2014 13th International Conference on Control, Automation, Robotics & Vision
Marina Bay Sands, Singapore, 10-12th December 2014 (ICARCV 2014)

Medical Image Classification with Convolutional Neural Network

Qing Li*, Weidong Cai*, Xiaogang Wang†, Yun Zhou‡, David Dagan Feng* and Mei Chen§
*Biomedical and Multimedia Information Technology (BMIT) Research Group, School of IT, University of Sydney, Australia
†Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong
‡Johns Hopkins University School of Medicine, United States
§Intel Science and Technology Center on Embedded Computing, Carnegie Mellon University, United States

***Abstract*—Image patch classification is an important task in many different medical imaging applications. In this work, we have designed a customized Convolutional Neural Networks (CNN) with shallow convolution layer to classify lung image patches with interstitial lung disease (ILD). While many feature descriptors have been proposed over the past years, they can be quite complicated and domain-specific. Our customized CNN framework can, on the other hand, automatically and efficiently learn the intrinsic image features from lung image patches that are most suitable for the classification purpose. The same architecture can be generalized to perform other medical image or texture classification tasks.**

I. Introduction

In image classification problems, the descriptiveness and discriminative power of features extracted are critical to achieve good classification performance. Feature extraction techniques commonly used in medical imaging include intensity histograms, filter-based features [18], [22], and the recently very popular scale-invariant feature transform (SIFT) [17], [22] and local binary patterns (LBP) [16], [18], [22]. The feature vectors extracted are normally used to train a classification model, e.g. the support vector machine (SVM) [4], [11], [12], [17], [23], [25], [26] and sparse representation [10], [15], [16], [18], [20], [24], to obtain the image label.

Artificial Neural Network (ANN) has been studied for many years to solve complex classification problems including image classification [7]. The distinct advantage of neural network is that the algorithm could be generalized to solve different kinds of problems using similar designs. Convolutional Neural Network (CNN) is a successful example of attempts to model mammal visual cortex using ANN. The architecture of CNN has strong biological plausible evidence support from Hubel and Wiesels early work on the cats visual cortex [5], [8]. It has demonstrated superior performance in solving many hard image classification problems. In certain applications, such as traffic sign detection, CNN-based system has even surpassed human capability in benchmarking tests [14].

Our aim of this study is to adapt CNN network to classify different categories of lung ILD patterns. ILD comprises a group of more than 150 different disorders of the lung parenchyma [19]. They cause scarring of the lung tissues and lead to breathing difficulties. High-resolution computed tomography (HRCT) imaging is usually used to differentiate the different categories of ILDs. However, due to high visual variation within the same class and high visual similarity between different classes, it is very challenging to achieve accurate classification. Such difficulties can be seen from Fig. 1. The main problem is thus how to design a highly discriminative feature set to effectively handle the within-class variation and between-class similarity. Recently customized feature design has been proposed with good classification performance [2], [4]. Unsupervised Restricted Boltzmann Machine (RBM) based neural network has also been experimented to automatically learn features for ILD classification [9].

In this work, we propose a customized CNN network for lung image patch classification. Rather than defining a set of features manually [2], [4], we designed a fully automatic neural-based machine learning framework to extract discriminative features from training samples and perform classification at the same time. Being not problem-specific, our method can be easily applied to any other imaging domains. Compared with unsupervised RBM neural network [9], CNN uses supervised learning algorithm and hence better classification result would be expected. In addition, since lung images do not exhibit distinct visual structures and have relatively small number of training samples, we customized the general CNN architecture to tackle these issues. In particular, we incorporated random neural node drop-out and used a single convolutional layer architecture to reduce the number of parameters in the CNN model to avoid the over-fitting problem. For flexible experimentation, we implemented our own neural network toolkit including CNN and RBM with performance acceleration using Advanced Vector Extensions (AVX).

II. Methods

CNN has been proved very successful in solving image classification problems. Research works based on CNN significantly improved the best performance for many image databases, including the MNIST database, the NORB database and the CIFAR10 dataset. It is very good at learning the local and global structures from image data. General image objects like hand written numbers or human faces have obvious local and global structures, hence simple local features such as edges and curves can be combined to become more complex features such as corners and shapes and eventually the objects. Recently, CNN has also been incorporated into medical imaging analysis, such as the knee cartilage segmentation [13]. Differ-

978-1-4799-5199-4/14/$31.00 ©2014 IEEE

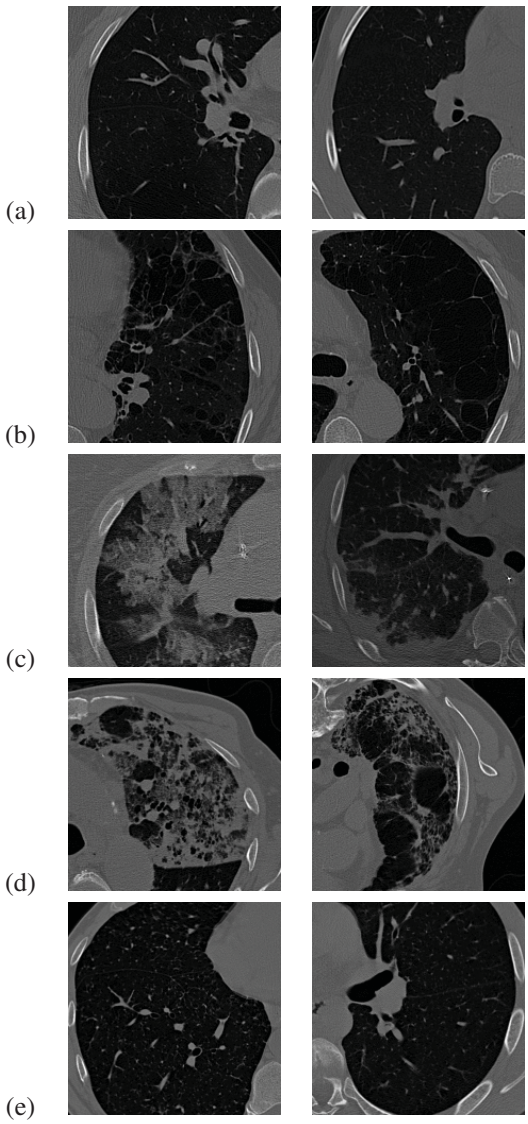


Fig. 1. Two example HRCT images are shown for each category of ILD tissues: (a) normal; (b) emphysema; (c) ground glass; (d) fibrosis; (e) micronodules.

ent from these published studies, we observe that lung image patches are more texture-like that have no distinct structures, hence deep layers in CNN would actually not perform well on such data. Over-fitting is another potential problem when training large neural network with many parameters, especially for medical image database with limited number of samples. In this research, we thus proposed a CNN network architecture which is specially adapted for texture-like multi-class image classification and turned to avoid over-fitting problem.

A. Network Architecture

The overall framework design is illustrated in Fig. 2. Input of the network is normalized lung image patch with unit variance and zero mean. The first layer is a convolutional layer with kernel size of 7×7 pixels and 16 output channels. The second layer is a max pooling layer with 2×2 kernel size. The following three layers are fully connected neural layers with 100-50-5 neurons in each layer. Compare to other applications of CNN, there is only a single convolutional layer in our design. As the input patches are textures-like images, there are no obvious large scale or high level features for the network to learn. The network with a single convolutional layer performs as good as networks with multiple convolution layers in ILD lung image patch classification tasks. The simpler network architecture also greatly reduced the number of parameters to be learned, which helps to avoid the over-fitting problem.

A number of techniques that improve and speed-up neural network training are applied. The learning parameters related to these learning techniques are selected based on typical values or chosen according to empirical studies. In the following sections, each layer of the CNN network is presented in detail.

B. Image Feature Extraction Using Convolution

Image pixels could be directly used as input to the standard feed-forward neural networks to solve image classification problems. However even relatively small image patches may have thousands of pixels, resulting in very large number of connection weight parameters to be trained. According to the VC dimension theory, large number of weight parameters result in more complex systems that need a lot more training samples in order to avoid over-fitting problem. In contrast, CNN models combine weights into much smaller kernel filters that dramatically simplify the learning model; and CNN networks are much faster and more robust than traditional fully-connected neural networks.

C. Convolutional Layer

Convolutional neuron layers are the key component of CNN. In image classification tasks, one or more 2D matrices (or channels) are treated as the input to the convolutional layer and multiple 2D matrices are generated as the output. The number of input and output matrices may be different. The process to compute a single output matrix is defined as:

$$A_j = f(\sum_{i=1}^{N} I_i * K_{i,j} + B_j) \tag{1}$$

Firstly each input matrix I_i is convoluted with a corresponding kernel matrix $K_{i,j}$. Then the sum of all convoluted matrices is computed and a bias value B_j is added to each element of the resulting matrix. Finally a non-linear activation function f is applied to each element of the previous matrix to produce one output matrix A_j.

Each set of kernel matrices represents a local feature extractor that extracts regional features from the input matrices. The aim of the learning procedure is to find sets of kernel matrices K that extract good discriminative features to be used for image classification. The back propagation algorithm that optimizes neural network connection weights can be applied here to train the kernel matrices and biases as shared neuron connection weights.

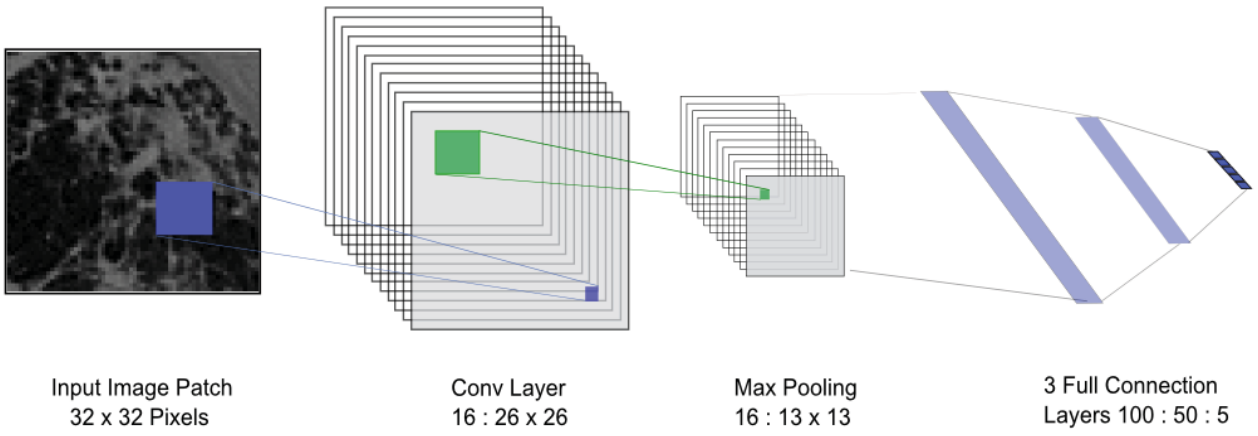


Fig. 2. Network architecture.

D. Pooling Layer

Pooling layer plays an important role in CNN for feature dimension reduction. In order to reduce the number of output neurons in the convolutional layer, pooling algorithms should be applied to combine the neighbouring elements in the convolution output matrices. Commonly used pooling algorithms include max-pooling and average-pooling. In this work, max-pooling layer with 2×2 kernel size selects the highest value from the 4 neighbouring elements of the input matrix to generate one element in the output matrix. During error back propagation process, the gradient signal must be only routed back to the neurons that contribute to the pooling output.

E. ReLU Neuron Activation

In ANN, non-linear transfer functions are used as the neuron activation function. For example, commonly used activation functions include sigmoid function $f(x) = 1/(1 + e^{-x})$ and hyperbolic tangent function $f(x) = \tanh(x)$. Both sigmoid and hyperbolic tangent are saturating non-linear functions that the output gradient drops close to zero as the input increases. Some recent studies suggested that non-saturating non-linear function like rectified linear function $f(x) = max(0, x)$ (ReLU) improves both learning speed and classification performance in CNN applications [6]. In our CNN model, the ReLU activation function is used in the convolutional layer. Testing results showed that the ReLU activation function improves classification performance by 2.5% and the network converges much faster than using the sigmoid activation function.

F. Back Propagation Details

Techniques for speeding up and stabilizing neural network training are applied, including batch learning, momentum and weight decay. Batch learning is applied to improve learning speed and accuracy. Instead of updating the connection weights after every single back propagation, we process 128 input samples in a batch and then perform one single update for the whole batch.

In order to further speed up the learning, momentum weight updating combined with weight decay is applied. The weight $\Delta\omega_i$ is updated with:

$$\Delta\omega_i(t+1) = \omega_i(t) - \eta\frac{\partial E}{\partial w_i} + \alpha\Delta\omega_i(t) - \lambda\eta\omega_i \qquad (2)$$

The $\omega_i(t) - \eta\frac{\partial E}{\partial w_i}$ part is the standard back propagation term, where $\omega_i(t)$ is the current weight vector, $\frac{\partial E}{\partial w_i}$ is the error gradient with respect to the weight vector and η is the learning rate. The $\alpha\Delta\omega_i(t)$ is the momentum part, where α is the momentum rate. The momentum weight update term helps to speed up learning. The $-\lambda\eta\omega_i$ is the weight decay part, where λ is the weight delay rate. It slightly reduces / decays the weight vector towards zero in each learning iteration, which helps stabilizing the learning process. In our experiment we choose learning rate $\eta = 0.001$, momentum rate of $\alpha = 0.9$ and decay rate $\lambda = 0.01$ for all layers. These learning parameters are selected through experimental tests.

G. Drop-out Algorithm

Drop-out algorithm is applied to improve the performance, by randomly disabling neurons in each layer during training [1]. A drop-out map with the same size of the neurons in each layer is randomly initialized to mark the on or off state of the corresponding neuron at the start of each training iteration. The neurons with off state are then removed from the network during the training iteration, by disabling the activation signal forward propagation and error signal backward propagation of the neuron. It is equivalent to switching between different models for each learning iteration, so that many different models are trained at the same time. During testing, all neurons are turned on, but with the activation signal attenuated to the probability of average turn on rate during the training phase. In our experimentation, drop-out rate of 0.6 is chosen during training, the neuron activation output was multiplied by 0.4 during testing.

H. Implementation

In order to experiment using different architectures and algorithms, a flexible neural computing software platform is required to adapt to very different designs include future ideas. Therefore, instead of using the existing open source CNN implementation, we have implemented our own neural network framework that is able to train different types of neural systems including CNN, RBM, autoencoder and more. The CNN network training is a very computational intensive task. Advanced Vector Extensions (AVX) optimized CNN and neural computing software is thus implemented for this research work. AVX acceleration is chosen over the popular GPGPU-based technology, because software that runs on CPU is more flexible in supporting various complex algorithms on general purpose CPU than the specialized GPU architecture. One hundred Giga-flop operations per second is achieved on a high-end Intel i7 processor for both convolution and full connection neuron computation.

III. Experimental Results

A. Datasets

The publicly available ILD database [3] is used for evaluation. The database contains 113 sets of HRCT images, with 2062 2D regions of interest (ROI) annotated indicting the ILD category. Following [3], we chose to classify image patches of five ILD categories: normal (N), emphysema (E), ground glass (G), fibrosis (F) and micronodules (M). The CT slices were divided into half-overlapping image patches of 32×32 pixels. Only image patches with at least 75% of its pixels falling inside of an annotated ROI were used in the experiment. The dataset thus contains 16220 image patches from 92 HRCT image sets, including 4348 N patches, 1047 E patches, 1953 G patches, 2591 F patches, and 6281 M patches.

The training samples are evenly divided into 10 groups. In each testing session, samples from one group is chosen as the testing data, and all other samples from the remaining 9 groups are used for training. Final results are collected after 10 testing sessions, each time with a different group as the testing data. Random image shifting is performed on the training image patches, to artificially increase the number of training samples to avoid the over-fitting issue.

B. Visualizing the Learned Features

Unlike the traditional fully-connected neural networks, which behave like a black box learning machine, features learned by CNN can be easily visualized and understood [21]. The kernel matrices in the convolutional Layer represent sets of features learned by the network. These features can be visualized in Fig.3. It is interesting that the feature filters look similar to two dimensional discrete cosine transformation (DCT) kernel functions. Based on this observation, it is clear that 2D spacial frequency information has been learned by the network as good discriminative features to distinguish the texture-like lung image patches.

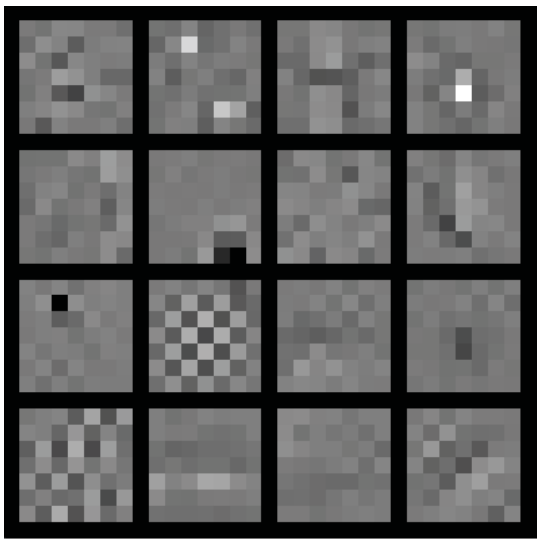


Fig. 3. DCT-like kernel matrices visualized as small images patches.

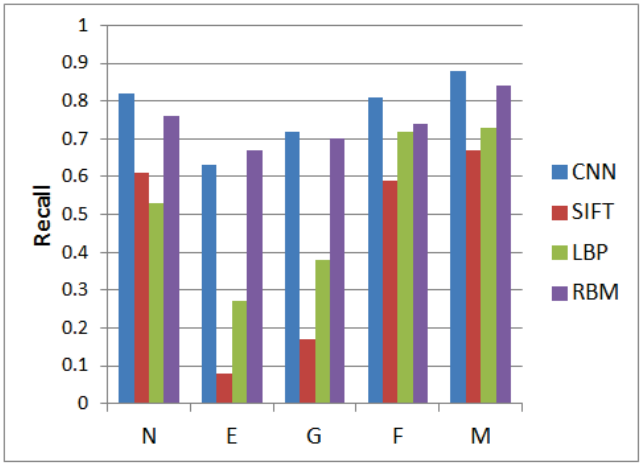
C. Classification Results

We compared our classification results with three other feature extraction approaches: (1) SIFT feature with keypoint located at the patch center; (2) rotation-invariant LBP feature with three resolutions; and (3) unsupervised feature learning using RBM [9]. All three compared approaches are coupled with SVM classification models. Note that our customized CNN method does not involve an additional classifier such as SVM, as the classification model is learned by the three-layer fully connected neural network layers. Using ANN for classification model learning has the distinct advantage that back-propagation algorithm can be applied to fine tune parameters in all layers to obtain a better final classification model. Classification results are evaluated using recall and precision.

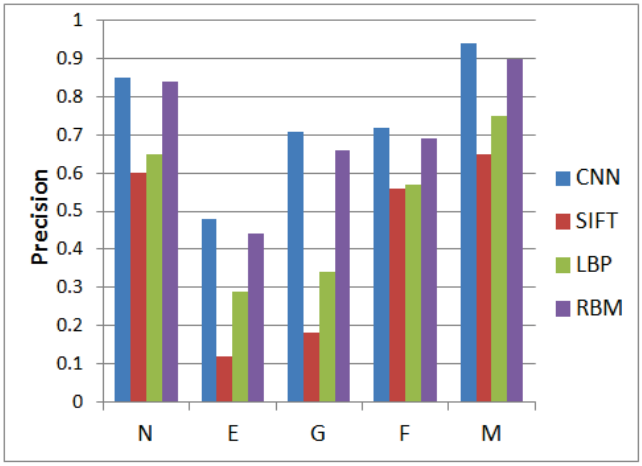
As shown in Fig.4, our customized CNN method achieved the best classification performance. The improvement over RBM demonstrates the advantage of the supervised feature learning, while in RBM the feature learning and classifier training are separated. Our method also exhibits large performance margin over the popular feature descriptors, i.e. SIFT and LBP, demonstrating the feasibility of having automatic feature learning for medical imaging. While we have not compared directly with approaches based on customized feature design specifically for the ILD images [2], [4], we suggest that the main attractiveness of our method is the customized adaptation of CNN, which is currently less studied in medical image computing.

IV. Conclusion

In this work, we have proposed a customized CNN architecture to classify HRCT lung image patches of ILD patterns. Our design with a single convolution layer learns the DCT-like patterns from training samples efficiently and yields good classification results. The results demonstrate that our design



(a)



(b)

Fig. 4. The classification results comparing our proposed customized CNN method with SIFT, LBP and RBM: (a) recall and (b) precision.

is capable of extracting discriminative features automatically without manual feature design, and achieving good benchmark performance. During our experiment we found that indistinct visual structure and limited size of training data were the major issues in adapting CNN to ILD image classification. However, with the properly designed network structure and applying techniques like intensive dropout and input distortion to suppress the over-fitting problem, we have addressed these problems effectively.

ACKNOWLEDGMENT

This work was supported in part by ARC grants.

REFERENCES

[1] Dahl, G.E., Sainath, T.N., Hinton, G.E.: Improving deep neural networks for lvcsr using rectified linear units and dropout. in Proc. ICASSP pp. 8609–8613 (2013)
[2] Depeursinge, A., Foncubierta-Rodriguez, A., de Ville, D.V., Muller, H.: Lung texture classification using locally-oriented riesz components. in MICCAI LNCS 6893, 231–238 (2011)
[3] Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.A., Muller, H.: Building a reference multimedia database for interstitial lung diseases. Comput. Med. Imaging Graph. 36(3), 227–238 (2012)
[4] Depeursinge, A., de Ville, D.V., Platon, A., Geissbuhler, A., Poletti, P.A., Muller, H.: Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames. IEEE Trans. Inf. Technol. Biomed. 16(4), 665–675 (2012)
[5] Hubel, D.H., Wiesel, T.N.: Receptive fields of single neurones in the cat's striate cortex. The Journal of physiology 148(3), 574–591 (1959)
[6] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. in Proc. NIPS pp. 1–9 (2012)
[7] Le Cun, Y., Jackel, L., Boser, B., Denker, J., Graf, H., Guyon, I., Henderson, D., Howard, R., Hubbard, W.: Handwritten digit recognition: Applications of neural network chips and automatic learning. IEEE Communications Magazine 27(11), 41–46 (1989)
[8] LeCun, Y., Bengio, Y.: Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks 3361 (1995)
[9] Li, Q., Cai, W., Feng, D.D.: Lung image patch classification with automatic feature learning. in Proc. EMBC pp. 6079–6082 (2013)
[10] Liu, M., Lu, L., Ye, X., Yu, S., Salganicoff, M.: Sparse classification for computer aided diagnosis using learned dictionaries. in MICCAI pp. 41–48 (2011)
[11] Liu, S., Cai, W., Song, Y., Pujol, S., Kikinis, R., Wen, L., Feng, D.: Localized sparse code gradient in alzheimer's disease staging. EMBC pp. 5398–5401 (2013)
[12] Nayak, N., Chang, H., Borowsky, A., Spellman, P., Parvin, B.: Classification of tumor histopathology via sparse feature learning. ISBI pp. 410–413 (2013)
[13] Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. in MICCAI LNCS 8150, 246–253 (2013)
[14] Sermanet, P., LeCun, Y.: Traffic sign recognition with multi-scale convolutional networks. in Proc. IJCNN pp. 2809–2813 (2011)
[15] Song, Y., Cai, W., Huang, H., Zhou, Y., Wang, Y., Feng, D.: Boosted multifold sparse representation with application to ILD classification. ISBI pp. 1023–1026 (2014)
[16] Song, Y., Cai, W., Huh, S., Chen, M., Kanade, T., Zhou, Y., Feng, D.: Discriminative data transform for image feature extraction and classification. MICCAI pp. 452–459 (2013)
[17] Song, Y., Cai, W., Wang, Y., Feng, D.: Location classification of lung nodules with optimized graph construction. ISBI pp. 1439–1442 (2012)
[18] Song, Y., Cai, W., Zhou, Y., Feng, D.: Feature-based image patch approximation for lung tissue classification. IEEE Trans. Medical Imaging 32(4), 797–808 (2013)
[19] Webb, W.R., Muller, N.L., Naidich, D.P.: High-resolution CT of the lung. Lippincott Williams Wilkins (2008)
[20] Xu, Y., Gao, X., Lin, S., Wong, D.W.K., Liu, J., Xu, D., Cheng, C., Cheung, C.Y., Wong, T.Y.: Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression. in MICCAI pp. 468–475 (2013)
[21] Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional neural networks. arXiv preprint arXiv:1311.2901 (2013)
[22] Zhang, F., Song, Y., Cai, W., Lee, M., Zhou, Y., Huang, H., Shan, S., Fulham, M., Feng, D.: Lung nodule classification with multi-level patch-based context analysis. IEEE Trans. Biomedical Engineering 61(4), 1155–1166 (2014)
[23] Zhang, F., Song, Y., Cai, W., Zhou, Y., Shan, S., Feng, D.: Context curves for classification of lung nodule images. DICTA pp. 1–7 (2013)
[24] Zhang, P., Wee, C., Niethammer, M., Shen, D., Yap, P.: Large deformation image classification using generalized locality-constrained linear coding. MICCAI pp. 292–299 (2013)
[25] Zhao, Q., Okada, K., Rosenbaum, K., Kehoe, L., Zand, D.J., Sze, R., Summar, M., Linguraru, M.G.: Digital facial dysmorphology for genetic screening: hierarchical constrained local model using ICA. Medical Image Analysis 18(5), 699–710 (2013)
[26] Zhou, L., Wang, L., Liu, L., Ogunbona, P., Shen, D.: Support vector machines for neuroimage analysis: interpretation from discrimination. Support Vector Machines Applications pp. 191–220 (2014)