ANATOMY-SPECIFIC CLASSIFICATION OF MEDICAL IMAGES USING DEEP CONVOLUTIONAL NETS

Holger R. Roth, Christopher T. Lee, Hoo-Chang Shin, Ari Seff, Lauren Kim, Jianhua Yao, Le Lu, Ronald M. Summers

Imaging Biomarkers and Computer-Aided Diagnosis Laboratory
Radiology and Imaging Sciences Department
National Institutes of Health Clinical Center
Bethesda, MD 20892, USA

ABSTRACT

Automated classification of human anatomy is an important prerequisite for many computer-aided diagnosis systems. The spatial complexity and variability of anatomy throughout the human body makes classification difficult. "Deep learning" methods such as convolutional networks (ConvNets) outperform other state-of-the-art methods in image classification tasks. In this work, we present a method for organ- or body-part-specific anatomical classification of medical images acquired using computed tomography (CT) with ConvNets. We train a ConvNet, using 4,298 separate axial 2D key-images to learn 5 anatomical classes. Key-images were mined from a hospital PACS archive, using a set of 1,675 patients. We show that a data augmentation approach can help to enrich the data set and improve classification performance. Using ConvNets and data augmentation, we achieve anatomy-specific classification error of 5.9 % and area-under-the-curve (AUC) values of an average of 0.998 in testing. We demonstrate that deep learning can be used to train very reliable and accurate classifiers that could initialize further computer-aided diagnosis.

Index Terms— Image Classification, Computed tomography (CT), Convolutional Networks, Deep Learning

1. INTRODUCTION

Medical image classification can be an important component of many computer aided detection (CADe) and diagnosis (CADx) systems. Achieving high accuracies for automated classification of anatomy is a challenging task, given the vast scope of anatomic variation. In this work, our aim is to automatically classify axial CT images into 5 anatomical classes (see Fig. 1). This aim is achieved by mining radiological reports that refer to key-images and associated DICOM image tags manually in order to establish a ground truth for training and testing. Using computer vision and medical image computing techniques, we were able to train the computer to replicate these classes with low error rates.

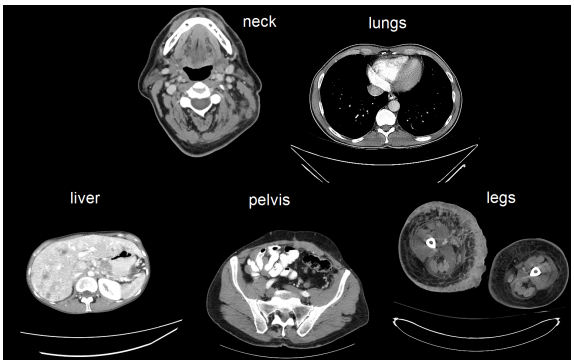

Fig. 1. Example key-images of 5 classes of anatomy in our data set: neck, lungs, liver, pelvis and legs.

2. METHOD

Recently, the availability of large annotated training sets and the accessibility of affordable parallel computing resources via GPUs have made it feasible to train "deep" convolutional networks (ConvNets). ConvNets have popularized the topic of "deep learning" in computer vision research [1]. Through the use of ConvNets, not only have great advances been made in the classification of natural images [2], but substantial advancements have also been made in biomedical applications, such as digital pathology [3]. Additionally, recent work has shown how the implementation of ConvNets can substantially improve the performance of state-of-the-art CADe systems [4, 5, 6, 7].

2.1. Convolutional networks

In this work, we apply ConvNets to build an anatomy-specific classifier for CT images. ConvNets are named for their convolutional filters which are used to compute image features

This work was supported by the Intramural Research Program of the NIH Clinical Center. Contact: holger.roth@nih.gov or rms@nih.gov

for classification. In this work, we use 5 cascaded layers of convolutional filters. All convolutional filter kernel elements are trained from the data in a supervised fashion. This has major advantages over more traditional CAD approaches that use hand-crafted features, designed from human experience. This means that ConvNets have a better chance of capturing the "essence" of the imaging data set used for training than when using hand-crafted features [1]. Examples of trained filters of the first convolutional layer can be seen in Fig. 2. These first-layer filters capture low spatial frequency signals. In contrast, a mixed set of low and high frequency patterns exists in the first convolutional layer shown in [5, 6]. This indicates that the essential information of this task of classifying holistic slice-based body regions lies in the low frequency spatial intensity contrasts. These automatically learned low frequency filters need no tuning by hand, which is different from using intensity histograms, e.g. [8, 9]. In-between

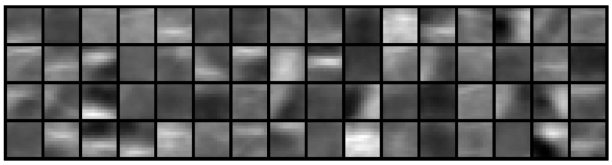

Fig. 2. The first layer of learned convolutional kernels of a ConvNet trained on medical CT images.

convolutional layers, the ConvNet performs *max-pooling* operations in order to summarize feature responses across non-overlapping neighboring pixels (see Fig. 3). This allows the ConvNet to learn features that are invariant to spatial variations of objects in the images. Feature responses after the 5th convolutional layer feed into a *fully-connected* neural network. This network learns how to interpret the feature responses and make anatomy-specific classifications. Our ConvNet uses a final *softmax* layer which provides a probability for each object class (see Fig. 3). In order to avoid overfitting, the fully-connected layers are constrained, using the *"DropOut"* method [10]. *DropOut* behaves as a regularizer when training the ConvNet by preventing co-adaptation of units in the neural network. We use an open-source implementation (*cuda-convnet2*[1]) by Krizhevsky et al. [2, 11] which efficiently trains the ConvNet, using GPU acceleration. Further speed-ups are achieved using rectified linear units as neuron activation function instead of the traditional neuron model $f(x) = \tanh(x)$ or $f(x) = (1 + e^{-x})^{-1}$ in both training and evaluation [2].

2.2. Data mining of key-images

We retrieve medical images (many related to liver disease) from the Picture Archiving and Communication System (PACS) of the Clinical Center of the National Institutes of

[1] https://code.google.com/p/cuda-convnet2

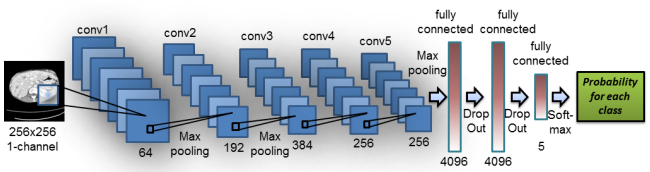

Fig. 3. ConvNet applied to an axial CT image. The number of convolutional filters and neural network connections for each layer are as shown.

Health by searching for a set of keywords in the radiological reports. Then, each image is assigned a ground truth label based on the *'StudyDescription'* and *'BodyPartExamined'* DICOM tags (manually corrected if necessary). This results in 5 classes of images as shown in Fig. 1. Images which show anatomies of multiple classes at once are duplicated and each image copy is assigned one of the class labels. This case commonly occurs at the transition region between lung and liver. Our ConvNet assigns equal probabilities for each class in these regions.

2.3. Data augmentation

We enrich our data set by applying spatial deformations to each image, using random translation, rotations and non-rigid deformations. Each non-rigid training deformation t is computed by fitting a thin-plate-spline (TPS) to a regular grid of 2D control points $\{\omega_i; i = 1, 2, , K\}$. These control points can be randomly transformed at the 2D slice level and a deformed image can be generated using a radial basis function $\phi(r)$:

$$t(x) = \sum_{i=1}^{K} c_i \phi\left(\|x - \omega_i\|\right). \quad (1)$$

We use $\phi(r) = r^2 log(r)$ which is commonly applied for TPS. A typical TPS deformation field and deformed variations of an example image grid are shown in Fig. 4. The variation of translation t, rotation r and non-rigid deformations d are a useful way to increase the variety and sample space of available training data, resulting in $N_{\text{aug.}} = N \times N_t \times N_r \times N_d$ variations of the imaging data. The maximum amounts of translation, rotation and non-rigid deformation are chosen such that the resulting deformations resemble plausible physical variations of the medical images. This approach is commonly referred to as data augmentation and can help avoid overfitting [2]. Our set of $N_{\text{aug.}}$ axial images are then rescaled to 256×256 and used to train a ConvNet with a standard architecture for multi-class image classification (as described in Sec. 2.1).

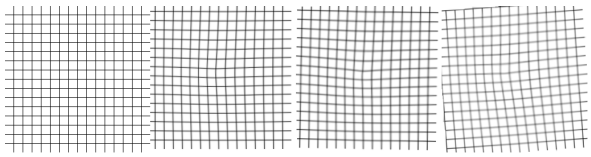

Fig. 4. Data augmentation using varying random transformations, rotations and non-rigid deformations, using thin-plate-spline (TPS) interpolations on an example image grid.

3. RESULTS

3.1. Key-image data set

We use 80 % of our total dataset for training a multi-class ConvNet as described in Sec. 2.1. and reserve 20 % for testing purposes. Our data augmentation step (see Sec 2.3) increases the amount of training and testing data drastically, as shown in Table 3.1. The number of deformations for each anatomical class is chosen so that the resulting augmented images build a more balanced and enriched data set. We use $N_t = 2$ and $N_r = 2$ while adjusting N_d for each class to achieve a balanced data set. Table 3.1 further shows that data augmentation helps to reduce classification errors from 9.6 % to 5.9 % in testing and furthermore improve the average area-under-the-curve (AUC) values from 0.994 to 0.998 using receiver-operating-characteristic (ROC) analysis. Confusion matrices shown in Fig. 5 show a clear reduction of misclassification after using data augmentation when testing on the original test set. We further illustrate the feature space of our trained ConvNet using t-SNE [12, 13] in Fig. 6. A clear separation of most classes can be observed. An overlapping cluster can be seen at the interface between the lungs and liver images. This is caused by key-images that show both lungs and livers being near the diaphragm region.

Table 1. Image data set before[1] and after[2] data augmentation. An improvement of both error rate and AUC values can be achieved by using data augmentation.

Organ	#[1]	#[2]	AUC[1]	AUC[2]
leg	477	24,804	1.000	1.000
pelvis	104	22,048	0.996	1.000
liver	2,684	32,208	0.994	0.999
lung	590	25,960	0.981	0.999
neck	443	23,036	0.999	1.000
Sum/Mean AUC	*4,298*	*12,8056*	**0.994**	**0.998**
Error	**9.6%**	**5.9%**		

3.2. Full torso CT volume

For qualitative evaluation, we also apply our trained ConvNet classifier on a full torso CT examination on a slice-by-slice basis (dimensions of $[512, 512, 652]$ and $[0.98, 0.98, 1.5]$ mm voxel spacing). The resulting anatomy-specific probabilities for each slice are plotted as profiles next to the coronal slice of the CT volume in Fig. 7. Note how the interface between the lungs and liver at the level of the diaphragm is captured by roughly equal probabilities of the ConvNet. This classification result is achieved in less than 1 minute on a modern desktop computer and GPU card (Dell Precision T7500, 24GB RAM, NVIDIA Titan Z).

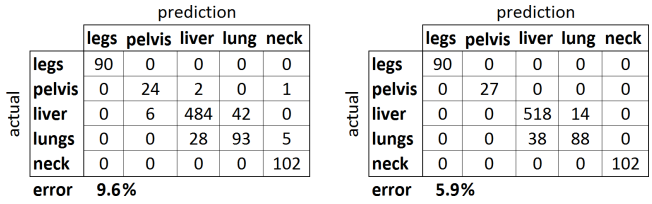

Fig. 5. Confusion matrices on the original test images before[1] and after[2] data augmentation.

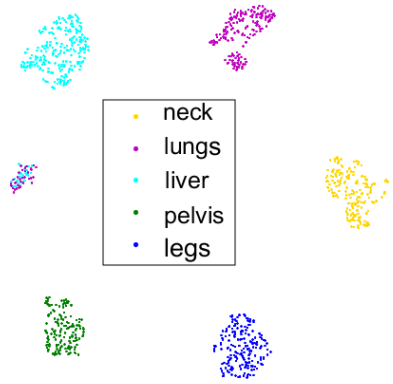

Fig. 6. 2D embedding of ConvNet features using t-SNE on a subset of test images. Each dot represents a key-image in feature space. The color-coding is based on the ground truth label for each key-image.

4. DISCUSSION

This work demonstrates how deep ConvNets can be applied to effective anatomy-specific classification of medical images. Similar motives to ours are explored in content-based image retrieval methods [14]. However, association based on clinical reports and image scans can be very loose. This makes retrieval based on clinical reports difficult. In this paper, we focus on manually labeled key-images that allow us to train a anatomy-specific classifier. Other related work includes the *ImageCLEF* medical image annotation tasks of 2005-2007. However, these tasks used highly subsampled 2D version of medical images (32×32 pixels) [15]. Methods applied to the *ImageCLEF* tasks included using local image descriptors and intensity histograms in a bag-of-features approach [16]. We

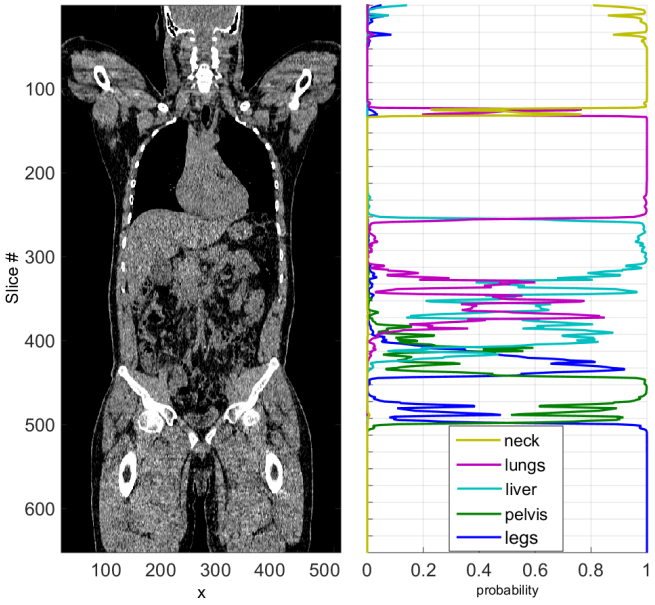

Fig. 7. Organ-specific probabilities for a whole-body CT scan.

concentrate on classifying images much closer to their original 512×512 resolution, namely rescaled to 256×256. We show that ConvNets can model this higher detail in the images and generalize well to large variations found in medical imaging data with promising quantitative and qualitative results. Some axial slices in the lower abdomen had erroneously high probabilities for lung or legs. Here, it could be beneficial to introduce an additional class of 'lower abdomen'. Our method could be easily extended to include further augmentation such as image scales in order to model variations in patient sizes. This type of anatomy classifier could be employed as an initialization step for further and more detailed analysis, such as disease and organ specific computer-aided detection and/or diagnosis.

5. REFERENCES

[1] J. N., "Computer science: The learning machines," *Nature*, vol. 505(7482), pp. 146–8, 2014.

[2] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *NIPS*, pp. 1097–1105, 2012.

[3] D. Cireşan, A. Giusti, L. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," *MICCAI*, pp. 411–418, 2013.

[4] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," *MICCAI*, pp. 246–253, 2013.

[5] H. Roth, L. Lu, A. Seff, K. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. Summers, "A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations," in *MICCAI*, pp. 520–527, Springer, 2014.

[6] H. Roth, J. Yao, L. Lu, J. Stieger, J. Burns, and R. Summers, "Detection of sclerotic spine metastases via random aggregation of deep convolutional neural network classifications," *MICCAI Spine Imaging Workshop, arXiv preprint arXiv:1407.5976*, 2014.

[7] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," *ICARCV*, 2014.

[8] J. Feulner, S. K. Zhou, S. Seifert, A. Cavallaro, J. Hornegger, and D. Comaniciu, "Estimating the body portion of ct volumes by matching histograms of visual words," in *SPIE Med. Imag.*, pp. 72591V–72591V, 2009.

[9] V. Dicken, B. Lindow, L. Bornemann, J. Drexl, A. Nikoubashman, and H. Peitgen, "Rapid image recognition of body parts scanned in computed tomography datasets," *IJCARS*, vol. 5, no. 5, pp. 527–535, 2010.

[10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[11] A. Krizhevsky, "One weird trick for parallelizing convolutional neural networks," *arXiv preprint arXiv:1404.5997*, 2014.

[12] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *JMLR*, vol. 9, no. 2579-2605, p. 85, 2008.

[13] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," *arXiv preprint arXiv:1310.1531*, 2013.

[14] C. B. Akgül, D. Rubin, S. Napel, C. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: current status and future directions," *Digital Img.*, pp. 208–222, 2011.

[15] H. Müller, T. Deselaers, T. Deserno, P. Clough, E. Kim, and W. Hersh, "Overview of the ImageCLEFmed 2006 medical retrieval and medical annotation tasks," *LNCS*, vol. 4730, pp. 595–608, 2007.

[16] T. Deselaers and H. Ney, "Deformations, patches, and discriminative models for automatic annotation of medical radiographs," *Pattern Recogn. Lett.*, vol. 29, no. 15, pp. 2003–2010, 2008.