Diagnosis of Degenerative Intervertebral Disc Disease with Deep Networks and SVM

Ayse Betul Oktay[1(✉)] and Yusuf Sinan Akgul[2]

[1] Department of Computer Engineering, Istanbul Medeniyet University, 34700 Istanbul, Turkey
abetul.oktay@medeniyet.edu.tr

[2] GTU Vision Lab, Gebze Technical University, Gebze, Kocaeli, Turkey
akgul@gtu.edu.tr

Abstract. Computer aided diagnosis of degenerative intervertebral disc disease is a challenging task which has been targeted many times by computer vision and image processing community. This paper proposes a deep network approach for the diagnosis of degenerative intervertebral disc disease. Different from the classical deep networks, our system uses non-linear filters between the network layers that introduce domain dependent information into the network training for a faster training with lesser amount of data. The proposed system takes advantage of the unsupervised feature extraction with deep networks while requiring only a small amount of training data, which is a major problem for medical image analysis where obtaining large amounts of patient data is very difficult. The method is validated on a dataset containing 102 lumbar MR images. State-of-the-art hand-crafted feature extraction algorithms are compared with the unsupervisedly learned features and the proposed method outperforms the hand-crafted features.

Keywords: Degenerative disc disease · Auto encoders · Deep network

1 Introduction

Low Back Pain (LBP) is the most common pain type with 27 % and it is the leading cause of activity limitation in USA under the age of 45 [7]. LBP is strongly associated with degenerative disc disease (DDD) [6]. Computer Aided Diagnosis (CAD) of DDD from MR images (Fig. 1) is crucial for many reasons. First, the inter-variability and intra-variability between the radiologists are high [12] and these variabilities affect diagnosis and treatment processes. A CAD system may reduce these variabilities. Second, the computer-based evaluation of an MRI sequence would help the radiologists in decreasing the costs and speeding up the evaluation process. In the literature, many machine learning based approaches with hand-crafted features have been proposed for CAD of various intervertebral disc diseases from MR images [1,4,5,9].

In recent years, deep networks have been widely used in many fields and they produce state-of-the-art results [3,10]. However, deep learning of medical

T. Czachórski et al. (Eds.): ISCIS 2016, CCIS 659, pp. 253–261, 2016.
DOI: 10.1007/978-3-319-47217-1_27

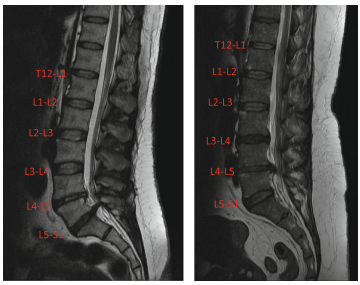


Fig. 1. Two MRI images that include the lumber region. The disc labels are shown on the images. The left image shows the discs L4-L5 and L5-S1. In the right image L3-L4 and L4-L5 discs are diagnosed as having DIDD

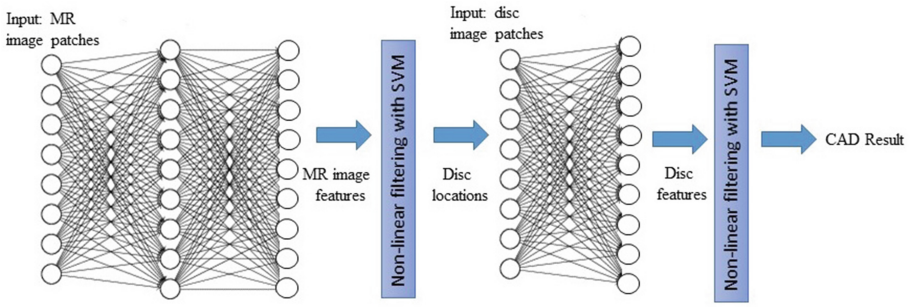


Fig. 2. The architecture of the system.

images has some domain-specific challenges. First, scaling the deep network for high dimensional medical images is mostly computationally intractable because of the large number of hidden neurons, often resulting in millions of parameters. Medical images have generally high resolution and the training needs high number of nodes. In addition, the large-scale data for training (even unlabeled) is not always available especially for many medical tasks where it is hard to gather data because of ethical issues. Furthermore, training data should involve many samples for different cases for CAD applications.

In this paper, we propose a novel deep learning architecture (Fig. 2) with non-linear filters that eliminates the requirement of large numbers of training data, network layers, and nodes. Instead of learning disc features with a traditional deep learning architecture, we propose to use non-linear filters together with auto-encoders [11]. The irrelevant input data is filtered with non-linear filters via SVM and only relevant data is fed to the succeeding layers. In this way, we restrict the upper layer to learn only the data that we consider valuable, which is very useful in reducing the training data size. Therefore, while the disc representations are learned with auto-encoders from the MR image patches, the non-linear filters reduce the domain of interest. Thus, with the first level

non-linear filters the system focus on the discs from the whole MR image where the second level non-linear filters consider the disc representations for the diagnosis of DDD.

The method is tested and validated on a dataset containing 102 MR images. We also implemented the state-of-the-art features used in the methods of [1,2,9] and compared them with the features learned with auto encoders.

2 Unsupervised Feature Learning with Auto-encoders

An auto-encoder is a symmetrical neural network that aims to minimize the reconstruction error between the input and output data to learn the features. Let $X = \{x_1, x_2, ..., x_m\}$ be the image input for a single hidden layered auto-encoder where m is the input size. The output nodes are the same as the input nodes, thus the auto-encoder learns a nonlinear approximation of the identity function for estimating the output $\hat{X} = \{\hat{x_1}, \hat{x_2}, ..., \hat{x_m}\}$. Let k be the size of the nodes in the hidden layer and $W^{(1)} = \{w_{11}^{(1)}, w_{12}^{(1)}, ..., w_{km}^{(1)}\}$ be the weights where $w_{km}^{(1)}$ is the weight between input node m to hidden node k at hidden layer 1. The value of a hidden layer node is calculated by

$$z_i = b_i^{(1)} + \sum_{j=1}^{m} w_{ij}^{(1)} x_j, \tag{1}$$

where $b_i^{(1)}$ is the bias term for the node i at hidden layer 1. Each hidden node outputs a nonlinear activation function $a = f(z_i)$. The output layer $\hat{X}$ is constructed using the activations a as input and decoding bias and weights similar to Eq. 1. Features are learned by minimizing the reconstruction error of the likelihood function between X and $\hat{X}$ and the features are encapsulated in weights W. Backpropagation via gradient descent algorithm is used for adjusting W. Stacked auto-encoders are formed by stacking auto encoders by wiring the learned weights to the next auto encoder's input.

2.1 Intervertebral Disc Detection

In the proposed architecture, first the lumbar MRI features are learned with stacked auto-encoders. Let $d = \{d_1, d_2, ..., d_6\}$ be the labels of the lumbar intervertebral discs in an MR image. Our goal is to identify the location $l_i \in \Re^2$ of each disc d_i on the image I. Randomly selected patches from image I are used for learning the features of the images. Let β be a patch of size $m \times n$ of image I where m and n varies between the minimum and maximum disc width and height in the training set, respectively. The image patch β is resized to $r \times r$ pixels and is formed into a $1 \times r^2$ vector to be used as an input of an autoencoder. Figure 3 shows the unsupervised learning of lumbar MR image features with an auto-encoder.

The stacked auto-encoder with $X = r^2$ input nodes is trained with the vectorized image patches β. The weights W of the final hidden layer are brought to

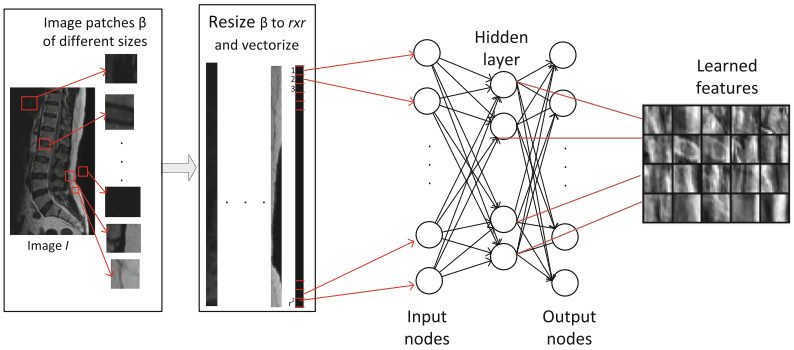


Fig. 3. An auto-encoder for learning MR image features. A single hidden layer auto-encoder trained with the vectorized image patches

square form (having $r \times r$ size) for building the feature set f of the MR images extracted in an unsupervised manner as explained in Sect. 2.

The feature set f includes the features of the whole MR image; however the objective of the proposed system is diagnosing the diseases related with the discs. To filter the irrelevant medical structures that exist in the image, we use nonlinear filtering with SVM. A sliding window approach is employed and each window $\Psi(p)$ enclosing the pixel p is convolved with the filter $f_i \in f$. The outputs of the convolution of each window with the filters in f are concatenated and the final feature vector is built. Each pixel p in the image I is given a score S_p with SVM that indicates the probability of being a location of disc d_i using f.

In order to locate and label the intervertebral lumbar discs, we follow the graphical model based labeling approach presented in [8] by enhancing the model with the unsupervised feature learning. We use a chain-like graphical model G consists of 6 nodes and 5 edges connecting the nodes where each lumbar intervertebral disc d_i is represented with a node. Our goal is to infer the optimal disc positions $d^* = \{d_1^*, d_2^*, ..., d_6^*\}$ where $d_i^* \in \Re^2$ and $1 \leq i \leq 6$ in the image I according to the given scores S_p and the spatial information between the discs in the training set. The optimal locations d^* of the discs are determined by using the maximum a posteriori estimate

$$d^* = \arg\max_d P(d|I, S_p, \alpha), \tag{2}$$

where I represents the image, S_p is the given score and α represents the parameters learned from the training set. The Gibbs distribution of $P(d|I, S_p, \alpha)$ is

$$P(d|I, P_s, \alpha) = \frac{1}{Z} exp\left\{-\left[\sum \psi_L(I, d_k) + \lambda \sum \psi_{spa}(d_k, d_{k+1}, \alpha)\right]\right\}. \tag{3}$$

The function $\psi_L(I, d_k)$ represents the scores S_p given via deep learning and the potential energy function $\psi_{spa}(d_k, d_{k+1}, \alpha)$ captures the geometrical information between the neighboring discs d_k and d_{k+1}. The optimal solution d^* is

gathered with dynamic programming in polynomial time. For the details of the graphical model G and inference, please refer to [8].

2.2 Diagnosis of DDD

After localizing the discs in the MR images, the disc features should be learned and they should be classified as healthy or not. The location l_i of each disc d_i is found with the Eq. 2. Since the window $\psi(p)$ enclosing the pixel p is known, these windows are directly used for CAD of degenerative disc disease. The windows $\Psi(p)$ of each located disc are used for training a sparse auto-encoder. The windows $\psi(p)$ are resized and vectorized to be used as input. The features are learned with sparse auto-encoders. The weights W of the final hidden layer of the auto-encoder are the used as the features f_d.

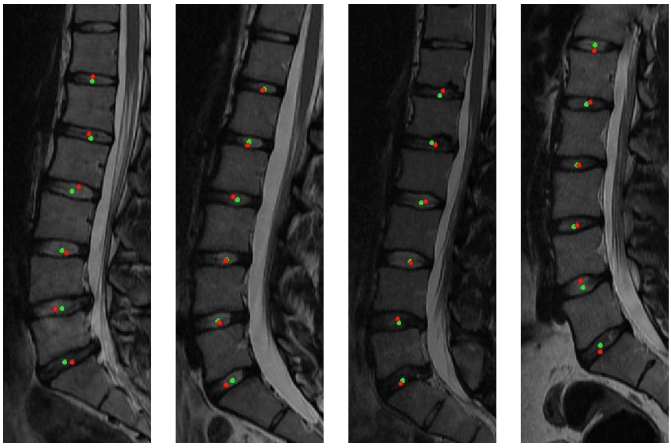
After determining the features of the discs, we again convolve the window $\psi(p)$ with the learned filter f_d. The output of the convolution operations are concatenated and the final feature vector is formed. These final feature vectors are trained and tested with SVM. Binary classification is performed and each window ψ is classified as having degenerative disc disease or not.

3 Experiments

In order to evaluate the proposed system, two different datasets, one with labeled and another with unlabeled discs, are used. First clinical MR image dataset contains the lumbar MR images of 102 subjects. The MR images are 512 × 512 pixels in size. In the images, there are 612 (102 subjects*6 discs) lumbar intervertebral discs where 349 of them are normal and 263 of them are diagnosed with degenerative disc disease. The disc boundaries are delineated and each disc is diagnosed having DDD or not by an experienced radiologist to be used as the ground truth. The second dataset includes the lumbar MR images of 43 subjects where the intervertebral discs are neither delineated nor diagnosed by an expert. This unlabeled dataset is used for providing data to the auto-encoder for unsupervised training. It is not used for testing the system since it does not include the ground truth.

For labeling process, randomly selected patches are used from the MR images. The width and height of the intervertebral discs are between 30–34 mm and 8–13 mm, respectively [13]. The patch size is selected in accordance with the intervertebral disc size. The total number of patches used for training is 10000. For preprocessing, the mean intensity value of the patch is subtracted from the image patch for normalization. The patches are resized to 15×15 pixels ($r = 15$) and the number of the input nodes X is 225. Two layers are used for the stacked auto encoder. The number of nodes in layer the first inner layer is 70 and the number of nodes in the second layer is 30.

The number of features f learned from the MR image patches is 30. Six-fold-cross-validation is used for SVR training. The parameters of the Eq. 3 are learned from the training set and the weighting parameter λ is selected as 0.5 empirically.



(a) (b) (c) (d)

Fig. 4. Labeling results of the lumbar MR images selected from the database. Green rectangles are the ground truth center points and the red rectangles are the disc centers determined by our system. The MR images are cropped for better visualization (Color figure online)

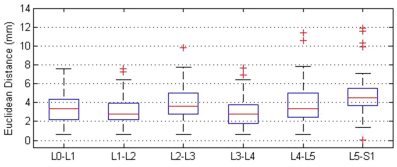


Fig. 5. Boxplot of the Euclidean distances of the disc centers determined by our system to the ground truth centers

Some of the visual labeling results of our system is shown in Fig. 4. In order to evaluate the performance of the labeling system with unsupervised feature learning, the Euclidean distances between the disc center point detected by our system and the ground truth are calculated. Figure 5 shows the boxplot of the Euclidean distances in mm.

For automated DDD diagnosis, a similar validation method is followed. Since the disc labels d determined for an image I and their enclosing windows ψ are determined in the labeling step, they are employed as the image patches for training and testing. Leave-one-out approach is used for training. Instead of using the whole window ψ, we use the half right side of the window ψ since the DDD including disc bulging and herniation occur at the right side. A two-layer stacked auto-encoder (70 nodes in the first layer, 40 nodes in the second layer) is employed for learning the features. The half right side of the labeled disc images are resized to 15×15 pixels in size and they are the input of the auto-encoder after vectorization.

Table 1. The accuracy, specificity, and sensitivity of the hand-crafted feature extraction methods and our method

Feature type	Number of features	Accuracy	Sensitivity	Specificity
Raw image intensity	1000	0.86	0.88	0.84
LBP	8	0.70	0.80	0.57
Gabor	1000	0.60	0.80	0.33
GLCM	5	0.71	0.78	0.62
Planar shape	3	0.55	1.0	0
Hu's moments	7	0.72	0.72	0.71
Intensity difference	12	0.89	**0.96**	0.82
Our method	40	**0.92**	0.94	**0.90**

After determining the features, each disc image is convolved with the features and the final feature vector for the final classification with binary SVM is created. The classification accuracy of the proposed system is 92 %.

In order to compare the unsupervised learned features with the hand-crafted features, popular feature types used in [1,9] are also implemented. The training is performed with six-fold-cross correlation and classification is performed via SVM. The number of features extracted and their accuracy, sensitivity, and specificity are reported in Table 1. The numerical results show that unsupervised learned features outperform hand-crafted features. The highest accuracy of the hand-crafted features 89.54 % for the intensity difference feature that calculates the numerical values (mean, standard deviation, etc.) of the intensities difference between T1-weighted and T2-weighted images. The accuracy of the unsupervised feature learning is higher than other hand-crafted features. In addition, the sensitivity and the specificity rates of the proposed system are higher than other state-of-the-art methods.

The experiments performed show that the DDD can be automatically diagnosed with a high accuracy with a few filters learned by auto-encoders. The unsupervised filters outperform other popular hand-crafted features even their number is lower than the hand-crafted features. In addition, the proposed system does not require a deep network structure including many hidden layers. The disc filters are efficiently learned with a two-layer auto-encoder with small training data.

4 Conclusions

In this paper, we present a novel method for CAD of the DDD with auto-encoders. The proposed architecture involves stacked auto-encoders and non-linear filters together for locating the intervertebral discs and diagnosis. The auto-encoders learns the image features effectively while the non-linear filters eliminates the irrelevant information. The system is validated on a real dataset

of 102 subjects. The results showed that unsupervised learning of features yields a better representation and the features could be extracted with minimal user intervention. The comparison with popular hand-crafted features show that the results are comparable with the state of the art.

References

1. Ghosh, S., Alomari, R.S., Chaudhary, V., Dhillon, G.: Composite features for automatic diagnosis of intervertebral disc herniation from lumbar MRI. In: Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5068–5071 (2011)
2. Ghosh, S., Alomari, R.S., Chaudhary, V., Dhillon, G.: Computer-aided diagnosis for lumbar MRI using heterogeneous classifiers. In: IEEE International Symposium on Biomedical Imaging, pp. 1179–1182 (2011)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
4. Koh, J., Chaudhary, V., Dhillon, G.: Diagnosis of disc herniation based on classifiers and features generated from spine MR images (2010)
5. Lootus, M., Kadir, T., Zisserman, A.: Radiological grading of spinal MRI. In: MICCAI Workshop: Computational Methods and Clinical Applications for Spine Imaging (2014)
6. Luoma, K., Riihimaumlki, H., Luukkonen, R., Raininko, R., Viikari-Juntura, E., Lamminen, A.: Low back pain in relation to lumbar disc degeneration. Spine **25**(4), 487–492 (2000)
7. National Centers for Health Statistics: Chartbook on trends in the health of Americans, special feature: pain (2011). http://www.cdc.gov/nchs/data/hus/hus06.pdf/
8. Oktay, A.B., Akgul, Y.S.: Simultaneous localization of lumbar vertebrae and intervertebral discs with SVM-based MRF. IEEE Trans. Biomed. Eng. **60**(9), 2375–2383 (2013)
9. Oktay, A.B., Albayrak, N.B., Akgul, Y.S.: Computer aided diagnosis of degenerative intervertebral disc diseases from lumbar MR images. Comput. Med. Imaging Graph. **38**(7), 613–619 (2014)
10. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: a unified embedding for face recognition and clustering. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015
11. Tang, Y.: Deep learning using support vector machines (2013)

12. Van Rijn, J.C., Klemetsouml, N., Reitsma, J.B., Majoie, C.B.L.M., Hulsmans, F.J., Peul, W.C., Stam, J., Bossuyt, P.M., den Heeten, G.J.: Observer variation in MRI evaluation of patients suspected of lumbar disk herniation. AJR Am. J. Roentgenol. **184**(1), 299–303 (2005)
13. Zhou, S., McCarthy, I., McGregor, A., Coombs, R., Hughes, S.: Geometrical dimensions of the lower lumbar vertebrae - analysis of data from digitised CT images. Eur. Spine J. **9**(3), 242–248 (2000)