Published in Medical Computer Vision: Algorithms for Big Data, MCV 2014

Unsupervised Pre-training across Image Domains Improves Lung Tissue Classification

Thomas Schlegl[1] *, Joachim Ofner[1] and Georg Langs[1]

Computational Imaging Research Lab, Department of Biomedical Imaging and Image-guided Therapy, Medical University Vienna, Austria
thomas.schlegl@meduniwien.ac.at, joachim.ofner@meduniwien.ac.at, georg.langs@meduniwien.ac.at

Abstract. The detection and classification of anomalies relevant for disease diagnosis or treatment monitoring is important during computational medical image analysis. Often, obtaining sufficient annotated training data to represent natural variability well is unfeasible. At the same time, data is frequently collected across multiple sites with heterogeneous medical imaging equipment. In this paper we propose and evaluate a semi-supervised learning approach that uses data from multiple sites (domains). Only for one small site annotations are available. We use convolutional neural networks to capture spatial appearance patterns and classify lung tissue in high-resolution computed tomography data. We perform domain adaptation via unsupervised pre-training of convolutional neural networks to inject information from sites or image classes for which no annotations are available. Results show that across site pre-training as well as pre-training on different image classes improves classification accuracy compared to random initialisation of the model parameters.

1 Introduction

Computer aided diagnosis often relies on automatic classification of observations made in medical imaging data. When constructing classifiers we face challenges such as limited annotated training data, the need to collect training data across multiple sites with potentially heterogeneous imaging hardware, and the choice of visual features that represent the data well, and are at the same time suited for discriminative learning. Existing methods often use handcrafted features. Manual feature engineering and feature selection requires appropriate expert knowledge. Apart from this, hand-crafted features are very domain-specific, thus applicable to other tasks or domains only to a limited extend. In contrast, shallow *convolutional neural networks* (*CNN*) [1] tend to learn low-level features, which

* This work has received funding from the European Union FP7 (KHRESMOI FP7-257528, VISCERAL FP7-318068), from the Austrian Science Fund (FWF P22578-B19, PULMARCH) and from the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development (OPTIMA).

are less domain-specific and thus yield better generalization to different domains. Before training of a CNN on annotated data starts, the model parameters have to be initialized with random values, or alternatively, with parameters pre-learned on non-annotated data. Random initialization is an appropriate approach only if large amounts of annotated data is available. In this paper we explore CNN as a means to learn low-level imaging features, to integrate non-annotated data, and to use data from multiple sites (*domains*) in the training process. A clinically highly relevant example where accurate classification of anomalies is crucial for treatment decisions are interstitial lung diseases. *Computed tomography* (*CT*) of the lung is an invaluable tool in the diagnostic process of interstitial lung diseases, which comprise a broad heterogeneous group of parenchymal lung disorders [2]. A wide range of anomalous patterns can be identified in high resolution CT scans of the lung, which correspond to specific lung diseases. Some of them show only subtly different appearance. Furthermore, the diagnosis of interstitial lung disease is a challenging task for computer-aided diagnosis systems as well as for human specialists, because various types of the disease occur with only low frequency (cf. [3]). In this paper, we study how CNN can be used to classify pathologies of the lung in CT imaging data and evaluate the ability of CNN to adapt across sites and image classes during unsupervised pre-training of feature extractors. We classify anomalous lung tissue (ground glass opacity, reticular interstitial pattern, honeycombing, emphysema) and normal lung parenchyma. Here, CNN allow to perform three tasks that are central to learning from partially annotated data obtained from different sites or different image classes. The algorithm learns representative and discriminative feature extractors based on the available partially annotated imaging data. It uses a substantial amount of non-annotated data to pre-train the network, and can inject data from different sites or different image classes to improve classification accuracy even if only data acquired on a single target site is annotated.

Related work. CNN were introduced in 1980 [1]. Since then, different variants of CNN were used to solve classification problems. The areas of successful application range from classification of handwritten digits (MNIST dataset) [4,5] to NORB [5] and ImageNet [6] datasets. Due to its ability to capture abstract representations deep learning applied successfully to unsupervised learning, transfer learning, domain adaptation and self-taught learning (cf. [7], [4]). Lee et al. presented in [4] an approach of self-taught learning for object recognition using CNN on non-medical images. But contrary to our proposed work, the target images were not taken from the medical domain. *Deep Belief Networks* (*DBN*) in general, and in particular CNN were successfully applied in detection tasks, such as mitosis detection in breast cancer histology images by means of supervised CNN [8]. In contrast to conventional DBN, CNN also use convolutional layers in the first few layers in addition to fully-connected layers. The beneficial effect of layer-wise unsupervised pre-training has been shown in [9]. *Convolutional Restricted Boltzmann Machines* (*CRBM*) [10] were used for manifold learning by reducing the dimensionality of 3D brain *magnetic resonance* (*MR*) images [11]. There, parameter learning in the frequency domain of the first CRBM layers was used to reduce the high dimensionality of the input images. Typically, DBN have been successfully applied in tasks processing relatively small images. Their application to image sizes characteristic for medical imaging (e.g., 256x256 or 512x512 for CT) remains challenging (cf. [4]).

Contribution. We use CNN to perform pixel-wise classification of lung tissue in 2D CT slices (images). Layer-wise unsupervised pre-training to initialize the parameters of a CNN is a well-known concept in deep learning theory. In the proposed work we aim to evaluate the beneficial effect on classification accuracy when this concept is applied to train a classifier on medical target images. The contributions of the paper are the data driven learning of spatial low-level features for lung tissue classification, the integration of unlabeled data in a pre-training phase, and the domain adaptation that allows the use of unlabeled data from different sites or different image classes to improve classification accuracy on medical images. The proposed approach focuses on classification tasks where only few labeled data from the target domain but large amounts of unlabeled data from different domains are available. We evaluate the algorithm on CT data of the lung and brain collected at the Vienna General Hospital, the STL-10 dataset [12] and the *Lung Tissue Research Consortium* (*LTRC*) dataset [13], and study the effect of unsupervised pre-training and supervised fine-tuning on different sites or image classes in detail.

2 Restricted Boltzmann Machines

A *Restricted Boltzmann Machine* (*RBM*) is an undirected graphical model with two layers. The first layer consists of a set of binary or real-valued input units v - also referred to as visible units - of dimension C and the second layer consists of a set of binary hidden units h of dimension B. The units of both layers are fully-connected by a weight matrix $W \in \mathbb{R}^{C \times B}$, i.e. every visible unit is connected with every hidden unit. The model parameteres of an RBM are trained to perform some kind of (non-linear) transformation between visible and hidden units. A DBN is a generative model that is constructed by stacking RBM on top of each other. Adjacent RBM within a DBN are in turn fully connected. Hinton et al. [14] showed that deep models can be efficiently trained by greedily training each layer as an RBM. The first RBM is trained with input samples. The single RBM from the second RBM upwards are trained by using the activations of the previous layer as inputs. RBM and DBN have an important limitation in common when using images as inputs. Both ignore the 2D structure of the input image. A CNN is a feature extractor that preserves the 2D structure of the input. The architecture of CNNs is motivated by biological vision [15].

A CNN is a feed-forward network that is hierarchically structured with one or more pairs of convolutional and max-pooling layers followed by one or more *fully-connected layers*. The *convolutional layers* act as detection layers. Each convolutional layer maps the input to Γ groups. Thus every convolution layer learns Γ different feature detectors. Because the weights of every group are

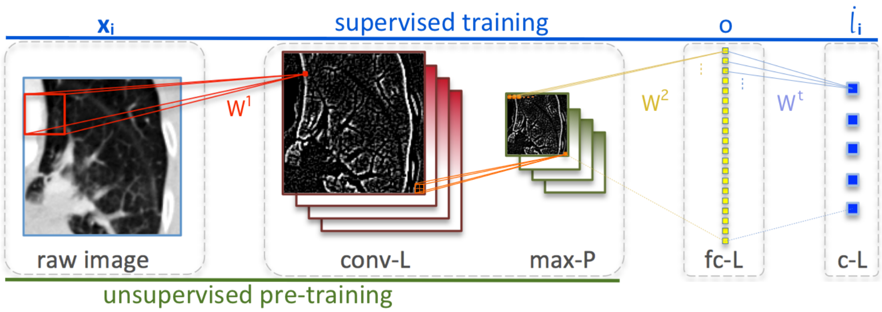

Fig. 1: Illustration of the architecture of our shallow CNN with one convolution (*conv-L*) and max-pooling layer (*max-P*), one fully-connected layer (*fc-L*) and a classification layer (*c-L*).

shared across the whole input image, the number of parameters to be learned is massively reduced so that CNNs scale well to full images. *Max-pooling* layers are stacked on top of every convolution layer. A pooling layer shrinks the size of the matrix of activations of the preceding detection layer by a constant factor by only taking the maximum activation within small non-overlapping regions. The pooling layer has no parameters that have to be trained. For classification tasks, a terminal *classification layer* is needed. We use softmax regression as classifier which enables our model to perform multi-class classification. Figure 1 illustrates the architecture of the CNN used in our experiments.

3 Domain Adaptation in Lung Tissue Classification

We are given two different datasets. The first dataset comprises pairs of 2D medical imaging data and corresponding pixel-wise class labels $\langle \mathbf{I}_m, \mathbf{L}_m \rangle$, with $m = 1, 2, \ldots, M$, where $\mathbf{I}_m \in \mathbb{R}^{n \times n}$ is an image of size $n \times n$ of pixel intensities, and $\mathbf{L}_m \in \{1, ..., K\}^{n \times n}$ is an array of the same size containing the corresponding class labels. The second dataset comprises only 2D image data $\mathbf{J}_u$, with $u = 1, 2, \ldots, U$ without corresponding class labels. We extract small input image patches $\mathbf{x}_i^I \in \mathbb{R}^{s \times s}$ with $s < n$ from image $\mathbf{I}_m$, corresponding patches of ground truth class labels $\mathbf{l}_i \in \{1, ..., K\}^{s \times s}$ from $\mathbf{L}_m$ and input image patches $\mathbf{x}_i^J \in \mathbb{R}^{s \times s}$ from $\mathbf{J}_u$ respectively, where s is the width and height and i the index of the centroid of the patch. The true class label $\dot{l}_i$, for patch $\mathbf{x}_i^I$ corresponds to the mode of given ground truth class labels in $\mathbf{l}_i$. The class label $\dot{l}_i$ is assigned to the whole image patch $\mathbf{x}_i^I$ centered at pixel position i. We use $\langle \mathbf{x}_i^I, \dot{l}_i \rangle$ pairs for supervised training of our model. Our objective is to learn a mapping $f : \mathbf{x}_i^I \longmapsto \dot{l}_i$ from image patches $\mathbf{x}_i^I$ to corresponding class labels $\dot{l}_i$ in a semi-supervised fashion. During testing we apply the mapping to new image patches in the test set.

During classification, an unseen image patch $\mathbf{x}_i$ causes an activation o of the fully-connected layer (fc-L) of the CNN (Figure 1). The activation o of the fully-connected layer is the input of the classification layer. The classification layer t has as its parameters a weight matrix W^t and a bias term a^t. We apply the softmax function on activations of the classification layer which gives us predictions $\hat{l}_i$ for the class k having the highest class membership probability:

$$\hat{l}_i = \operatorname*{argmax}_k P(\dot{l}_i = k | o, W^t, a^t). \tag{1}$$

During training, all weights and bias terms of the whole model are optimized by minimizing the misclassification error on image patches of the training set.

3.1 Unsupervised Pre-training

We can pre-train convolutional neural networks on unlabeled data to improve the training procedure [4]. Before supervised training starts, we use CRBM layers as a means for unsupervised pre-training of the CNN parameters of the convolutional layers. CRBM consist of an input layer and $\gamma = 1, ..., \Gamma$ groups of hidden units. Its parameters are learned via block Gibbs sampling [4] using the conditional distributions for the hidden units h in group γ

$$P(h_\gamma = 1 | v) = \sigma\left((\tilde{W}_\gamma * v) + b\right) \tag{2}$$

and for the visible units v

$$P(v|h) = \sum_\gamma (W_\gamma * h_\gamma) + c. \tag{3}$$

Here, $*$ is the convolution operation, b is the bias term of the hidden units in group γ, c is the bias term of the visible units and $\sigma(q) = \frac{1}{1+e^{-q}}$ is the sigmoid function. The tilde above the weight matrix $\tilde{W}$ denotes the usage of a horizontally and vertically flipped version of matrix W. Lee et al. [4] proposed a technique, also referred to as probabilistic max-pooling, which allows to stack CRBM into a multilayer model that is referred to as *Convolutional Deep Belief Network* (*CDBN*). This model integrates the information whether a pooling unit is on or off. The random variables for the hidden units h are sampled from a multinomial distribution including this information.

Nair et al. [16] showed that the hidden units of a Restricted Boltzmann machine can be approximated efficiently by noisy rectified linear units. Thus, instead of sampling the hidden units based on the conditional distribution given in equation (2) we use the noisy rectified activation $A(h)$ of hidden units h which is given by

$$A(h_\gamma) = max\left(0, x + N(0, \sigma(x))\right), \tag{4}$$

where $N(0, V)$ is the added Gaussian noise with zero mean and variance V. The variance V is the sigmoid activation function σ applied to the convolved input

$$x = (\tilde{W}_\gamma * v) + b. \tag{5}$$

At this point, instead of using unsupervised pre-training exclusively from the same samples $\mathbf{x}_i^I$ as used for supervised training, we evaluate the accuracy of the classifier f by using additional unlabeled samples $\mathbf{x}_i^J$ of either the same, or other domains during pre-training of the CRBM.

4 Experiments

Data. We perform experiments on two clinical CT datasets of the lung, on a clinical CT dataset of the brain and on a natural image dataset (see Figure 2).

The clinical lung datasets comprise clinical high-resolution lung CT scans of different patients from two different sites. The first lung dataset (V) comprises unlabeled clinical data from the Vienna General Hospital. The second lung dataset (L) comprises a subset of data from the LTRC dataset [13].

Dataset-L contains 20,000 2D image patches extracted from axial slices of 380 scans from the LTRC dataset provided by the Lung Tissue Research Consortium of the National Heart Lung and Blood Institute (NHLBI). It contains data of patients predominantly suffering from Chronic Obstructive Pulmonary Disease (COPD) or Interstitial Lung Disease (ILD).

Dataset-LL contains 1000 labeled 2D image patches extracted from axial slices of a randomly selected subset of the 380 scans of the LTRC dataset. These image patches are used for unsupervised pre-training (without using the corresponding labels) and for supervised fine-tuning (including the corresponding patch labels). This dataset contains pixel-wise annotations of all patterns listed above. There is no overlap between samples of dataset-L and dataset-LL regarding pixel locations of image patch centers.

Dataset-V consists of 20,000 unlabeled 2D image patches extracted from axial slices of 65 randomly selected clinical lung CT scans, not restricted to the labeled pathologies in L but containing a large variation of different pathologies such as chronic obstructive pulmonary disease, cyst, pneumonia, bronchiectasis or space-occupying lesions.

The brain dataset, ***Dataset-B***, contains 20,000 unlabeled 2D image patches extracted from axial slices of 427 randomly selected clinical high-resolution head CT scans.

Finally, we also use non-medical unlabeled natural images. ***Dataset-S*** contains 20,000 unlabeled image patches extracted from the STL-10 dataset [12].

Data selection and preprocessing We use 2D axial slices from the clinical datasets and grayscale versions of the STL-10 dataset for pre-training or supervised fine-tuning of the classifier. Images of CT scans have typically image resolutions of 512×512 pixels. From these images we extract 40×40 pixel patches x_i centered at pixel positions i. The positions of the patch centers as well as the slice positions within the CT scan (in the case of clinical datasets) are sampled randomly. The image patches are preprocessed by transforming the data to zero-mean and unit variance. Standardization of real valued data is a common requirement in deep learning.

Evaluation. We train a classifier to differentiate between five tissue classes (ground glass opacity, reticular interstitial pattern, honeycombing, emphysema and healthy lung tissue). We perform two experiments.

(1) In our first experiment we perform supervised training using dataset-LL (the target domain) and evaluate the classification accuracy when pre-training is

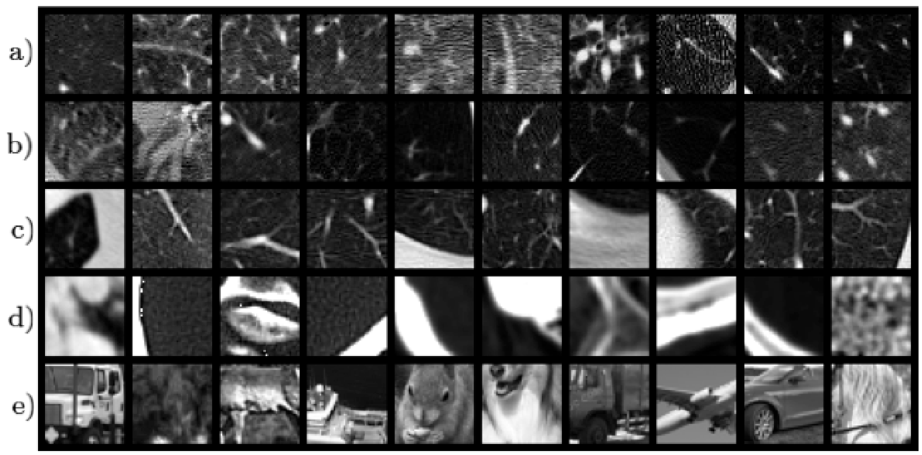

Fig. 2: Datasets used in the experiments: a) dataset-LL, b) dataset-L, c) dataset-V, d) dataset-B, e) dataset-S. The labels which correspond to the patches from left to right of dataset-LL (a) are as follows: 'normal', 'normal', 'ground glass', 'ground glass', 'reticular', 'reticular', 'honeycomb', 'honeycomb', 'emphysema', 'emphysema'.

performed on different unlabeled datasets. Supervised fine-tuning is performed on 50, 100, 150 or 200 samples of dataset-LL, which corresponds to 0.25%, 0.5%, 0.75% and 1% of the samples used for unsupervised pre-training (in experiments 1b-e) respectively. These samples are always already included in the pre-training phase of all scenarios without using the class labels. We evaluate the minimum classification error over the fine-tuning epochs. We use a random split of these samples to perform 5-fold-cross-validation. The performance measure of the classifier is the average of misclassification errors over the 5-fold-cross-validation runs. The images in the test set of the supervised fine-tuning step are also excluded from the training set of the pre-training phase. We evaluate the following pre-training scenarios. **(1a)** We evaluate the classification accuracy when pre-training is only performed on samples (LL) that are used for fine-tuning. **(1b)** Unsupervised pre-training on dataset-L evaluates the classification accuracy when a larger unlabeled image set of the target site is available. **(1c)** By including dataset-V in the pre-training we evaluated if the unsupervised pre-training across different sites yields comparable classification accuracy. **(1d)** Instead of pre-training on images of the same type (lung CT scans) we perform pre-training on brain CT scans (B) and **(1e)** on natural images (S) to evaluate if the learning of image feature extractors generalizes across image types. The last two scenarios simulate cases where only very little imaging data of a specific anatomical site is available. **(1f)** The classification accuracy is also evaluated for training the classifier without performing any pre-training.

(2) In the second experiment we vary the training set size for supervised fine-tuning from 10 patches to $20,000$ patches, to understand how even with minimal sets of annotated images, we can train classifiers with reasonable accuracy, if pre-training has been performed. Furthermore, we evaluate the performance of the classifier using all patches for supervised fine-tuning that are used for unsupervised pre-training. In this experiment we use all samples of dataset-L for unsupervised pre-training and samples of dataset-LL for supervised fine-tuning. Where necessary we also use additional samples of dataset-L and corresponding class labels in the supervised fine-tuning phase.

All experiments are implemented in Python 2.7 using the Theano [17] library and run on a graphics processing unit (GPU).

4.1 Model Parameters

We use gaussian visible units to model real valued data with the CRBM. Each CRBM pre-training is performed for 200 epochs. We choose a simple CNN architecture, intentionally. The shallowness of the network allows to focus on the extraction of low-level features, learned on different datasets. The *CNN* (see Figure 1) is hierarchically structured with 1 convolution layer, 1 fully-connected layer and a classification layer with 5 classification units. The convolution layer contains 32 groups of hidden units and is followed by a max-pooling layer. The filter size of the convolution layer is set to 5×5. The fully connected layer consists of 1000 neurons. For all classification experiments there is no overlap between the training set and the test set regarding pixel locations of image patch centers. But some patches share few pixels. The pre-training with different datasets results in corresponding parameter sets of learned weights W and bias terms b of the hidden units of the CRBMs. Before fine-tuning starts, we initialize the convolution layer of different CNNs with the pre-learned parameters. Because of normalizing the input data to zero mean and unit variance the bias terms c of the visible units have not to be learned. Fine-tuning of each CNN is performed for 400 epochs.

4.2 Classification Results

In experiment **(1)** we evaluate domain adaptation on dataset-V, dataset-B and dataset-S. Pre-training only on dataset-LL serves as the most restricted case, where no additional training samples are available or simply not used. We evaluate also the performance of the CNN without performing prior pre-training. This scenario also serves as reference scenario. Pre-training on dataset-L involves no domain adaptation of model parameters between unsupervised pre-training and supervised fine-tuning but this is an obvious approach when only a restricted amount of labeled data and a larger amount of unlabeled data of the same domain is available. Results are shown in Table 1.

Results show that the beneficial effect of unsupervised pre-training depends on the domain of the data used for pre-training. Pre-training on datasets of similar domains, both clinical CT datasets of the lung, result in similar performance. Dataset-L and dataset-V improve classification accuracy over only supervised training on dataset-LL. The performance of the fine-tuned CNN after pre-training on dataset-LL is slightly better than the performance of the

Table 1: Misclassification errors of the CNN pre-trained on different datasets and the corresponding sample sizes of dataset-LL used for supervised fine-tuning. The results of pre-training datasets with overall minimum misclassification errors are highlighted.

	no pre-training	dataset-LL	dataset-L	dataset-V	dataset-B	dataset-S
50	0.2853	0.2709	0.2600	0.2737	0.3008	**0.2499**
100	0.2076	0.1929	**0.1724**	0.1924	0.2242	0.1736
150	0.2057	0.1901	0.1656	0.1792	0.2148	**0.1628**
200	0.1977	0.1893	0.1633	0.1773	0.2138	**0.1604**

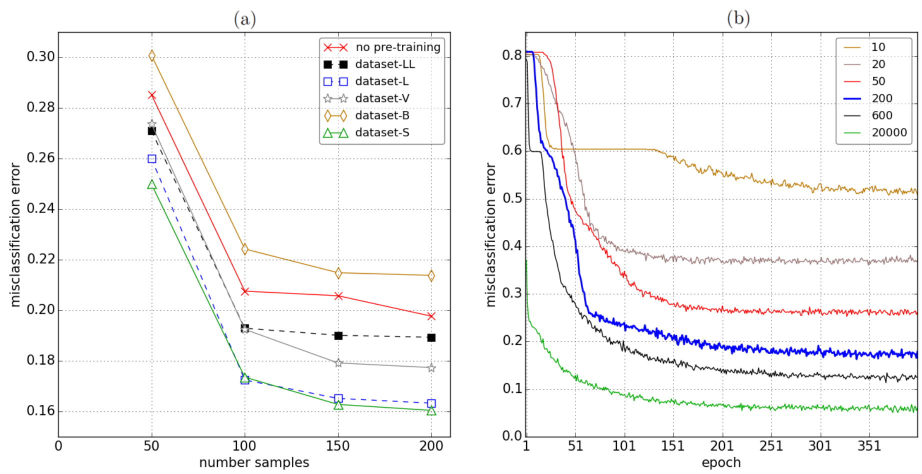

Fig. 3: (a) Misclassification error as a function of number of samples used for supervised fine-tuning using unsupervised pre-training on different datasets. (b) Misclassification error of a single run of the 5-fold-cross-validation setup as a function of training epochs using different sample sizes for supervised fine-tuning after pre-training on dataset-L.

CNN with no prior pre-training. Unsurprisingly, pre-training on dataset-B does not improve the performance of the classifier. The performance of the CNN initialized with random weights (*no pre-training*) is even better than the performance of the CNN that is pre-trained on dataset-B. All the more surprising, pre-training on dataset-S, a domain of natural images, performs comparably or even slightly better than the pre-training on medical images of the lung (dataset-L) in the classification of visual lung tissue patterns. A summary of the results of experiment **(1)** is given in Table 1.

Figure 3 (a) shows the misclassification errors as a function of the number of samples used for supervised fine-tuning and different pre-training datasets. Increasing the number of annotated data in the fine-tuning phase reduces the misclassification error. This holds true for all datasets used for pre-training. The choice of the dataset used for pre-training as well as the number of samples of the target site used for fine-tuning influence the classification accuracy.

Table 2: Misclassification error of the CNN for different numbers of annotated images used for supervised fine-tuning after pre-training on dataset-L.

Sample size	Error
10	0.5048
20	0.4145
50	0.2600
200	0.1633
300	0.1190
600	0.1130
1000	0.1094
1500	0.1029
3000	0.0878
10000	0.0674
20000	0.0531

(2) We investigate to what extend the sample size of labeled data of the target domain can be reduced when pre-training is performed. Unsupervised pre-training is always performed on the entire dataset-L. Results show that a model that is fine-tuned with only 10 randomly chosen image patches from the dataset-LL already performs better than chance. Fine-tuning the model with all samples of dataset-LL and dataset-L (20,000 patches) results in a minimum misclassification error of 5.3% on the 5 class classification task. The number of lung image patches used for supervised fine-tuning and the corresponding misclassification errors are summarized in Table 2. Figure 3 (b) shows the misclassification error of the CNN as a function of training epochs for different sample sizes of lung images used for supervised fine-tuning. Again, the performance measure corresponds to the results of the 5-class classification task.

5 Discussion

We propose methodology for improving lung tissue classification by unsupervised pre-training of CNN with samples drawn from different data and clinical sites. In contrast to previous work, we perform unsupervised pre-training of a CNN not (only) on medical images sampled from the distribution which is also used for fine-tuning the model. The overall classification performance can be improved via unsupervised pre-training by using large amounts of additional data, that is similar to the target domain. This is relevant in cases where only part of the training data is annotated, and data has to be collected across different sites to obtain sufficient training set size. It indicates that injecting data from different sites during pre-training can improve results, even if their characteristics are slightly different from the target site. The proposed classification outperforms previous approaches on the LTRC data. Zavaletta et al. [18] presented an approach for 5-class classification on the LTRC dataset using canonical signatures based on an adaptive histogram binning algorithm. Their classification of cubic image patches ($15 \times 15 \times 15$ voxels) of the lung into the same classes we use in our experiments yielded an overall misclassification error of 27.33%. Pre-training results in good classifier performance, even if only very small numbers of annotated data are available for supervised fine-tuning. Surprisingly, we observe that a domain of natural images outperforms the pre-training on clinical CT scans of the lung in the classification of visual lung tissue patterns. This can be explained due to the fact that pre-training of a single convolution layer leads to learning of low-level features which are not as domain-specific as for example features learned by pre-training and stacking more than one convolution layer on top of each other. Furthermore, natural images show fine textures which are advantageous for the learning of low-level features. In contrast, pre-training on clinical CT datasets of the brain lead to a worse performance of the CNN on the lung tissue classification task. Contrary to natural images and CT images of the lung as well, CT images of the brain comprise large regions of homogeneous gray values and less texture information. Thus they are not as beneficial for pre-training of a low-level feature extractor. The proposed domain adaptation is relevant in cases where only few labeled training data but large amounts of unlabeled data from a similar domain is available. We conclude, that unsupervised pre-training using additional data from a slightly different domain can lead to better generalization from the training data set, and improves classification performance.

References

1. Fukushima, K.: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological cybernetics **36**(4) (1980) 193–202
2. Ryu, J.H., Daniels, C.E., Hartman, T.E., Yi, E.S.: Diagnosis of interstitial lung diseases. In: Mayo Clinic Proceedings. Volume 82., Elsevier (2007) 976–986
3. Depeursinge, A., Sage, D., Hidki, A., Platon, A., Poletti, P.A., Unser, M., Muller, H.: Lung tissue classification using wavelet frames. In: Engineering in Medicine and Biology Society. 29th Annual International Conference of the IEEE. (2007) 6259–6262
4. Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Unsupervised learning of hierarchical representations with convolutional deep belief networks. Communications of the ACM **54**(10) (2011) 95–103
5. Ciresan, D., Meier, U., Schmidhuber, J.: Multi-column deep neural networks for image classification. In: Conference on Computer Vision and Pattern Recognition, IEEE (2012) 3642–3649
6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. Volume 1. (2012) 4
7. Bengio, Y.: Deep learning of representations for unsupervised and transfer learning. Journal of Machine Learning Research-Proceedings Track **27** (2012) 17–36
8. Ciresan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: Medical Image Computing and Computer-Assisted Intervention. Volume 2. (2013) 411–418

9. Erhan, D., Bengio, Y., Courville, A., Manzagol, P.A., Vincent, P., Bengio, S.: Why does unsupervised pre-training help deep learning? The Journal of Machine Learning Research **11** (2010) 625–660
10. Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th Annual International Conference on Machine Learning. (2009) 609–616
11. Brosch, T., Tam, R.: Manifold learning of brain MRIs by deep learning. Medical Image Computing and Computer-Assisted Intervention (2013) 633–640
12. Coates, A., Ng, A.Y., Lee, H.: An analysis of single-layer networks in unsupervised feature learning. In: International Conference on Artificial Intelligence and Statistics. (2011) 215–223
13. Holmes III, D., Bartholmai, B., Karwoski, R., Zavaletta, V., Robb, R.: The lung tissue research consortium: an extensive open database containing histological, clinical, and radiological data to study chronic lung disease. In: The Insight Journal MICCAI Open Science Workshop. (2006)
14. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural computation **18**(7) (2006) 1527–1554
15. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11) (1998) 2278–2324
16. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning. (2010) 807–814
17. Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., Bengio, Y.: Theano: a cpu and gpu math expression compiler. In: Proceedings of the Python for scientific computing conference (SciPy). Volume 4. (2010)
18. Zavaletta, V.A., Bartholmai, B.J., Robb, R.A.: High resolution multidetector ct-aided tissue analysis and quantification of lung fibrosis. Academic radiology **14**(7) (2007) 772–787