

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/304360689>

Discriminative feature extraction from X-ray images using deep convolutional neural networks

Conference Paper · March 2016

DOI: 10.1109/ICASSP.2016.7471809

CITATIONS

8

READS

274

3 authors:



Mettu Srinivas

Indian Institute of Technology Hyderabad

16 PUBLICATIONS 91 CITATIONS

SEE PROFILE



Debaditya Roy

Nihon University

14 PUBLICATIONS 47 CITATIONS

SEE PROFILE



Krishna Mohan Chalavadi

Indian Institute of Technology Hyderabad

57 PUBLICATIONS 351 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Scalable Methods for Visual Big Data Analytics [View project](#)



Unsupervised feature extraction for video understanding [View project](#)

Discriminative Feature Extraction from X-ray Images using Deep Convolutional Neural Networks

M. Srinivas, Debaditya Roy, C. Krishna Mohan
Visual learning and Intelligence (VIGIL) group
Dept. of Computer Science and Engineering
Indian Institute of Technology Hyderabad
Email: { cs10p002, cs13p1001, ckm } @iith.ac.in

Abstract—Feature extraction is one of the most important phases of medical image classification which requires extensive domain knowledge. Convolutional Neural Networks (CNN) have been successfully used for feature extraction in images from different domains involving a lot of classes. In this paper, CNNs are exploited to extract a hierarchical and discriminative representation of X-ray images. This representation is then used for classification of the X-ray images as various parts of the body. Visualization of the feature maps in the hidden layers show that features learnt by the CNN resemble the essential features which help discern the discrimination among different body parts. A comparison on the standard IRMA X-ray image dataset demonstrates that the CNNs easily outperform classifiers with hand-engineered features.

Keywords—Convolutional Neural Networks (CNN), X-ray image, Feature Extraction

I. INTRODUCTION

X-ray images constitute one of the most important categories of medical imaging and are extensively used in clinical decision making. This has opened the field of X-ray image classification based on modality and body parts for efficient cataloging and ease of search. On one hand, domain knowledge is expensive to obtain on massive X-ray image databases as it requires extensive effort to catalog and index the images manually. On the other hand, same features cannot be used for different types of classification tasks even on the same image database. However, efficient classification of X-ray images still relies on region of interest detection and careful feature extraction requiring extensive domain knowledge. CNN provides the flexibility of extracting intrinsic and discriminative features from X-ray images which are most suitable for classification. Hence, to develop an application-agnostic classifier for medical images, CNNs are a natural choice.

X-ray imaging has vastly improved due to the advent of digital X-ray imaging techniques leading to more accurate and clear images. One of the earliest attempts was by Avni et al. [1] who proposed an X-ray image categorization and retrieval method using patch-based visual word representations. Local patches extracted from an image were clustered and converted to visual words which were used to form a bag-of-features for each image. The bag-of-features extracted from the images of each class was different from others and a kernel-support vector machine (SVM) was able to classify the various categories. An attempt to combine the benefits of local features which were patch based with global shape features was introduced by Fulidong et al. [2]. Edge density histogram

descriptor (EDHD) combines edge information extracted from the whole image with edge density of sub-images which was then classified using a linear SVM. Automatic learning of local binary patterns (LBP) feature was proposed in [3]. An LBP operator was used to encode the image pixels with binary labels by thresholding neighborhood of each pixel with the center value. Classification was done using a maximum margin SVM. In [4], an ensemble of features was used for medical image indexing. Adaptive learning based heartbeat classification method is proposed in [5]. Time and frequency domains based new feature extraction method and furthermore, feature normalization techniques to reduce inter-patient and intra-patient variations in heartbeat cycles are proposed.

From literature, it is evident that a careful combination of both classifier and features is required for decent classification performance. Various authors have devised features which exploit different aspects of X-ray images like spatial frequency coefficients, spatial correlation etc. which have their own strengths but no common underlying similarity. A CNN can efficiently produce a set of discriminative features without any expert to demarcate a region of interest (ROI). Also, a few examples are also sufficient for training a CNN [6], which makes it more useful for most medical datasets as they suffer from data imbalance. In [7], wavelet features extracted from an image provide discrimination useful for classification of different sensors (modalities) based medical images. And with the help of On-Line dictionary learning and sparse representation methods are used for classification approach. Dictionary learning based clustering method is proposed in [8]. Each image portioned into concentric circular regions around the center, and consider the mean and variance of pixel intensities in each region as components in the feature vector. And dictionary learning based clustering method developed for image retrieval on unlabeled data.

The rest of the paper is organized as follows. Section 2 introduces CNN and its prevalence in medical imaging. Section 3 presents the proposed CNN architecture and section 4 provides the experimental details with results on X-ray image classification. Finally, in section 5 we discuss the conclusions drawn from this work.

II. CNN IN MEDICAL IMAGING: PRIOR ART

CNNs were first introduced by Fukushima et al. [9] and the architecture was later improved by LeCun et al. [10]. CNNs have been shown to be close to human recognition [11] on

the ImageNet Large Scale Visual Recognition Challenge. In medical imaging, CNNs were introduced by Sahiner et al. [12] for classification of mass and normal breast tissue. Each mammogram image was marked with an ROI by an expert and the same region was then divided into small patches. These patches were then subjected to averaging and sub-sampling in one case and texture based feature extraction in the other giving the features which were used to train a CNN. Suitability of CNNs for mammogram image classification was established in this paper but manual annotation of ROI was still key to the process. In a review by Jiang et al. [13] regarding the applicability of neural networks for medical image analysis, the authors note that ROI guides the use of CNN. Further, in most of the approaches, binary classification is desired which makes the job of extracting features simpler.

CNNs have also been used for pre-processing, especially for bone segmentation in X-ray images by Cernazanu-Glavan et al. [14]. CNNs were used to classify bone and non-bone areas in an X-ray image to help speed up extraction of features. Prasoon et al. [15] used three 2D-CNNs to map the xy , yz and zx planes of a 3D Magnetic Resonance Imaging (MRI) scan in order to segment the tibial cartilage in the knee. This approach worked better than multi-scale 3D features.

III. PROPOSED ARCHITECTURE

A typical CNN is different from a multi-layer neural network as it retains the spatial correlation of the image which can be very useful to extract meaningful features. The three main types of layers in a CNN are : convolution, max-pooling/sub-sampling and fully connected. Typically, the convolution and sub-sampling/max-pooling layers are interleaved to obtain spatial and configuration invariance [16].

A. Convolution Layers

At every convolution layer, the previous layers feature maps convolve with kernels and are then passed through the activation function to form the output feature maps. Further, each output map can also be formed as a result of the combination of many input maps. Mathematically, we can write that as:

$$\mathbf{x}_j^l = f(\sum_{i \in M_j} \mathbf{x}_i^{l-1} * k_{ij}^l + b_j^l),$$

where $f(\cdot)$ is a non-linear function like softmax, \mathbf{x}_j^l represents the j^{th} output feature map of the l^{th} layer, \mathbf{x}_i^{l-1} is the i^{th} input map of the $(l-1)^{th}$ layer, M_j represents a selection of input maps, k_{ij}^l is the kernel for input map i and output map k in the l^{th} layer and b_j^l is the additive bias associated with j^{th} output map.

B. Sub-sampling/Max-pooling Layers

A sub-sampling or max-pooling layer produces the down sampled versions of the input maps in different ways. Average activation over the neighbourhood patch is considered in sub-sampling whereas max-pooling only consider the highest activation value in the input map. Formally,

$$\mathbf{x}_j^l = f(\beta_j^l \text{down}(\mathbf{x}_i^{l-1} * k_{ij}^l) + b_j^l),$$

where β_j^l is the multiplicative bias of each output feature map j to scale the output back to the original range, $\text{down}(\cdot)$ can either be replaced by $\text{max}(\cdot)$ or $\text{avg}(\cdot)$ over a $n \times n$ window effectively reducing the size of the input map by n times in each dimension.

After the required number of convolution and sub sampling layers, the output is flattened as a vector for the fully connected layer, where generally *softmax* is used to classify the obtained features into the corresponding classes. Also, support vector machines (SVM) can be used to the same effect.

Careful tuning of kernel size and number of hidden layers is essential to achieve good classification performance. The proposed architecture as can be seen in fig 1 involves four convolutional layers with kernel sizes 9×9 , 5×5 , 5×5 and 2×2 , respectively, and it obtains impressive classification results. Sub-sampling was done after each convolution layer and the window size chosen was 2×2 . The final layer or the fully connected layer was used to classify the extracted features into one of the 12 classes.

C. Patch vs Entire image

The user community on CNN generally employs a patch based approach for larger images where the various patches taken from the same image are supplied to a CNN and feature maps corresponding to those patches are learnt and then combined in subsequent layers into more meaningful feature maps. Li et al. [17] use a 32×32 patch from a high-resolution computed tomography (HRCT) images of the lungs to classify various interstitial lung diseases (ILD). Since the task is to identify ILD which occurs locally in HRCT images, the patch based approach is plausible. However, when applied to classification of X-ray images, it fails miserably as the learnt feature maps convey no discriminative information for distinguishing various body parts which can only be achieved in a global setting. Hence, the entire image of size 120×120 is directly employed for classification.

IV. RESULTS AND DISCUSSION

Over the years, various works like [18], [19], [20], [21], [22], [23] have considered 1617 images in 6 classes, 9100 images in 40 classes, 6231 images, 5000 images in 20 classes, 1500 images in 17 classes and 4937 images in 28 classes, respectively, from the IRMA dataset for classification. Hence, it is very difficult to compare the proposed method with the previous works directly. Instead, the classification prowess of CNNs is compared with various classifiers used in the previous works which have been proven to be effective. For the experiments, we used 826 training and 300 testing images spanning 12 classes of human organs. The performance of the proposed system is evaluated by measuring classification accuracy (Acc) Accuracy defined as follows:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FN+FP)},$$

where TP = true positive, FN = false negative, FP = false positive, and TN = true negative.

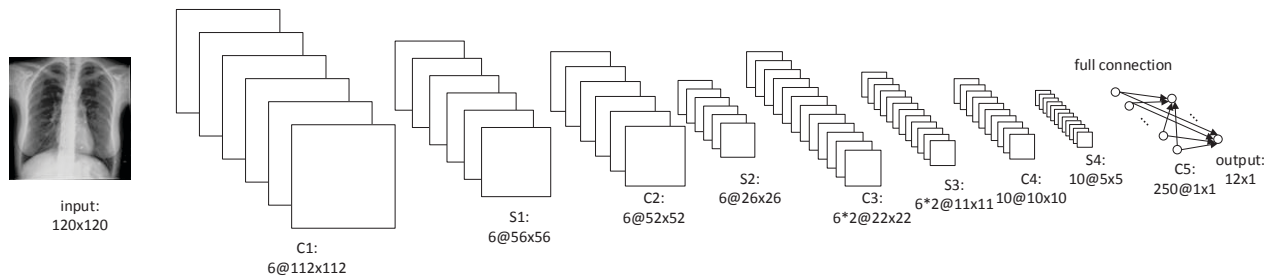


Fig. 1: Proposed CNN Architecture

A. Comparison with other classifiers

A comparison of various features used for content based image retrieval systems for X-ray images is presented in [24], in which wavelet and texture based features have been shown to be effective. Hence, for the purpose of comparison with other classifiers, we chose these features. In table ??, wavelet, texture and HoG features were used with an SVM classifier and the observed classification performance was 67%, 85% and 93.6%, respectively. On the same dataset, the proposed method gives classification results of 97.6%. Further, the proposed CNN architecture is evaluated on various body parts whose results are shown in terms of a confusion matrix in figure 3. It can be seen that the CNN is able to identify most of the examples correctly except in few cases. This can be attributed to the fact that X-ray images of certain body parts in some orientations resemble certain other body parts. In such cases, discrimination can only be done faithfully by a medical expert.

In figure 4, it can be observed that training of our CNN classifier slowly converges and almost reaches negligible error on the validation examples. This shows good generalization capability of the network. The decay in error rate is exponential to begin with and asymptotically depends on the iterations. Hence, with reasonable bounds on accuracy the network can be trained well.

B. Visualizing the CNN outputs

The proposed deep CNN is trained using the medal-master toolbox [25] and the outputs of the first two layers of the CNN for some images in few of the classes are displayed in figure 2. It can be clearly observed that the two layers learn almost complementary information. Further inspection reveals that the general structure of an image is captured in the first layer whereas smoother image regions are learnt in the next layer. This is analogous to other CNN where edges are learnt in the first layer and more and more coherent features are learnt in the next few layers. Hence, the proposed approach indeed learns hierarchical representations from X-ray images even if the entire image is used instead of patches for training.

V. CONCLUSION

In this paper, CNNs were explored to extract representative and discriminative features from X-ray images for classification into various body parts. The inherent capability of CNN to capture the structure of the image in its feature maps was

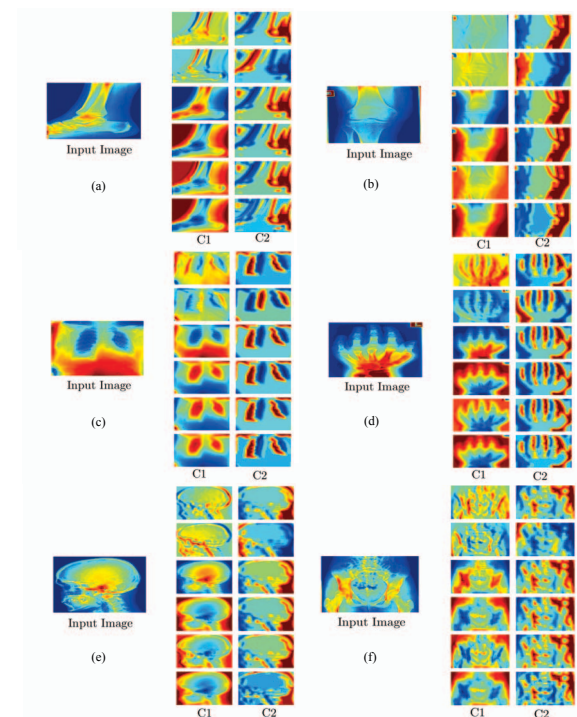


Fig. 2: Feature maps of first and second convolutional layers for images of various classes: (a) ankle, (b) knee, (c) chest, (d) palm, (e) skull, (f) pelvis.

portrayed. CNNs easily outperform hand-engineered features coupled with a classifier as was shown in the experimental evaluation. Since this work is purely empirical in terms of estimating the number of layers and kernel size for each layers, it can be extended to standardize the training of CNNs on all kinds of X-ray images. Further, suitability of CNN can be investigated for different kinds of medical images like ultrasound, magnetic resonance imaging (MRI) etc.

ACKNOWLEDGEMENT

The authors would like to thank Dustin Stansbury for making the code for CNN available on github. We would

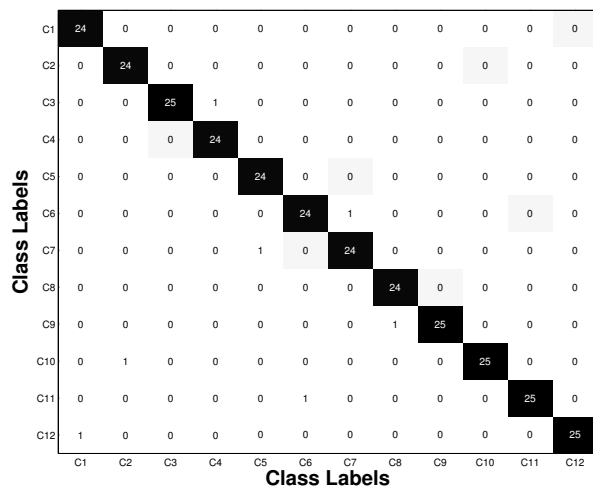


Fig. 3: Confusion matrix on the IRMA dataset for the proposed method.

Table I: Performance comparison of the proposed method with different classifiers on the IRMA dataset.

Classes	Wavelet features	Texture features	HoG features	CNNs
C1	96	76	100	96
C2	96	64	84	96
C3	100	56	100	100
C4	100	80	96	96
C5	96	16	100	96
C6	100	68	84	96
C7	92	28	88	96
C8	92	80	92	96
C9	0	84	100	100
C10	48	84	92	100
C11	100	20	100	100
C12	100	84	92	100
Avg. (%)	85	67	93.6	97.6

also like to thank Dr. T.M. Deserno, Department of Medical Informatics, RWTH Aachen, Germany for making the original IRMA Database available for research purposes.

REFERENCES

- [1] U. Avni et al., X-ray image categorization and retrieval using patch-based visual words representation. In Proc. IEEE Int. Symp. Biomed. Image : 1(2): 350-353, 2009.
- [2] Fu. Li-dong, Z. Yi-fei. Medical image retrieval and classification based on morphological shape feature. In Proc. IEEE Int. Conf. Intelligent Networks and Intelligent systems : 116-119, 2010.
- [3] M. R. Zare, M. Awedh, A. Mueen, W. C. Seng. Medical scheme based classification of medical X-ray images. In Proc. IEEE Int. Conf. Computational Intelligence, Modeling and Simulation : 253-258, 2011.
- [4] M. Srinivas, C. Krishna Mohan. Medical Image Indexing and Retrieval Using Multiple Features. In Proceedings of CIIT conference 2013; Elsevier.
- [5] M. Srinivas, Tony Basil, C. Krishna Mohan. "Adaptive learning based heartbeat classification." Bio-medical materials and engineering (2015), Vol.26.1-2 : 49-55.
- [6] Raimar Wagner, Markus Thom, Roland Schweiger, Gunther Palm, and Albrecht Rothermel. Learning convolutional neural networks from few samples. In *Neural Networks (IJCNN), The 2013 International Joint Conference on*, pages 1-7. IEEE, 2013.
- [7] Srinivas, M and Mohan, C Krishna. Medical images modality classification using multi-scale dictionary learning. 19th International

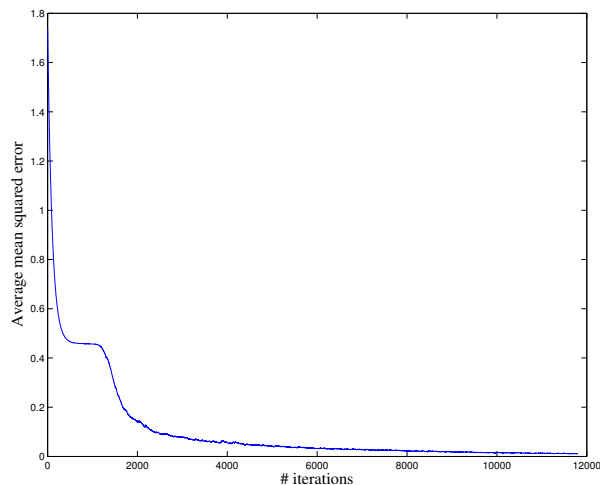


Fig. 4: Average mean squared error vs. number of iterations on the validation set for IRMA dataset. One iteration comprises of presenting a mini-batch (14 images) to the network

- Conference on Digital Signal Processing (DSP), 2014, pp. 621-625, 2014.
- [8] M. Srinivas, R. Ramu Naidu, C. S. Sastry, Mohan, C Krishna. Content Based Medical Image Retrieval Using Dictionary Learning. *Neurocomputing* (2015), Vol.168, : 880-895.
- [9] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193-202, 1980.
- [10] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278-2324, 1998.
- [11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097-1105, 2012.
- [12] Berkman Sahiner, Heang-Ping Chan, Nicholas Petrick, Datong Wei, Mark A Helvie, Dorit D Adler, and Mitchell M Goodsitt. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *Medical Imaging, IEEE Transactions on*, 15(5):598-610, 1996.
- [13] Jianmin Jiang, P Trundle, and Jinchang Ren. Medical image analysis with artificial neural networks. *Computerized Medical Imaging and Graphics*, 34(8):617-631, 2010.
- [14] C. Cernazanu-Glavan, and S. Holban. "Segmentation of bone structure in X-ray images using convolutional neural network." *Adv. Electr. Comput. Eng* 13.1 87-94, 2013.
- [15] Adhish Prasoon, Kersten Petersen, Christian Igel, François Lauze, Erik Dam, and Mads Nielsen. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2013*, pages 246-253. Springer, 2013.
- [16] Jake Bouvrie. Notes on convolutional neural networks. 2006.
- [17] Qing Li, Weidong Cai, Xiaogang Wang, Yun Zhou, Dagan Feng, Mei Chen. Medical Image Classification with Convolutional Neural Network, The 13th International Conference on Control, Automation, Robotics and Vision (ICARCV 2014), December 10-12, 2014.
- [18] D. Keysers, J. Dahmen, H. Ney, B. B. Wein, T. M. Lehmann. Statistical frame work for model based image retrieval in medical applications. *J Electronic Image* : 12(1): 59-68, 2003.
- [19] H. Pourghassem, H. Ghassemian. Content based medical image classification using a new hierarchical merging scheme. *Computerized Medical Imaging and Graphics* : 32(8), 651-661, 2008.

- [20] T. M. Lehmann, M. O. Guld, T. Deselaers, D. Keysers, H. Schubert, K. Spitzer, H. Ney, B. B. Wein. Automatic categorization of medical images for content-based retrieval and data mining. *Comput.Med.Imag. Graph* : (29): 143-155, 2005.
- [21] M. M. Rahman, P. Bhattacharya, B. C. Desai. A frame work for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. *IEEE Trans. Inf Technology Biomed* : 11(1): 58-69, 2007.
- [22] H. Greenspan, A. T. Pinhas. Medical image categorization and retrieval for PACS using the GMM-KL frame work. *IEEE Trans. Inf Technology Biomed* : 11(2): 190-202, 2007.
- [23] N. J. Fesharaki, H. Pourghassem. Medical X-ray Images Classification Based on Shape Features and Bayesian Rule. In *Proc. IEEE Int. Conf. Computational Intelligence and Communication Networks* : 369-373, 2012.
- [24] Ceyhun Burak Akgl, Daniel L Rubin, Sandy Napel, Christopher F Beaulieu, Hayit Greenspan, and Burak Acar1. Content-Based Image Retrieval in Radiology: Current Status and Future Directions. *Journal of Digital Imaging*, 24(2):208–222, 2011.
- [25] Dustin Stansbury. Medal-master toolbox. <https://github.com/dustinstansbury/medal>.