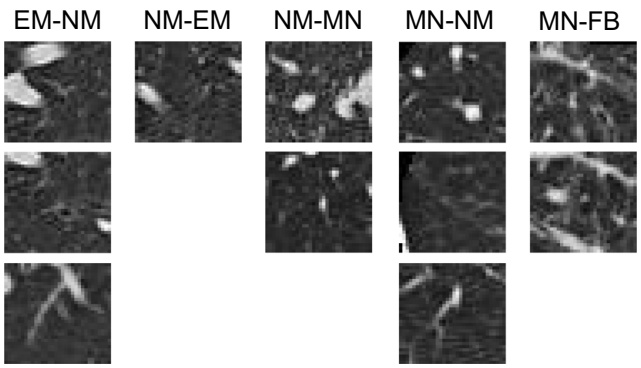


Fig. 7 Misclassified samples. Each column lists the image patches in the form of groundtruth-prediction

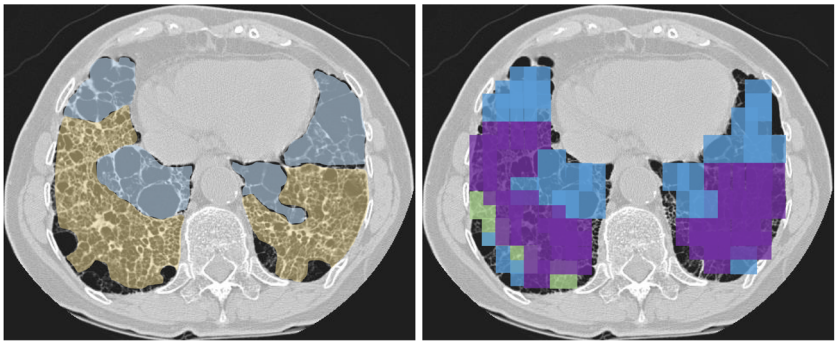


Fig. 8 Typical output example for DCAE on a whole slice from dataset C. **a** Ground truth. **b** Prediction area of ILD pathology pattern. This example shows the pattern of fibrosis (purple) and emphysema (blue)

on a whole slice. From the result of this example, we can observe that the area of ILD pathology pattern can be roughly delineated. And more precise results can be obtained by further synthesis decision-making, which would be helpful during the diagnosis.

4.3 Unsupervised representation learning

To have an insight into the effectiveness of different surrogate objective functions for unsupervised representation learning, we conduct another series of experiments that pre-training the CNN and DCAE on the intermediate domain dataset and then fine-tuning for the target task. The separated unsupervised pre-training transfer approaches are investigated by optimizing the single or combination of the different surrogate objective functions. Specifically, the prediction confidence (P) and mutual information maximization (M) are investigated. And the reconstruction error (R) is also included when the unsupervised learning performed by DCAE.

From the results shown in Table 5, we observe that the network optimized by mutual information maximization (M) achieves slight better result in F_1-score than the prediction confidence (P). And the improvement can be realized by optimizing these two surrogate objective functions simultaneously. This indicates that combining the prediction confidence and mutual information maximization for unsupervised representation learning is more effective. Comparing with the results of Table 2, we can find that no matter which surrogate objective function is applied, the unsupervised pre-training transfer approaches show better classification performance than training from scratch. This reveals that the intermediate domain dataset is helpful for pre-training and transfer to the target task even though training in unsupervised manner. On the other hand, the result of CNN with P+M optimization and DCAE with R+P+M optimization are slight worse than the corresponding TSTL strategy. It can be concluded that the pre-training on source domain data is benefit for the subsequent unsupervised representation learning.

Table 5 The comparison of different unsupervised transfer approaches

Network	Objective function	A	B	C
CNN	P	0.9677	0.9550	0.9681
	M	0.9752	0.9612	0.9733
	P+M	0.9767	0.9632	0.9753
DCAE	R	0.9688	0.9561	0.9685
	R+P	0.9696	0.9591	0.9720
	R+M	0.9771	0.9630	0.9721
	R+P+M	0.9777	0.9638	0.9764

4.4 Comparison with state-of-the-art methods

To verify the advance of proposed method, we compare with other CNNs from similar studies in recent literatures. For fair comparison, we implement the experiment of each method under the same training data and test data processed by ourselves. Table 6 provides the comparison of evaluation results.

Table 6 The comparison results of the proposed method with related literatures

Study	A	B	C
Li [20]	0.7725	0.7644	0.7987
LeNet [19]	0.8219	0.8176	0.8532
Anthimopoulos [2]	0.9465	0.9215	0.9392
Guo [11]	0.9483	0.9269	0.9554
AlexNet [17]	0.8962	0.8821	0.9226
Pretrained AlexNet [17]	0.9471	0.9337	0.9609
Pretrained VGG16 [30]	0.9603	0.9435	0.9664
Pretrained ResNet18 [12]	0.9581	0.9443	0.9651
Pretrained DenseNet121 [15]	0.9741	0.9579	0.9715
Proposed CNN	0.9674	0.9532	0.9663
TSTL-CNN	0.9791	0.9641	0.9768
TSTL-DCAE	0.9810	0.9654	0.9786

The first two rows correspond to the performance of two fairly shallow architectures. The results of them are undesirable due to their limited learning ability for highly non-linear features. The results of next two rows are more competitive since these two networks are customized for ILD pattern classification. However, they are still inferior to the proposed CNN. Furthermore, we evaluate the performances of the well-known architectures such as AlexNet, VGGNet, ResNet and DensNet, which are the state-of-the-art architectures in recent years and their pre-trained models from ImageNet [8] are fine-tuned using target training set in this series of experiments. From the results, we notice that the pre-trained AlexNet, VGG16, and ResNet18 achieve fair high performance but are still worse than the proposed CNN. Note that the results are somewhat different from the experiment presented in [2], e.g., pre-trained AlexNet achieves higher than the CNN model proposed by Anthimopoulos et al. [2]. This phenomenon may due to the different training data and test data. The pre-trained DenseNet121 outperforms all the models including the proposed CNN. However, the final best results are obtained by our proposed two-stage transfer learning method performed by CNN or DCAE. Especially, the two-stage transfer learning performed by DCAE (TSTL-DCAE) outperforms the state-of-the-art DenseNet121 by 0.006 in average F_1-score.

5 Discussion

In this study, we aim to gain a high accuracy on the classification task of ILD patterns under the condition of limited training samples. Correspondingly, we design a new CNN architecture and propose a novel two-stage transfer learning strategy.

The first finding to discuss is the performance of proposed CNN architecture. From the results of Table 6, we can find that the proposed CNN shows the highest classification accuracy. Our CNN outperforms the one proposed by Anthimopoulos et al. [2] mainly due to its depth and the batch normalization technique, which make it has enough fitting ability and easier to train. This CNN architecture is elaborately designed for the task of ILD pattern classification, so it surpasses those shallower networks [2, 19, 20] and some more complicated networks [11, 12, 17, 30] even without transfer learning. From this perspective, it is important to select or design a suitable network for a specific task when the deep learning applied to the medical imaging analysis.

The second issue to analyze is the comparison of training from scratch and transfer learning. The training data is fairly small in the series of experiments of this paper. It can be seen that, all of networks pre-trained using another domain dataset outperforms the one training from scratch. The typical example is the comparison of the AlexNet before and after pre-training as shown in the Table 6. Theoretically, AlexNet has more than 60 million parameters and requires millions of training data to prevent over-fitting. Its pre-trained model from ImageNet [8] helps it work well after fine-tuning with target training data. So when the training data is scarce, especially in the biomedical field, transfer learning is an effective way to avoid over-fitting and improve the performance of networks. Beyond single-stage transfer, we also investigate the effectiveness of unsupervised pre-training transfer and two-stage transfer. These strategies are successful since they follow the mechanism of inheriting or preserving the knowledge learned from a data-rich source domain.

It is also important to discuss where should the knowledge be transferred from. In this paper, the classification task is texture-dominated and we consider lightweight CNN architecture, so we choose the texture benchmark database KTD [18] as the source domain. There are also many similar datasets to choose from, such as the other five databases investigated in the [7]. These databases are public available and more suitable to transfer for some specific pathology analysis tasks, such as the ILD pattern classification in this paper. The main value of this paper is exploring the effectiveness and practicability of the data extracted from the unannotated areas of the lung field. These data share highly similarity with target domain and are more considerable than the texture benchmark databases. Even though these data lack of label information, effective knowledge transfer can be realized through rational unsupervised learning algorithm and transfer strategy. Our work is a good example for the future research since there are huge amount of unlabeled data can be used in the field of biomedical. The successful of unsupervised learning and transfer learning can help reduce the cost for collecting and labeling the training data, and promote the application of deep learning in the field of biomedical.

The transfer strategy we applied in this paper is two-stage transfer learning. Transferring the knowledge from unlabel intermediate data by unsupervised training is an important stage. As we know, unsupervised learning technique has not yet fulfilled the promise results comparing to supervised methods under most of situation. Hence, it is wise to combine two or more unsupervised learning method, such as entropy minimization and mutual information maximization. As shown in Tables 2 and 5, our unsupervised training method can learn useful knowledge from intermediate auxiliary data and thus promoting the transfer learning. And the best results gain by the DCAE indicate that the autoencoder architecture training by minimizing the reconstruction error is still an important method in unsupervised learning. The decoding process of DCAE helps the encoder to learn

hidden presentation of the unlabeled data. And our strategy make it possible to train in entire model rather than the conventional stack layer-wise.

The limitations of the proposed method for ILD pattern classification is less efficient than the slice-wise or semantic segmentation manner. But this limitation is caused by the database itself. For the ILD pattern classification task, we wish there will be a public available database with fully semantic labeling in the future.

6 Conclusion

For the purpose of improving the ILD pattern classification performance under the condition of limited training data, we propose a new deep CNN architecture and utilize a two-stage transfer learning strategy that successively training the model on the source domain, intermediate domain and target domain dataset. Experimental results show that our proposed CNN achieves good benchmark performance and surpasses most the state-of-the-art one. Furthermore, the performance can be further improved by the proposed two-stage transfer learning method which is capable of transferring knowledge learned from source domain and intermediate domain. The experimental results reveal that transfer learning is an effective way to address the data hungry problem. And most importantly, the auxiliary unlabeled data, which extracted from the unannotated areas of the lung field, are feasible to serve as the pre-training domain in transfer learning process. Meanwhile, the proposed unsupervised leaning method which incorporates the prediction confidence and mutual information into the objective function is proved effective in this task.

References

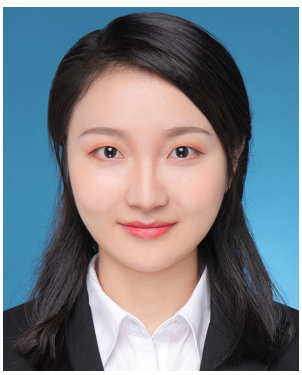
1. Altaf F, Islam S, Akhtar N, Janjua NK (2019) Going deep in medical image analysis: Concepts, methods, challenges and future directions. arXiv:1902.05655
2. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S (2016) Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imag 35(5):1207–1216
3. Aziz ZA, Wells AU, Hansell DM, Bain G, Copley SJ, Desai SR, Ellis SM, Gleeson FV, Grubnic S, Nicholson AG, et al. (2004) Hrct diagnosis of diffuse parenchymal lung disease: inter-observer variation. Thorax 59(6):506–511
4. Azizpour H, Razavian AS, Sullivan J, Maki A, Carlsson S (2016) Factors of transferability for a generic ConvNet representation. IEEE Trans Pattern Anal Mach Intell 38(9):1790–1802
5. Chen G, Zhang J, Zhuo D, Pan Y, Pang C (2019) Identification of pulmonary nodules via CT images with hierarchical fully convolutional networks. Medical & Biological Engineering & Computing:1–14
6. Cheplygina V, Pena IP, Pedersen JH, Lynch DA, Sørensen L., de Bruijne M (2018) Transfer learning for multicenter classification of chronic obstructive pulmonary disease. IEEE J Biomed Health Informat 22(5):1486–1496
7. Christodoulidis S, Anthimopoulos M, Ebner L, Christe A, Mougiakakou S (2017) Multisource transfer learning with convolutional neural networks for lung pattern analysis. IEEE J Biomed Health Informat 21(1):76–84
8. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. Proc IEEE Conf Comput Vis Pattern Recognit
9. Depeursinge A, Vargas A, Platon A, Geissbuhler A, Poletti PA, Müller H. (2012) Building a reference multimedia database for interstitial lung diseases. Comput Med Imag Graph 36(3):227–238
10. Gao M, Bagci U, Lu L, Wu A, Buty M, Shin HC, Roth H, Papadakis GZ, Depeursinge A, Summers RM (2015) Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks. 1st Workshop Deep Learn. Med. Image Anal:41–48
11. Guo W, Xu Z, Zhang H (2018) Interstitial lung disease classification using improved densenet. Multimed Tools Appl:1–12
12. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 770–778
13. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507
14. Hjelm RD, Fedorov A, Lavoie-Marchildon S, Grewal K, Trischler A, Bengio Y (2018) Learning deep representations by mutual information estimation and maximization. arXiv:1808.06670
15. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 4700–4708
16. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv:1412.6980
17. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Proc Adv Neural Inf Process Syst, pp 1097–1105
18. Kylberg G (2011) The Kylberg texture dataset v. 1.0, centre for image analysis, Swedish University of Agricultural Sciences and Uppsala University, external report (blue series) no 35
19. LeCun Y, Bottou L, Bengio Y, Haffner P, et al. (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
20. Li Q, Cai W, Wang X, Zhou Y, Feng DD, Chen M (2014) Medical image classification with convolutional neural network. In: Proc 13th Int Conf Control Automat Robot Vis. IEEE, pp 844–848
21. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Imag Anal 42:60–88
22. Lu Y, Chen L, Saidi A (2017) Optimal transport for deep joint transfer learning. arXiv:1709.02995
23. O'Neil A, Shepherd M, Beveridge E, Goatman K (2017) A comparison of texture features versus deep learning for image classification in interstitial lung disease. In: Proc Ann Conf Med Imag Und Anal. Springer, pp 743–753
24. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359
25. Pang S, Du A, Orgun MA, Yu Z (2019) A novel fused convolutional neural network for biomedical image classification. Medical & Biological Engineering & Computing 57(1):107–121
26. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A (2017) Automatic differentiation in pytorch

27. Samala RK, Chan HP, Hadjiiski L, Helvie MA, Richter CD, Cha KH (2018) Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets. IEEE Trans Med Imag
28. Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S (2014) Cnn features off-the-shelf: an astounding baseline for recognition. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp. 806–813
29. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imag 35(5):1285–1298
30. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
31. Sluimer I, Schilham A, Prokop M, Van Ginneken B (2006) Computer analysis of computed tomography scans of the lung: a survey. IEEE Trans Med Imag 25(4):385–405
32. Society B, Committee S (1999) The diagnosis, assessment and treatment of diffuse parenchymal lung disease in adults. Thorax 54(Suppl 1):S1–S28
33. Suzuki A, Sakanashi H, Kido S, Shouno H (2018) Feature representation analysis of deep convolutional neural network using two-stage feature transfer-an application for diffuse lung disease classification. arXiv:1810.06282
34. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 1–9
35. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: full training or fine tuning. IEEE Trans Med Imag 35(5):1299–1312
36. Tan B, Zhang Y, Pan SJ, Yang Q (2017) Distant domain transfer learning. In: Proc. 31th AAAI Conf Artif Intell
37. Tarando SR, Fetita C, Faccinetto A, Brillet PY (2016) Increasing cad system efficacy for lung texture analysis using a convolutional network. In: Medical imaging 2016: Computer-aided diagnosis, vol 9785, p 97850Q
38. Van Ginneken B, Armato S. GIII, de Hoop B, van Amelsvoortvan de Vorst S, Duindam T, Niemeijer M, Murphy K, Schilham A, Retico A, Fantacci ME, et al. (2010) Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. Med Imag Anal 14(6):707–722
39. Wei X, Chen J, Cai C (2017) Using deep convolutional neural networks and transfer learning for mammography mass lesion classification. J Comput Theor Nanos 14(8):3802–3806
40. Yap MH, Pons G, Martí J, Ganau S, Sentís M, Zwiggelaar R, Davison AK, Martí R (2018) Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J Biomed Health Informat 22(4):1218–1226
41. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? In: Proc Adv Neural Inf Process Syst, pp 3320–3328
42. Zheng L, Zhao Y, Wang S, Wang J, Tian Q (2016) Good practice in CNN feature transfer. arXiv:1604.00133
43. Zhuang F, Cheng X, Luo P, Pan SJ, He Q (2015) Supervised representation learning: transfer learning with deep autoencoders. In: Proc 24th Int Conf Artif Intell

Sheng Huang graduated from the University of Shanghai for Science and Technology and acquired the degree of Bachelor of Engineering in 2017. He is currently a master student at School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology. Research interests include machine learning and medical images analysis.

Feifei Lee received her Ph.D. degree in electronic engineering from Tohoku University in Japan, in 2007. She is currently a professor at the University of Shanghai for Science and Technology. Her research interests include pattern recognition, video indexing, and image processing.

Ran Miao graduated from the Xi'an Polytechnic University and acquired the degree of Bachelor of Engineering. He is currently a master student at School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, researching on machine learning and scene recognition task specifically



Qin Si graduated from Tongling University and acquired the degree of Bachelor of Engineering in 2017. She is currently a master student at school of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology. Research interests is machine learning.

Chaowen Lu is currently a master student at School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology. His current research interests are in the areas of image retrieval.

Qiu Chen received Ph.D. degree in electronic engineering from Tohoku University, Japan, in 2004. Since then, he has been an assistant professor and an associate professor at Tohoku University. He is currently a professor at Kogakuin University. His research interests include pattern recognition, computer vision, information retrieval and their applications. He is also a guest professor at the University of Shanghai for Science and Technology. Dr. Chen serves on the editorial boards of several journals, as well as committees for a number of international conferences.