Medical Image Analysis 45 (2018) 41–54

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

3D multi-scale FCN with random modality voxel dropout learning for Intervertebral Disc Localization and Segmentation from Multi-modality MR Images

Xiaomeng Li[a], Qi Dou[a], Hao Chen[a,*], Chi-Wing Fu[a], Xiaojuan Qi[a], Daniel L. Belavý[b,c], Gabriele Armbrecht[c], Dieter Felsenberg[c], Guoyan Zheng[d,1,*], Pheng-Ann Heng[a,1]

[a] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China
[b] Institute of Physical Activity and Nutrition Research, Deakin University, Burwood, Victoria, Australia
[c] Charité University Medical School, Berlin, Germany
[d] Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland

ARTICLE INFO

Article history:
Received 31 May 2017
Revised 23 December 2017
Accepted 16 January 2018
Available online 2 February 2018

Keywords:
Multi-modality
Magnetic resonance imaging
Intervertebral discs
Localization
Segmentation
Deep learning
Dropout

ABSTRACT

Intervertebral discs (IVDs) are small joints that lie between adjacent vertebrae. The localization and segmentation of IVDs are important for spine disease diagnosis and measurement quantification. However, manual annotation is time-consuming and error-prone with limited reproducibility, particularly for volumetric data. In this work, our goal is to *develop an automatic and accurate method based on fully convolutional networks (FCN) for the localization and segmentation of IVDs from multi-modality 3D MR data*. Compared with single modality data, multi-modality MR images provide complementary contextual information, which contributes to better recognition performance. However, how to effectively integrate such multi-modality information to generate accurate segmentation results remains to be further explored. In this paper, we present a novel multi-scale and modality dropout learning framework to locate and segment IVDs from four-modality MR images. First, we design a 3D multi-scale context fully convolutional network, which processes the input data in multiple scales of context and then merges the high-level features to enhance the representation capability of the network for handling the scale variation of anatomical structures. Second, to harness the complementary information from different modalities, we present a random modality voxel dropout strategy which alleviates the co-adaption issue and increases the discriminative capability of the network. Our method achieved the 1st place in the MICCAI challenge on automatic localization and segmentation of IVDs from multi-modality MR images, with a mean segmentation Dice coefficient of 91.2% and a mean localization error of 0.62 mm. We further conduct extensive experiments on the extended dataset to validate our method. We demonstrate that the proposed modality dropout strategy with multi-modality images as contextual information improved the segmentation accuracy significantly. Furthermore, experiments conducted on extended data collected from two different time points demonstrate the efficacy of our method on tracking the morphological changes in a longitudinal study.

1. Introduction

Intervertebral discs (IVDs) are spine components that lie between each pair of adjacent vertebrae. They serve as shock absorbers in the spine and are crucial for vertebral movement. Disc degeneration (An et al., 2004; Urban and Roberts, 2003) is a common cause of back pain and stiffness for adults, and is a major public health problem in modern societies. Traditionally, studies on disc degeneration were done mainly by means of manual segmentation of the discs. Such a manual approach is, however, rather tedious and time-consuming, and is often subject to inter- and intra-observer variabilities (Violas et al., 2007; Niemeläinen et al., 2008). In this regard, automatic localization and segmentation of intervertebral disc can help to reduce manual labor work and assist in the disease treatment by providing quantitative parameters, which improves the efficiency and accuracy for spine pathologies diagnosis.

[*] Corresponding authors.
E-mail addresses: hchen@cse.cuhk.edu.hk (H. Chen), guoyan.zheng@istb.unibe.ch (G. Zheng).
[1] Authors share the senior authorship.

https://doi.org/10.1016/j.media.2018.01.004
1361-8415/© 2018 Elsevier B.V. All rights reserved.

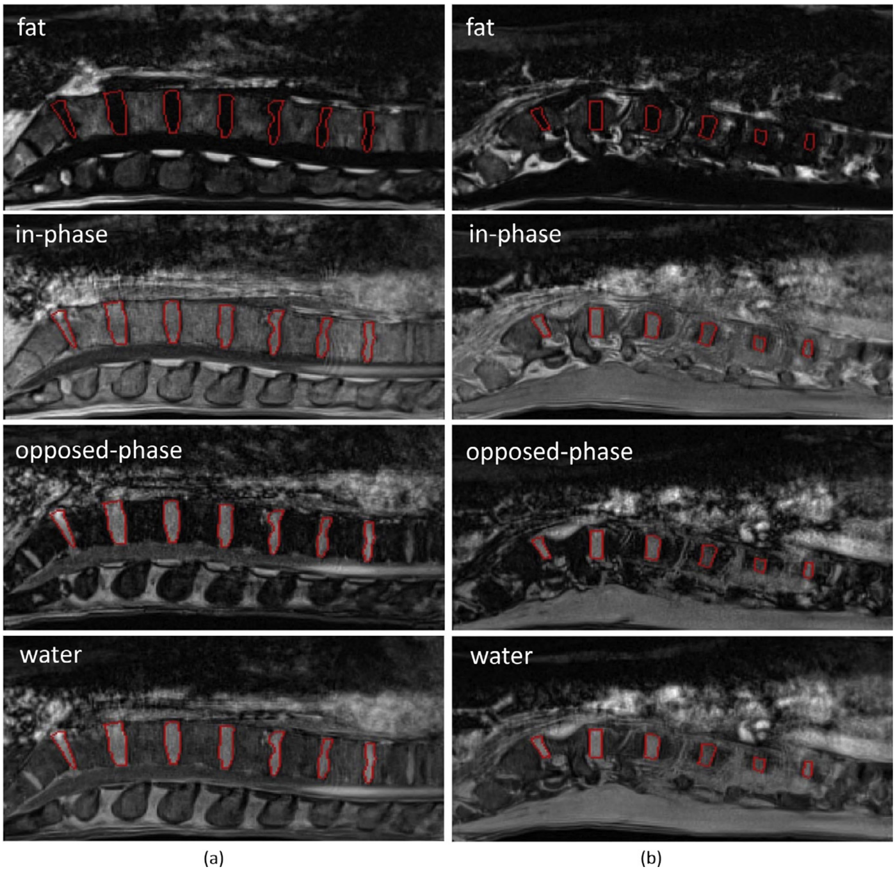


Fig. 1. Examples of 3D multi-modality input data. (a) and (b) show two data sets scanned from two different patients, each including four 3D modalities: fat, in-phase, opposed-phase, and water (top to bottom). In these figures, we show the 18th slice in the 3D images; red contours indicate the boundary of the IVDs, which are the ground truth marked by radiologists. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Magnetic resonance imaging (MRI) is an excellent non-invasive technique, commonly used in spine disease diagnosis such as disc herniation degeneration and spinal stenosis (Tertti et al., 1991; Schneiderman et al., 1987; Hamanishi et al., 1994; BenEliyahu, 1995), due to its excellency in soft tissue contrast without ionizing radiation. Multi-modality MR images can be obtained with different scanning configurations for the same subject (see Fig. 1). Hence, it can provide more comprehensive information for robust diagnosis and treatment, as demonstrated in the recent work by Misri (2013). For example, the in-phase and water image modalities in Fig. 1 have low intensity contrast between the IVDs and their neighboring regions, while the fat and opposed-phase image modalities have high intensity contrast. The effective integration of these multi-modal information facilitates more accurate delineation of the IVD boundary.

In this work, we are interested in the *automatic localization and segmentation of IVDs* from 3D multi-modality spine MR images. Localization refers to the identification of the centroid of each IVD, while segmentation refers to the generation of a binary mask to indicate the IVD regions in the image domain, where a 3D surface can be constructed for the IVD boundary.

Automatic localization and segmentation of IVDs from volumetric data are difficult due to following challenges. First, the IVDs have large variations in shape, even for the same subject, thus hindering robust localization and segmentation as illustrated in Fig. 1. Second, the intensity resemblance between IVDs and their neighboring structures interferes the detection of disc boundary. Lastly, how to take full advantage of multi-modality information to improve the segmentation performance remains to be fully explored.

1.1. Previous work

Most of previous methods localized and segmented the IVDs using hand-crafted features derived based on intensity and shape information (Schmidt et al., 2007; Chevrefils et al., 2007; Shi et al., 2007; Corso et al., 2008; Chevrefils et al., 2009; Raja'S et al., 2011; Neubert et al., 2011; Ayed et al., 2011; Law et al., 2013; Haq et al., 2014; Korez et al., 2015). For localization, Schmidt et al. (2007) proposed a graphical model based on image intensity and geometric constraints for spine detection and labeling. Specifically, they employed a part-based graphical model to represent both the shape of local parts and the anatomical structures between the parts. Corso et al. (2008) and Raja'S et al. (2011) independently proposed two different graphical models to improve the localization accuracy by capturing both pixel- and object-level features.

For segmentation, different types of graph-based methods are very popular in the segmentation of vertebrae or discs. For example, Carballido-Gamio et al. (2004) proposed the normalized cut to segment vertebral bodies from MR images. Another new form of graph cuts was proposed by Ayed et al. (2011). They developed new object interaction priors for graph cut image segmentation and employed the method to delineate the IVDs in spine MR images. Recently, Neubert et al. (2011) segmented the IVDs and vertebral bodies from high-resolution spine MR images by using a statistical shape model based method.

Machine learning-based methods have gained increasing interest in the field of medical image analysis. Great successes have been validated in different medical image analysis problems. For example, Kelm et al. (2013) detected spine in CT and MR images by marginal space learning (MSL), which was proposed by Zheng et al. (2008) to localize the heart chamber in 3D CT data at first. In the study of Kelm et al. (2013), the IVDs were detected and segmented based on Haar-like features under a MSL scheme. Huang et al. (2009) proposed an enhanced Adaboost classifier with an over-complete wavelet representation to detect vertebra and segment vertebra by iterative normalized-cut method. Another two regression-based methods were proposed by Chen et al. (2014, 2015a) and Wang et al. (2015). Chen et al. (2014, 2015a) proposed a unified data-driven estimation framework to estimate the image displacements to localize IVDs and then segment IVDs by predicting foreground and background probability of each pixel in which the neighborhood intensity vector were used as visual features. While Wang et al. (2015) designed a sparse kernel machine based regression method taking hand-crafted features including texture and shape as input to segment disc and vertebral structures from both MRI and CT modalities. However, these hand-crafted features tend to suffer from limited representation capability compared with the automatically end-to-end learned features.

More recently, with the advance of deep learning techniques (Simonyan and Zisserman, 2014; Long et al., 2015; Chen et al., 2016b; He et al., 2016; Dou et al., 2017), many researches have proposed deep learning based methods to localize and segment IVDs or vertebrae from volumetric data or 2D images (Cai et al., 2015; Suzani et al., 2015; Chen et al., 2015c; Jamaludin et al., 2016; Chen et al., 2016a; Zheng et al., 2017). For example, Cai et al. (2015) recognized vertebra by a 3D deformable hierarchical model from multi-modality images and achieved the detection by using multi-modality features extracted from deep neural networks. Although both Suzani et al. (2015) and Chen et al. (2015c) employed deep learning methods for vertebrae identification in spine images. The method proposed by Suzani et al. (2015) was based on feed-forward networks and the method designed by Chen et al. (2015c) was based on convolutional neural networks (CNN). Very recently, Jamaludin et al. (2016) proposed a CNN based framework to automatically label each IVD and the surrounding vertebrae with a number of radiological scores. They demonstrated that radiological scores and pathology hotspots can be predicted to an excellent standard using only the “weak” supervision of class labels. Chen et al. (2016a) introduced a 3D fully convolutional network (FCN) to localize and segment IVDs, which has achieved the state-of-the-art localization performance in MICCAI 2015 IVD localization and segmentation challenge.

FCNs have become the back-bone of state of the art medical image segmentation systems and a lot of variant FCNs have been proposed to advance this stream, including multi-scale FCN, multi-path fusion and multi-modality FCN. For example, Kamnitsas et al. (2017) proposed a multi-scale 3D FCN with two convolutional pathways for brain lesion segmentation, where the multi-scale information is fed into the network by using the low resolution and the normal resolution input. Sun et al. (2017) designed a multi-channel FCN to segment liver tumors from CT images, in which the probability map was generated by features fusion from multiple channels. All these multi-scale FCNs and multi-path fusion FCN achieved remarkable improvement. We also propose a multi-scale FCN with three convolutional pathways, which shares the same spirit with the multi-scale FCN. However, the network structure in each pathway is different, providing various kernels’ field-of-view in each pathway and enabling feature extraction for different scale of contexts.

With various modality images being available in the medical imaging community (e.g., CT, MRI, etc.), multi-modality FCNs have also been developed and the contribution of multi-modality images was also verified by some very recent deep learning based research work on vertebrae (Cai et al., 2016), brain (Zhang et al., 2015; Chen et al., 2017a), and brain tumor segmentation (Havaei et al., 2016), where the segmentation performance can be significantly improved based on the multi-modality data. Despite of the improvements, none of them have yet explored how to effectively harness multi-modality images for segmentation performance gains. In this regard, we proposed a random voxel dropout learning strategy which showed to be effective in learning with multi-modality images.

1.2. Our contributions

We propose a 3D multi-scale and modality dropout learning framework for localizing and segmenting IVDs from multi-modality MR images. Experimental results on the *MICCAI 2016 Challenge on Automatic Intervertebral Disc Localization and Segmentation from 3D Multi-modality MR Images* demonstrated the superiority of our proposed framework.

Our main contributions can be summarized as follows:

1. We propose a novel multi-scale 3D fully convolutional network, which consists of three pathways to integrate spatial information of multiple scales input. This network is inherently general and can be adopted for other medical image segmentation tasks to handle objects with large variations, such as tumor segmentation.
2. We propose a modality drop strategy for maximizing the utilization of complementary information from multi-modality MR data. This is the first study we are aware of adopting dropout strategy for segmenting multi-modality images. Experiments show that, compared with a network trained without dropout strategy, the network with dropout strategy can generate more discriminative features and achieve more accurate segmentation results.
3. We applied our method to the datasets of MICCAI 2016 challenge, which consists of 24 sets of 3D multi-modality MR images acquired from two different time points. The results achieved by our method demonstrated the efficacy of our method on tracking the morphological changes in a longitudinal study.

This paper is organized as follows. We first present the details of our method in Section 2. Dataset and extensive experimental results are described in Section 3. We then discuss the significance of our work from the perspectives of both clinical application and computational analysis in Section 4. Finally, conclusions are draw in Section 5.

2. Methodology

Fig. 2 presents an overview of our proposed multi-scale FCN with random modality voxel dropout learning framework for IVD localization and segmentation from multi-modality MR images. To handle the scale variations of IVDs, our multi-scale fully convolutional network (MsFCN) consists of three pathways, which take volumetric regions extracted from the same location to harness contextual information in different scales. To enhance the training efficacy, modality dropout strategy is employed on the input multi-modality data to reduce the feature co-adaption and encourage each single modality image to provide discriminative information independently. In the following, we will first present the architecture of 3D FCN, then detail our proposed MsFCN framework and finally we elaborate the modality dropout strategy for effective multi-modality learning.

2.1. 3D FCN for end-to-end IVD segmentation

CNNs have achieved remarkable successes in 2D medical image analysis tasks (Shin et al., 2016; Shen et al., 2017; Chen et al., 2015b; Sirinukunwattana et al., 2015; Greenspan et al., 2016; Chen et al., 2017b; Kong et al., 2016; Ronneberger et al., 2015), however,

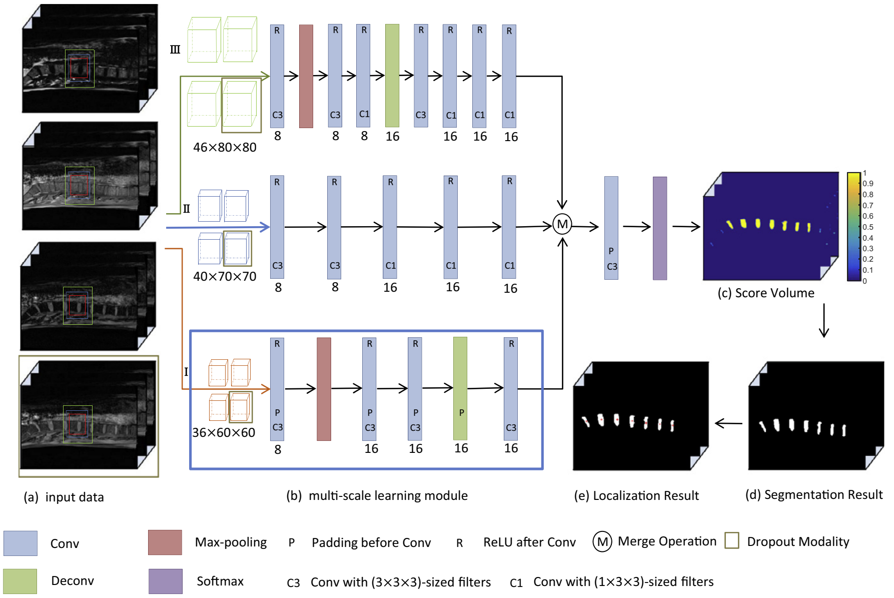


Fig. 2. Our proposed MsFCN framework for IVDs localization and segmentation. (a) Four modality input images. (b) Our multi-scale learning module. Each pathway contains several convolutional layers, ReLU layers, and deconvolutional layers. (c) Score volume. (d) Segmentation results obtained by setting a threshold on the probability map. (e) Localization results are the centroids of the IVDs identified in each segmentation mask (shown as the red crosses). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

how to effectively employ CNNs to handle volumetric data still remains an open problem in the field of medical image computing. One straightforward way is to employ conventional 2D CNNs based on a single slice and process the slices sequentially (Jamaludin et al., 2016; Ji et al., 2016b; Chen et al., 2015c) or aggregate orthogonal planes (Roth et al., 2014). However, this solution may disregard the volumetric contextual information, which would impede the representation capability of network and thus degrade the performance. One alternative approach is 3D FCN for end-to-end IVD segmentation, which takes a volumetric image as input and outputs the segmented mask directly (Chen et al., 2016a). Instead of using 2D convolutional kernels, 3D FCN encodes richer spatial information of the volumetric data by employing volumetric kernels. For max-pooling layers, max-pooling operation is performed in a 3D fashion, where activations with a cubic neighborhood are abstracted and transmitted to higher layers. In addition, 3D deconvolution layers are adopted to bridge the coarse feature volumes into dense predictions with the same size to the input image. With volumetric operations in 3D FCN, the spatial information in three dimensions can be fully explored and contributes to the segmentation performance improvement. In our work, we adopt a modified 3D FCN for IVD localization and segmentation as illustrated in Fig. 2.

2.2. 3D multi-scale FCN for context fusion

One limitation of previous methods on IVD localization and segmentation (Chen et al., 2016a; Ji et al., 2016a) is that they usually considered a single-scale of spatial information surrounding the discs, which lacks the capability of handling different scales of anatomical structures. In contrast, multi-scale contextual information can contribute to better recognition performance (Kamnitsas et al., 2015; Chen et al., 2016c; Zhao et al., 2016; Kamnitsas et al., 2017). Therefore, we propose to use multi-scale FCN for incorporating contextual information in different scales to conquer the scale and shape variations of IVDs. Fig. 2 shows the architecture of our proposed method. Specifically, given four modality data in Fig. 2(a), we crop three sizes, i.e., $36 \times 60 \times 60$, $40 \times 70 \times 70$, $46 \times 80 \times 80$, of volumetric data centered on the same position from the four modality images. The selection of three scales is based on the observation that the IVD horizontally occupies at most 50 voxels in all training samples, shown in Fig. 3. The smallest scale should ensure that each IVD can be enclosed within the crop patch and it is not necessary to exceed the main spine region too much for introducing redundant information. With this consideration, three sizes are cropped, i.e., $36 \times 60 \times 60$, $40 \times 70 \times 70$, $46 \times 80 \times 80$. These cropped four-modality patches are fed into the network in four channels. The multi-scale learning module is used to gather contextual information in different scales for the centered region, as shown in Fig. 2(b). In our framework, the multi-scale learning module contains three pathways. Each pathway consists of several convolutional layers with kernel size $1 \times 3 \times 3$ or $3 \times 3 \times 3$ interleaved with activation function, i.e., rectified linear units (ReLU). To enlarge the receptive field of the network, max-pooling layers are employed in the intermediate layers. The three pathways are delicately designed such that the output feature maps are with the same size even though the input sizes are different as shown in

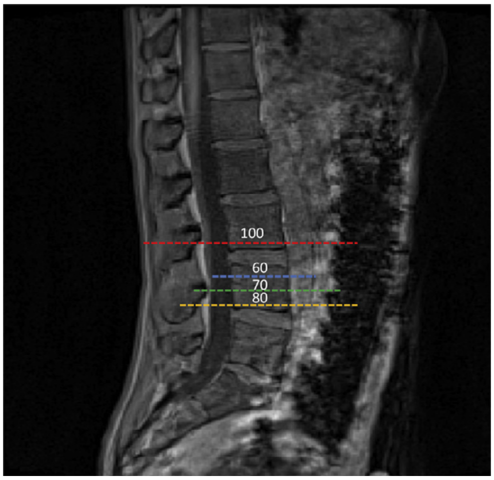


Fig. 3. Scale illustration on one spine image.

Fig. 2(b). Specifically, the size of feature maps in pathway I remains unchanged since we apply padding operation before each convolution layer. However, without padding process, the size of feature maps will be reduced in pathway II and III. In this way, the feature map in pathway II and III turns to size 36 * 60 * 60 at the end, which keeps the consistency with that in pathway I. The feature maps derived from the three pathways are then concatenated together and the following convolutional layers are applied to generate the final probability map as shown in Fig. 2(c). The final IVD segmentation results (see Fig. 2(d)) can be generated from the score volume by thresholding (set as 0.9 in our experiments) and the localization results are derived as the centroids of each connected component of the segmentation mask. In summary, our architecture provides a solution to gather multi-scale contextual information for voxel-wise prediction. Different levels of representation are fused together for the final prediction which can incorporate different scales of context for better recognition.

2.3. Random modality voxel dropout for effective multi-modality learning

Multi-modality images have been utilized in many medical image analysis tasks, which can provide complementary information for improving recognition performance. For example, Zhang et al. (2015) trained their network on the input of different modality images to achieve infant brain image segmentation. Their experimental results showed that training with multi-modality data can boost the segmentation performance. However, simply combining all modality volumes together for training may cause too much dependency among modalities, which may lead to feature co-adaption and thus degrades the performance. Recently, Havaei et al. (2016) proposed a deep learning segmentation framework that can be applied on incomplete multi-modality images. However, how to effectively employ the multi-modality images has not been fully explored.

Dropout has been proven with great success in training deep neural networks by randomly zeroing out the outputs of neurons (Hinton et al., 2012; Srivastava et al., 2014; Wan et al., 2013; Li et al., 2016b). It has been recognized as an effective way to prevent co-adaption of feature detectors and alleviate the over-fitting issue. Specifically, each hidden unit in a network trained with dropout must learn to work with randomly selected hidden units, which should make each hidden unit more robust and force it to create useful features without relying on other specific units to rectify its mistakes.

In the field of multi-modality images, the traditional methods take all multi-modality images as the input to the neural network in different channels. The voxels in the same locations at different modality images may rely on each other. This dependency would lead to data co-adaption problem in which the neurons detect the same feature repeatedly, indicating the network has not achieved its full capacity efficiently. To alleviate this problem, we propose a simple but effective strategy, referred as Random Modality Voxel Dropout (RMVD), which shares the similar spirit with dropout method (Srivastava et al., 2014), to enhance the feature learning process from different modalities. To be more specific, with four modality images as input shown in Fig. 1, we randomly zero out voxels with a certain ratio in the randomly selected modality image during each training iteration. Then, three modality images together with one randomly selected modality image performed RMVD strategy are input into the network for training. The motivation of this operation is that the random disappearance of the limited amount of voxels would force the network to avoid generating redundant features, thus improving the feature representation capability of our network. Different from traditional dropout method, our RMVD strategy is performed on multi-modality input images. At test time, we can approximate the effect of averaging the probability maps from all these dropout neural networks by simply using a neural network without dropout input data (Srivastava et al., 2014).

2.4. Weighted loss function

The number of background voxels is much larger than the number of foreground voxels (i.e., IVD) with a ratio of approximately 16:1. The imbalance of training samples would inevitably prohibit the network learning effective representations for the foreground classes. To solve this issue, we adopt the weighted loss function as following:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}[-\lambda \cdot t_{x_i}\log p(x_i) - (1 - t_{x_i})\log(1 - p(x_i))] \tag{1}$$

where λ is the weight for strengthening the importance of foreground voxels; N denotes the total number of voxels at the prediction volume; $p(x_i)$ denotes the corresponding probability value of IVD for voxel x_i; t_{x_i} denotes the label at voxel x_i. Note that t_{x_i} is 1 when x_i is a foreground voxel, otherwise 0.

We employed the Adam (Kingma and Ba, 2014) optimization algorithm for training the whole network, which has been demonstrated well in training deep neural networks compared to other optimization methods. During the inference stage, we directly segmented the original spine image and the probability map of the whole image was generated in an overlap-tile way. It is worthwhile noting that the localization and segmentation of IVDs are seamlessly integrated in our framework and the whole process was performed in an automated way without any manual intervention.

3. Experiments

3.1. Dataset and pre-processing

We evaluated our proposed method on the dataset from the *2016 MICCAI Challenge on Automatic Localization and Segmentation of IVDs from Multi-modality MR Images* [2] (Yao et al., 2017), which consists of 24 sets of 3D multi-modality MR images acquired in two different time points from 12 patients involved in the second Berlin BedRest study (Belavy et al., 2010). All the images were scanned with 1.5 Tesla MR scanner of Siemens (Siemens Healthcare, Erlangen, Germany) with following protocol: Slice thickness 2.0 mm, Pixel Spacing 1.25mm, Repetition Time (TR) = 10.6 ms, Echo time (TE) = 4.76 ms. The aims of this study were to understand the effects of inactivity on the human body and to simulate the effects of microgravity on human body by space agencies (Belavy et al., 2010; 2011; 2012). Each set of data consists of four modality MR images, i.e., in-phase, opposed-phase, fat, and water as well as a binary ground truth image. Thus, in total we have 12 subjects ×2 stages ×4 modalities = 96 volume data, with voxel spacing of 2 mm×1.25 mm×1.25 mm. The four multi-modality images of each subject at a time point are acquired in the same space and thus are aligned with each other. The ground truth segmentation for each set of data were then manually annotated. Table 1 summarizes the demographic statistics of the 12 subjects.

Table 2 shows the details of the training data and test data. During the MICCAI challenge stage, the organizer released *Training set of IVD challenge* as the training data. The ground truth (manual annotation) of the test data (*Test set of IVD challenge*) was held by the organizer for independent evaluation. After the challenge, we obtained *Additional training set* and conducted extensive experiments

[2] http://ivdm3seg.weebly.com.

Table 1
Demographic statistics of subjects involved in our study.

Subject characteristics	Mean ± SD	Min	Max
Age (year)	35.1 ± 8.5	21	45
Weight (kg)	69.8 ± 8.0	59	81.8
Height (cm)	176.0 ± 0.06	169	190

Table 2
Dataset details.

Dataset	Subject index
Training set of IVD challenge	A_S1, B_S1, C_S1, D_S1, E_S1, F_S1, G_S1, H_S1
Test set of IVD challenge	B_S2, F_S2, G_S2, I_S2, J_S2, K_S1
Additional training set	A_S2, C_S2, D_S2, E_S2, H_S2, I_S1, J_S1, K_S2, L_S1, L_S2

A ,..., L denote the subject index, respectively; S1 and S2 denotes the different time points.

on the extended dataset. In the data preprocessing stage, we subtracted the input data by the mean of intensity values of the whole dataset.

3.2. Evaluation metrics

We employed the challenge evaluation metrics to evaluate the performance of our method regarding the localization and segmentation respectively. We denote m as the number of testing subjects and d as the number of IVDs in one spine image. Since the challenge only evaluates 7 IVDs (T11-S1) in each spine image, hence $d = 7$ (Zheng et al., 2017).

1) *Localization*: Mean Localization Distance (MLD) is used to measure the accuracy of localization results and standard deviation(SD) quantifies the degree of variation (Zheng et al., 2017).

$$MLD = \frac{\sum_{i=1}^{m}\sum_{j=1}^{d} R_{ij}}{d \cdot m},$$
$$SD = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{d}(R_{ij} - MLD)^2}{d \cdot m - 1}}. \qquad (2)$$

where $R_{ij} = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}$ measures the localization difference for jth IVD in the ith test image; Δx, Δy, Δz denote the absolute location difference of the identified and ground truth IVD centers along X, Y, Z axes, respectively. MLD measures the average localization difference for IVDs. Lower MLD and SD values indicate better localization accuracy and stability.

2) *Segmentation*: For the segmentation, the evaluation criteria is Mean Dice Overlap Coefficients (MDOC) with standard deviation (SDDOC). MDOC and SDDOC are defined as:

$$MDOC = \frac{\sum_{i=1}^{m}\sum_{j=1}^{d} Dice_{ij}}{d \cdot m},$$
$$SDDOC = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{d}(Dice_{ij} - MDOC)^2}{d \cdot m - 1}}, \qquad (3)$$

where $Dice_{ij} = \frac{2|A \bigcap B|}{|A| + |B|} \times 100\%$ denotes the dice overlap coefficients (DOC) between the ground truth annotation A and segmentation result B for the jth IVD of the ith test subject. Larger MDOC value indicates the higher segmentation accuracy.

3.3. Results of MICCAI 2016 on-site challenge

We first present the MICCAI 2016 challenge results of IVD localization and segmentation, as shown in Table 3. A total of three teams participated in the challenge, and our method ranked the first among them, with MLD of 0.62 mm for localization and MDOC of 91.2% for segmentation. Fig. 4 showed one example of our results, which demonstrates that our method can accurately localize and segment IVDs from volumetric data.

Table 3
Results of IVDs localization and segmentation challenge on MICCAI 2016. (8 training samples and 6 test samples).

Method	Localization MLD(mm) ± SD(mm)	Segmentation MDOC(%) ± SDDOC(%)	Time cost Seconds (s)
Ours (Li et al., 2016a)	0.62 ± 0.38	91.2 ± 1.8	10
Regression forest and CNNs (Ji et al., 2016b)	0.64 ± 0.50	90.8 ± 3.9	30

In fact, the second-place team (Ji et al., 2016b) also employed CNNs, demonstrating the popularity, as well as performance gains of CNNs for IVD analysis. More specifically, they first coarsely localized the center of the first IVD by training the random forest regression (Gao and Shen, 2014) and then sequentially trained the kth CNN classifier to segment the kth IVD. The center of $(k+1)$th IVD can be predicted by the mean shape model and previous segmented IVDs. During the on-site competition, their method achieved MLD of 0.64 mm for localization and MDOC of 90.8% for segmentation.

In comparison with other methods, experimental results showed that our method is more accurate than Ji et al. (2016b), with 0.02 mm and 0.04% improvement for localization and segmentation, respectively. This is because our method is an end-to-end voxelwise segmentation system and we trained the CNN classifier for all IVDs at the same time, in which the CNN classifier is capable to recognize the feature relations among IVDs and thus improves the training efficiency. For localization part, instead of coarsely detecting IVDs first, we generate the centroids of each IVD on the segmentation mask as localization centers.

The other participant (Heinrich and Oktay, 2016) combined the vantage point forests, Hough aggregation and a simple graphic model to localize and segment IVDs. However, their method failed to segment one case, leading to much larger error in segmentation and was dropped out from the on-site competition. As reported in Heinrich and Oktay (2016), their method achieved segmentation accuracy of 89% (MDOC) and localization error of 0.69 mm (MLD) in the leave-one-out experiments.

Regarding the computational performance, our method can be quite efficient leveraging the fully convolutional architecture. For a set of multi-modality test images, it approximately took 9s to get the segmentation result and then an additional 1s to obtain the localization result using a Titan X GPU. However, the proposed method in Ji et al. (2016b) took about 30s to obtain the final results including 25s for segmentation and 5s for localization based on a Tesla K80 GPU. Even though the configuration of hardware is different, our method is inherently more efficient given the network architecture.

3.4. Ablation study

In this section, we conduct experiments to investigate the effectiveness of using multi-modality input images for automatic IVD localization and segmentation. Moreover, to demonstrate the superiority of our method, experiments are conducted to show the effectiveness of each proposed component in our method. Please note that the training data in this section is *Training set of IVD challenge* and *Additional training set* while the test data is *Test set of IVD challenge* in Table 2.

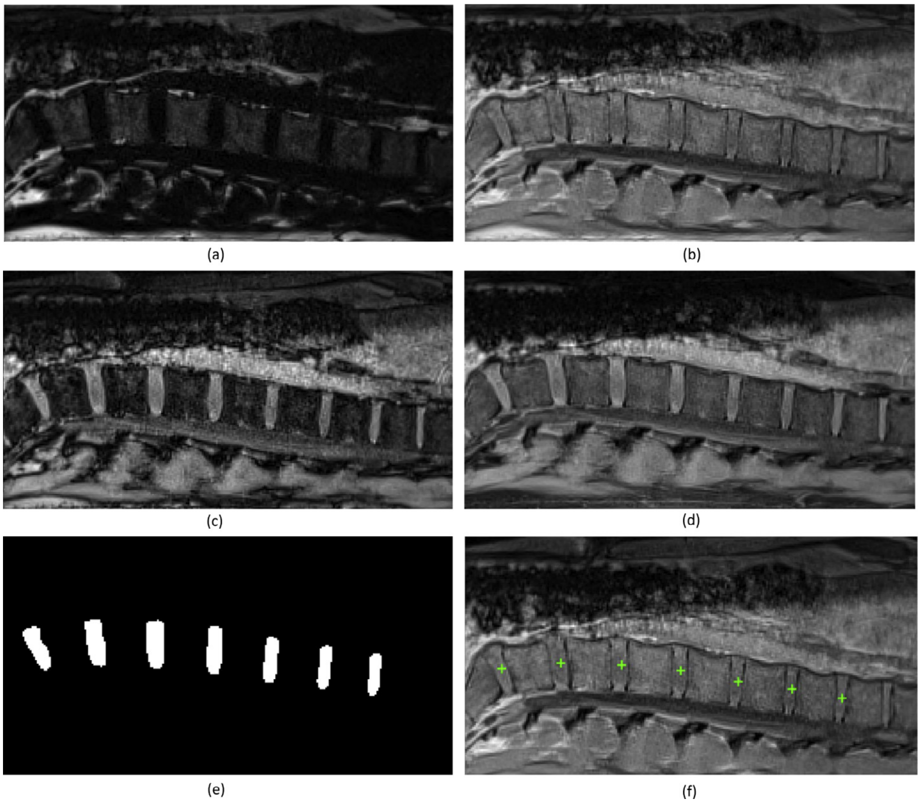


Fig. 4. One example result from the test data. (a)-(d) are the input images while (e) and (f) are the segmentation result and the localization result, receptively. We present the middle sagittal slice of 3D volumetric data for easy visualization.

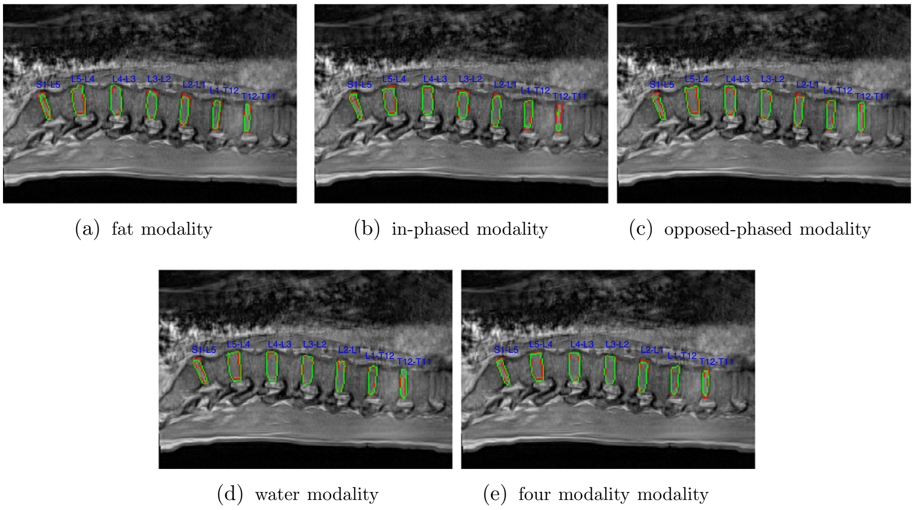
Table 4
Comparison of IVDs localization and segmentation results produced by the network with single-modality data input and multi-modality data input (18 training samples and 6 test samples).

Modality		Localization MLD(mm) ± SD(mm)	Segmentation MDOC(%) ± SDDOC(%)
Single modality	fat	0.77 ± 0.49	86.74 ± 3.68
	in-phase	0.87 ± 0.84	86.68 ± 5.13
	opposed-phase	0.44 ± 0.28	89.48 ± 2.61
	water	0.49 ± 0.39	89.68 ± 2.81
Four modality		**0.42** ± 0.29	**90.48** ± **1.97**

3.4.1. *Effectiveness of multi-modality input images*

We conducted comparative experiments by using networks trained with input of each single modality images (i.e., fat, in-phase, opposed-phase, and water) and multi-modality images, respectively. The network employed in this section is the pathway I of the proposed multi-scale framework (see Fig. 2), which is the *baseline*. All of these experiments were carried out using the same network architecture, training strategies and data augmentation strategies. Table 4 lists the experimental results. It is observed that training with multi-modality images input improves the localization and segmentation accuracy, compared to that with single-modality image input. This demonstrated that, by providing richer complementary information, network trained with multi-modality images can generate more discriminative features, hence improving IVD segmentation and localization accuracy. It is worthy to point out that results generated from networks trained with opposed-phase image and water image input have much lower localization error and higher segmentation accuracy than that with fat and in-phased image input. The reason is that opposed-phase and water images have larger intensity contrast around the IVDs and its neighboring regions than fat and in-phased images, which ease the difficulties for IVDs recognition.

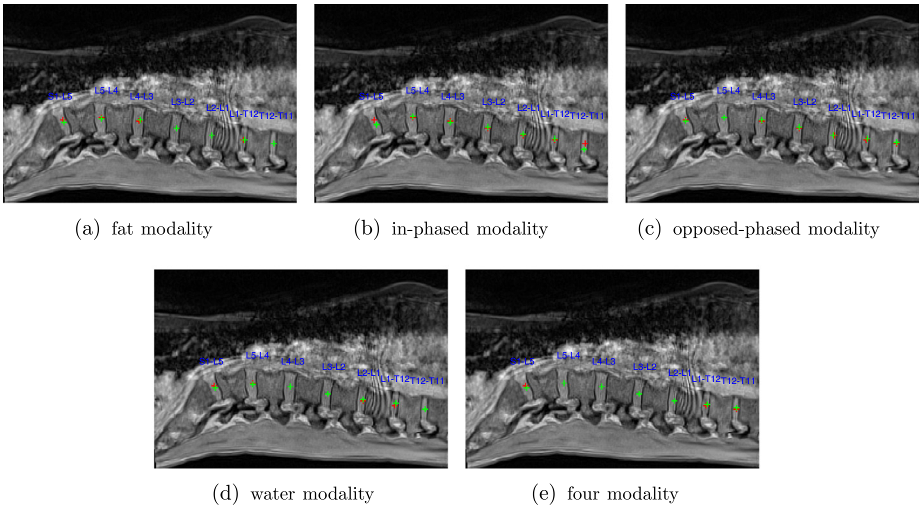
Figs. 5 and 6 present several examples of localization and segmentation results for different experimental settings, including training with fat, in-phase, opposed-phase, water modality image and multi-modality images. Green contour indicates the segmentation boundary of the IVD region while red contour is the ground truth boundary of that region. Green cross mark is the predicted IVD centers while red one is the IVD centers on ground truth images. If the segmentation and localization are perfect, we will only see the green contour or cross mark as it occludes the red one. It is observed that all the experimental settings can segment and localize IVDs L3-L2 and L4-L3 with reasonable accuracy. This is because L3-L2 and L4-L3 have regular shape appearance and normal size. However, we observed that the segmentation results on most of IVDs are more accurate in the last experiment than in other experiments, especially on the 12th sagittal slices in Fig. 5. Specifically, segmentation results is not consistent with the ground truth boundaries in some cases, such as for IVDs S1-L5, L5-L4, L2-L1 in the first experiment, IVD L1-T12 in the second experiment, IVD S1-L5 in the third and fourth experiments, but the prediction boundaries generated from the last experiment can achieve reasonable accuracy in these IVD cases. For localization, networks trained with input of fat and in-phased images are not able to localize S1-L5 and L1-T12 in an accurate manner. The localization of S1-L5 and L5-L4 were better predicted in the networks with four modality images input rather than the input of opposed-phase or water images only. These observations confirmed that training with multi-modality images could achieve better localization and segmentation results, which can increase the capability of networks to handle more challenging targets than with input of single-modality images.



(a) fat modality (b) in-phased modality (c) opposed-phased modality

(d) water modality (e) four modality modality

Fig. 5. Examples of the segmentation results for test subject G_S2. For clear visualization, we show results on the 12th slice of in-phase modality image. (a), (b), (c), (d) and (e) are the segmentation results generated from networks with fat, in-phased, opposed-phased, water images and four-modality images, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a) fat modality (b) in-phased modality (c) opposed-phased modality

(d) water modality (e) four modality

Fig. 6. Examples of the localization results for test subject B_S2. For clear visualization, we show results on the 25th slice of in-phase modality image. (a), (b), (c), (d) and (e) are the localization results generated from networks with fat, in-phased, opposed-phased, water images and four-modality images, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4.2. *Effectiveness of learning techniques*

To investigate the effectiveness of our proposed learning techniques including both 3D multi-scale learning module (MsFCN) and random modality voxel drop strategy (RMVD), we compare the segmentation results achieved by the baseline, MsFCN, MsFCN+RMVD, MsFCN-A + RMVD and 3D U-Net (Çiçek et al., 2016). All experiments have the same training strategies and data augmentation strategies for fair comparison.

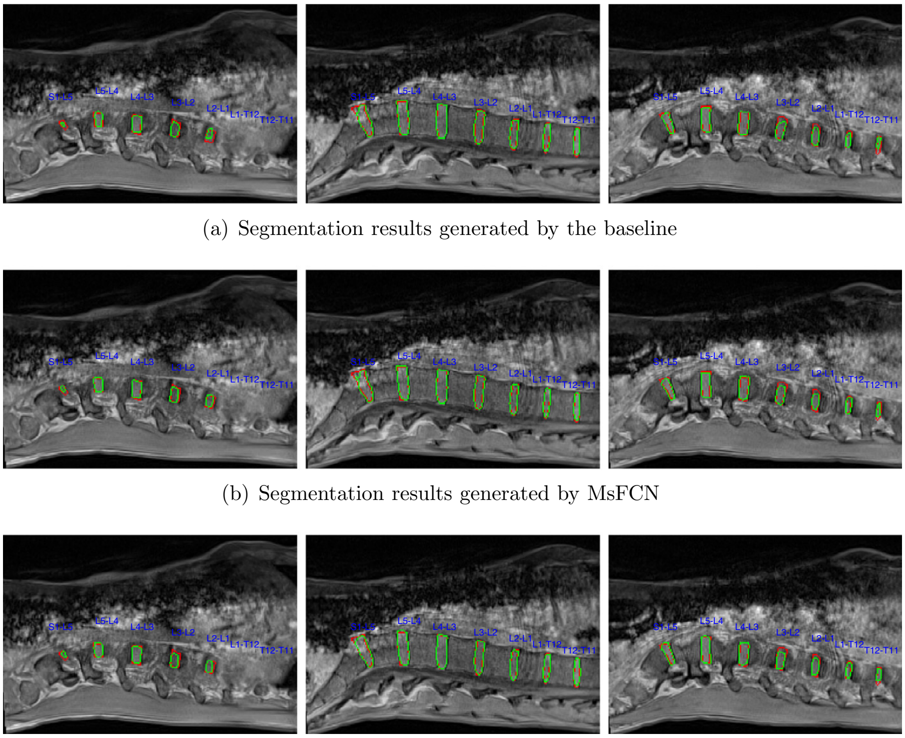
Comparison of baseline and MsFCN. Table 5 presents the localization and segmentation results of these three experimental configurations. It is observed that our MsFCN can generate much better results than directly employing the single-scale input deep neural network, with a localization error of 0.03 mm improvement on the MLD metric and a segmentation accuracy of 0.69% improvement on the MDOC evaluation. This is because most of the IVDs have varying shapes (see Fig. 1). Training the segmentation network with 3D MsFCN instead of the single-scale input neural network can effectively integrate the different levels of contextual information, hence improving the capability to discriminate features for better recognition.

Table 5
Comparison and effectiveness of methods on IVD localization and segmentation (18 training samples and 6 test samples).

Method	Localization MLD(mm) ± SD(mm)	Segmentation MDOC(%) ± SDDOC(%)
Baseline	0.42 ± 0.29	90.48 ± 1.97
U-Net (Çiçek et al., 2016)	0.37 ± 0.20	90.97 ± 2.27
MsFCN	0.39 ± 0.24	91.17 ± 2.07
MsFCN-A + RMVD	0.37 ± 0.21	90.72 ± 2.54
MsFCN + RMVD	**0.36 ± 0.21**	**91.34** ± 2.16

Notice: MsFCN-A denotes the MsFCN framework with same input size in each pathway.

Figs. 7 and 8 show some examples of the segmentation results achieved by the baseline, MsFCN and MsFCN with RMVD on two different test subjects. In Fig. 7, we can clearly see that green contours are more consistent with red contours for most of IVDs in Fig. 7(b) than that in Fig. 7(a), especially for IVDs S1-L5, L5-L4, L2-L1, T12-T11. Also, it is observed in Fig. 8 that IVDs S1-L5, L2-L4, L1-T12, L5-L4 were segmented more accurately in Fig. 8(b) than

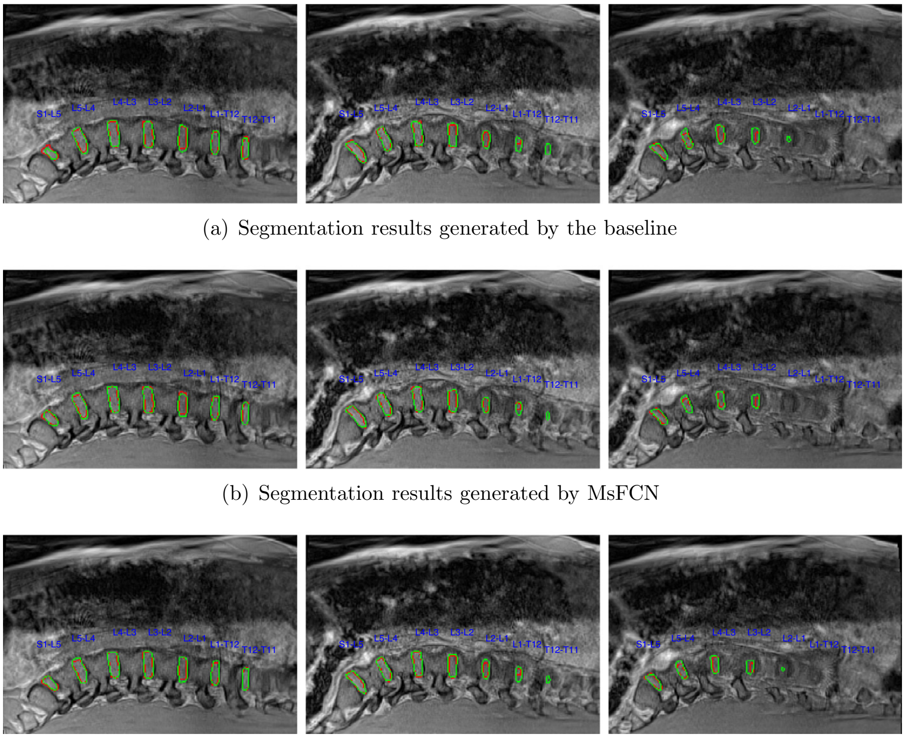


(a) Segmentation results generated by the baseline

(b) Segmentation results generated by MsFCN

(c) Segmentation results generated by MsFCN+RMVD

Fig. 7. Examples of segmentation results from one test subject B_S2. From left to right, each column shows the segmentation results on the 6th, 20th and 27th sagittal slices. (a), (b) and (c) are the segmentation results generated by the baseline, MsFCN and MsFCN+RMVD, respectively. Red: the ground-truth contour. Green: our segmentation contour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

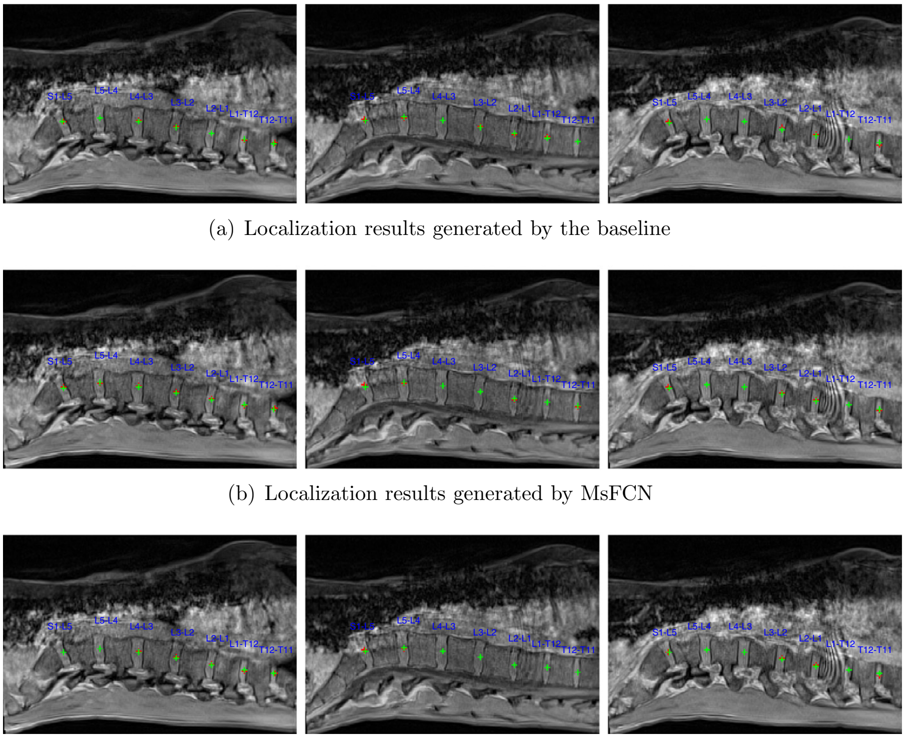


(a) Segmentation results generated by the baseline

(b) Segmentation results generated by MsFCN

(c) Segmentation results generated by MsFCN+RMVD

Fig. 8. Examples of segmentation results from test subject G_S2. From left to right, each column shows the segmentation results on the 10th, 28th and 30th sagittal slices. (a), (b) and (c) are the segmentation results generated by the baseline, MsFCN and MsFCN+RMVD, respectively. Red: the ground-truth contour. Green: the our segmentation contour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a) Localization results generated by the baseline

(b) Localization results generated by MsFCN

(c) Localization results generated by MsFCN+RMVD

Fig. 9. Examples of localization results from test subject B_S2. From left to right, each column shows the localization results on the 10th, 20th and 26th sagittal slices. (a), (b) and (c) are localization results generated by the baseline, MsFCN and MsFCN+RMVD, respectively. Red: the ground-truth. Green: our localization results. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

those generated by the baseline. These observations demonstrated that MsFCN is more powerful in the IVD segmentation than network with single-scale input image, indicating that MsFCN has the promising ability of discriminative feature generation. Examples of localization results on one test image were shown in Fig. 9.

Comparison of MsFCN and MsFCN+RMVD. From Table 5, it is observed that the results achieved by MsFCN+RMVD have higher segmentation and localization accuracy than MsFCN, with segmentation accuracy of 0.17% improvement on the MDOC metric and localization error of 0.03 mm improvement on the MLD metric. The performance improvement is attributed to that drop random voxels on multi-modality images in each training propagation can alleviate the co-adaption issue among multi-modality images, which benefits the optimization of multi-modality learning.

From the comparison in Fig. 7(b) and (c), green contours achieve a better consistency with red contours for most of IVDs in Fig. 7(c) than Fig. 7(b). The same situation can also be observed from the segmentation results on the other test images. For example, IVDs L4-L3 and L3-L2 were segmented more accurately in Fig. 8(c) than that in Fig. 8(b). Similarly, the comparison of localization results in Fig. 9(b) and (c) also indicates that MsFCN with RMVD attained better performance in the IVD localization than that without.

Comparison of MsFCN+RMVD and MsFCN-A+RMVD. To explore the key factor in our proposed MsFCN, we compare the MsFCN+RMVD with MsFCN-A+RMVD, where MsFCN-A denotes the MsFCN with input of the same-scale patches in each pathway. From this comparison, we can explore whether the different scale contexts contribute to the performance gains. Both two experiments were conducted under same environmental settings. From Table 5, it is observed that our method consistently surpassed the MsFCN-A counterpart, which indicates that the MsFCN improves the segmentation results by incorporating different scale contexts, not only the model combination.

Comparison of MsFCN+RMVD and 3D U-Net. We also compare our method with 3D U-Net (Çiçek et al., 2016) for volumetric segmentation, one of the most known frameworks in medical image community. To keep consistency with our architecture, we cropped raw images with size $32 \times 64 \times 64$. From Table 5, it is observed that our method consistently outperformed U-Net architecture. Compared with U-Net, our architecture has at least two advantages. First, we shed light on how to effectively extract features from multi-modality images, which has not been explored in U-Net. Second, our method incorporates different scales of contexts in the architecture and improves the feature representation ability, which also contributes to the performance gains.

3.5. System implementation

We implemented the proposed method with Python based on Theano [3] library on a workstation equipped with a GPU of Nvidia GeForce GTX Titan X. The networks were trained with Adam method (Kingma and Ba, 2014) (we set the batch size as 10 and the learning rate as 0.001 initially, and then gradually reduced the learning rate by a factor of 5 every 3000 iterations). The weights were randomly initialized from a Gaussian distribution ($\mu = 0$, $\sigma = 0.01$) and updated with a standard back-propagation. In the testing stage, the prediction results of the whole image were generated with an overlap-tile strategy. Leveraging the architecture of fully convolutional neural network, our method was quite efficient during the inference stage. It took about 10s on average to process one subject for IVD localization and segmentation. For the determi-

[3] Theano: http://deeplearning.net/software/theano.

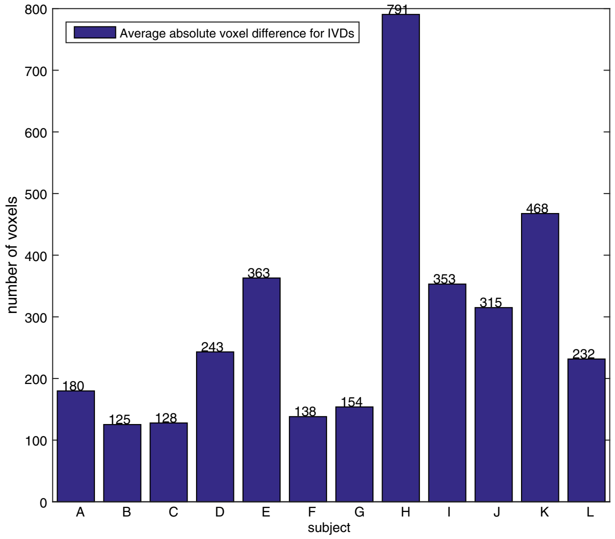


Fig. 10. Visualization of the average absolute voxel changes for each IVD in 12 subjects within two time points of the prolonged bed test study.

nation of dropout ratio, we tried three ranges of ratio, i.e., 0.1, 0.01, 0.001 and found the segmentation performance achieved the best performance when drop ratio was set as 0.01. Too large ratio may lead to contextual information loss and degrades the performance.

3.6. Two-stage study

The MICCAI challenge dataset was collected at two different time points of the prolonged bed rest study (Belavy et al., 2010) with the aims of understanding the effects of inactivity on the human body and to simulate the effects of microgravity on IVD morphology profile by space agencies (Belavy et al., 2010; 2011; 2012). Thus, it is important to know the changes of IVD morphology profile in different time points of the prolonged bed rest (spaceflight simulation). Fig. 10 presents the average absolute volume difference for IVD in each subject during two time points. It is observed that the dataset has large volume difference variations at two time points, which is a challenging dataset to evaluate the capability of our methods in modeling the morphology profile changes.

To measure the accuracy of our method on modeling the volume difference at two individual time points, we designed the absolute value of *relative volume difference* (arvd) metric:

$$arvd = \left| \frac{\left|V_{g1} - V_{t2}\right| - \left|V_{g1} - V_{g2}\right|}{V_{g1}} \right| \quad (4)$$

where V_{g1} and V_{g2} denote the IVD volume size on ground truth images collected at time point 1 and 2, respectively; V_{t2} is the produced IVD volume size from our method on test subjects collected at time point 2. The smaller the *arvd* is, the better performance our method achieves on modeling the changes. Table 6 lists the *arvd* value on the experiments which employs subjects at time point 1 as training data while subjects at time point 2 as test data. It is observed that the value of *arvd* is quite small in most subjects; this observation demonstrated that our method has promising performance in the volume changes prediction, and can help understand the effects of inactivity on the human body as well as the IVD simulation in microgravity environment. However, we observed that the *arvd* value in subject H is a bit high, due to that the IVD volume size in this subject is much larger than other subjects as shown in Fig. 10. This situation also presents in the 1st and 7th IVD in some subjects (e.g., subject E,J and K) since T12-T11, S1-L5 and L1-T12 IVDs are usually small. The results in Table 7 present the high accuracy for IVD localization and segmentation prediction achieved by our method in spine images on IVD morphology profiling, demonstrating the excellent robustness of our method.

4. Discussion

IVD localization and segmentation have great significance in spine pathologies diagnosis. For example, IVD degeneration has been proven to be associated with the low back pain (LBP), one of the most prevalent health problems amongst the world's population, in many clinical studies (Fraser et al., 1993; Luoma et al., 2000; Kjaer et al., 2005). However, in current clinical routine, the manual labeling is time-consuming, laborious and error-prone. To relieve the workload of radiologists, we proposed an integrated 3D multi-scale FCN with random modality voxel dropout learning for IVD localization and segmentation framework. Regarding to the random modality voxel dropout strategy, to the best of our knowledge, this is the first study that explores modality related method in clinical IVD segmentation applications from multi-modality images. Extensive experimental results demonstrated the efficacy and robustness of our method. It also surpassed other methods presented in *MICCAI 2016 IVD localization and segmentation challenge*.

In the training of deep neural networks, it usually demands a large number of training samples due to the plenty of parameters in the network. The larger amount of training data can contribute to better segmentation and localization results. We trained the network with *Training set of IVD challenge* during the on-site competition as shown in Table 3 while used the *Training set of IVD challenge* and *Additional training set* as training data in the experiments shown in Table 5. Compared with the results achieved in experiments shown in Table 3, network with additional training set input has 0.26 mm relative improvement on the MLD metric for localization and 0.14% relative improvement on the MDOC metric for segmentation, indicating that more training data can contribute to better segmentation and localization performance (see Table 5). From these results, we anticipate that the performance of our method will be further improved with more training data, but the space of performance improvement is becoming diminished.

Although our method achieved appealing localization and segmentation results in most cases, there are still some limitations. Firstly, it is observed in Fig. 11 that the unsatisfactory segmentations occur at boundary slices (the start and the end slice of IVDs). The blurred and noisy IVDs as well as the lack of spatial information on these positions lead to the low accuracy of the segmentation. In the future, we shall investigate how to utilize some image processing techniques (e.g., image deblurring and image normalization) to further improve the performance. Moreover, it is important to acquire 3D MR data with a high resolution on the third dimension. With a large number of slices available in each 3D volume, the more detailed structure can be shown on the third dimension, which would also contribute to the performance gains. Furthermore, due to the limited data used in our longitudinal study, we conducted an experiment using dataset acquired at one time point as the training data and the dataset at the other time point as the test data. A more realistic setup would be to conduct a leave-one-subject-out study if a large number of longitudinal data would be available. Nevertheless, the results as we reported in Table 7 demonstrate the potential of our method in tracking morphological changes in a longitudinal study.

Another clinical importance is that our method holds the potential to save time and manual cost, and allow for a true 3D quantification avoiding problems caused by 2D measurements. Previous studies (Belavy et al., 2011; 2012) have shown that changes in IVD morphology profile persisted 5 months after 21-day bed rest and that the recovery of the lumbar intervertebral discs after 60-day bed rest was a prolonged process and incomplete within 2 years.

Table 6
The *arvd* value between ground-truth IVDs of time point 1 and segmented IVDs of time point 2 in the prolonged bed rest study (12 training samples and 12 test samples).

Subject												
IVD index	A	B	C	D	E	F	G	H	I	J	K	L
1	2.9%	2.7%	1.0%	0.8%	11.0%	0.78%	10.7%	11.7%	3.4%	8.8%	10.3%	0.1%
2	2.1%	1.6%	1.4%	2.5%	5.9%	0.5%	1.7%	7.1%	13.2%	1.6%	1.3%	0.1%
3	0.8%	0.2%	2.6%	1.2%	6.6%	4.3%	2.3%	14.6%	10.9%	3.6%	0.04%	0.3%
4	1.9%	5.2%	4.4%	0.4%	4.4%	1.5%	1.8%	17.9%	7.1%	8.2%	0.6%	5.4%
5	2.0%	4.8%	8.3%	4.0%	4.3%	1.6%	2.1%	21.5%	4.8%	2.3%	4.1%	3.5%
6	3.7%	2.4%	4.8%	2.7%	4.1%	3.8%	7.1%	17.4%	4.2%	1.6%	8.0%	0.1%
7	1.6%	21.9%	6.7%	3.3%	14.8%	2.1%	9.4%	21.1%	4.4%	2.7%	8.7%	5.7%

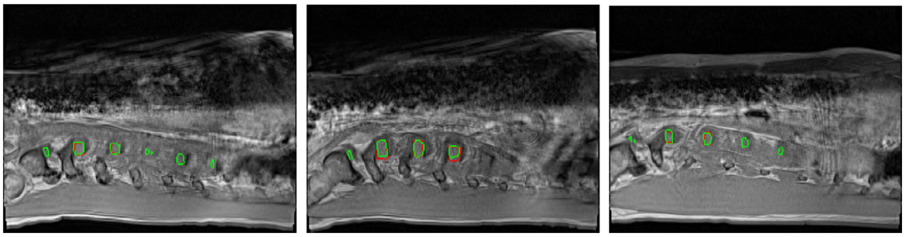


Fig. 11. Worse cases generated by our method.

Table 7
Evaluation of IVD localization and segmentation on datasets acquired from two different time points (12 training samples and 12 test samples).

Training data	Test data	Localization MLD(mm) ± SD(mm)	Segmentation MDOC(%) ± SDDOC(%)
Time point 1	Time point 2	0.47 ± 0.45	90.56 ± 2.48
Time point 2	Time point 1	0.46 ± 0.34	90.48 ± 0.02

The limitation of these studies, however, lies in the tools that they used to measure the IVD morphology changes. At this moment, clinicians lack tools to conduct a true 3D quantification even when 3D MR image data are available. Instead, they seek to use 2D surrogate measurements measured from selected 2D slices to quantify 3D IVD morphology (Belavy et al., 2011; 2012). In some extent, our proposed method may alleviate manual labor work and also advanced the 3D quantification of the IVD morphology changes.

5. Conclusion

In this paper, we present a novel system that achieved state-of-the-art IVD segmentation performance from multi-modality images. Compared with network trained with single-scale context image, the proposed 3D multi-scale FCN can generate features with high discrimination capability, and hence improve the performance on IVD localization and segmentation tasks. We also employed random modality voxel dropout strategy in the training phase to further effectively integrate the multi-modality information and enhance the learning capability. The results of 2016 MICCAI challenge on IVD localization and segmentation demonstrated the effectiveness of our proposed method. Extensive experiments were conducted to validate the robustness of our method on morphology profiling volume changes at individual collection time points. This plays an important role in the study of the effects of inactivity as well as the IVD simulation in microgravity environment at different time points.

Acknowledgments

This work is supported by Hong Kong Research Grants Council General Research Fund (Project No. GRF 14203115 and Project No. CUHK 14202514) and a grant from the National Natural Science Foundation of China (Project No. 61233012). The challenge data preparation is partially supported by the Swiss National Science Foundation Project no. 205321 - 1572071. The authors also gratefully thank the challenge organizers for helping the evaluation.

References

An, H.S., Anderson, P.A., Haughton, V.M., Iatridis, J.C., Kang, J.D., Lotz, J.C., Natarajan, R.N., Oegema Jr, T.R., Roughley, P., Setton, L.A., 2004. Introduction: disc degeneration: summary. Spine 29 (23), 2677–2678.

Ayed, I.B., Punithakumar, K., Garvin, G., Romano, W., Li, S., 2011. Graph cuts with invariant object-interaction priors: application to intervertebral disc segmentation. In: Biennial International Conference on Information Processing in Medical Imaging. Springer, pp. 221–232.

Belavy, D.L., Armbrecht, G., Felsenberg, D., 2012. Incomplete recovery of lumbar intervertebral discs 2 years after 60-day bed rest. Spine 37 (4), 1245–1251.

Belavy, D.L., Bansmann, P.M., Böhme, G., Frings-Meuthen, P., Heer, M., Rittweger, J., Zange, J., Felsenberg, D., 2011. Changes in intervertebral disc morphology persist 5 mo after 21-day bed rest. J. Appl. Physiol. 111 (5), 1304–1314.

Belavy, D.L., Bock, O., Börst, H., 2010. The 2nd berlin bedrest study: protocol and implementation. J. Musculoskeletal Neuronal interact. 10 (3), 207–219.

BenEliyahu, D.J., 1995. Magnetic resonance imaging and clinical follow-up: study of 27 patients receiving chiropractic care for cervical and lumbar disc herniations. J. Manipulative. Physiol. Ther. 19 (9), 597–606.

Cai, Y., Landis, M., Laidley, D.T., Kornecki, A., Lum, A., Li, S., 2016. Multi-modal vertebrae recognition using transformed deep convolution network. Comput. Med. Imaging Graphics 51, 11–19.

Cai, Y., Osman, S., Sharma, M., Landis, M., Li, S., 2015. Multi-modality vertebra recognition in arbitrary views using 3d deformable hierarchical model. IEEE Trans. Med. Imaging 34 (8), 1676–1693.

Carballido-Gamio, J., Belongie, S.J., Majumdar, S., 2004. Normalized cuts in 3-d for spinal mri segmentation. IEEE Trans. Med. Imaging 23 (1), 36–44.

Chen, C., Belavy, D., Yu, W., Chu, C., Armbrecht, G., Bansmann, M., Felsenberg, D., Zheng, G., 2015a. Localization and segmentation of 3d intervertebral discs in mr images by data driven estimation. IEEE Trans. Med. Imaging 34 (8), 1719–1729.

Chen, C., Belavy, D., Zheng, G., 2014. 3d intervertebral disc localization and segmentation from mr images by data-driven regression and classification. In: Proceedings of MICCAI-MLMI 2014. Springer, pp. 50–58.

Chen, H., Dou, Q., Wang, X., Qin, J., Cheng, J.C., Heng, P.A., 2016a. 3d fully convolutional networks for intervertebral disc localization and segmentation. In: International Conference on Medical Imaging and Virtual Reality. Springer, pp. 375–382.

Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.-A., 2017a. Voxresnet: deep voxelwise residual networks for brain segmentation from 3d mr images. NeuroImage.

Chen, H., Ni, D., Qin, J., Li, S., Yang, X., Wang, T., Heng, P.A., 2015b. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE J. Biomed. Health Inf. 19 (5), 1627–1636.

Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., Heng, P.-A., 2017b. Dcan: deep contour-aware networks for object instance segmentation from histology images. Med. Image Anal. 36, 135–146.

Chen, H., Shen, C., Qin, J., Ni, D., Shi, L., Cheng, J.C., Heng, P.A., 2015c. Automatic localization and identification of vertebrae in spine ct via a joint learning model with deep neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 515–522.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L., 2016b. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv:1606.00915.

Chen, L.-C., Yang, Y., Wang, J., Xu, W., Yuille, A.L., 2016c. Attention to scale: Scale-aware semantic image segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3640–3649.

Chevrefils, C., Cheriet, F., Aubin, C.-É., Grimard, G., 2009. Texture analysis for automatic segmentation of intervertebral disks of scoliotic spines from mr images. IEEE Trans. Inf. Technol. Biomed. 13 (4), 608–620.

Chevrefils, C., Chériet, F., Grimard, G., Aubin, C.-E., 2007. Watershed segmentation of intervertebral disk and spinal canal from mri images. In: International Conference Image Analysis and Recognition. Springer, pp. 1017–1027.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 424–432.

Corso, J.J., Raja'S, A., Chaudhary, V., 2008. Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 202–210.

Dou, Q., Yu, L., Chen, H., Jin, Y., Yang, X., Qin, J., Heng, P.-A., 2017. 3d deeply supervised network for automated segmentation of volumetric medical images. Med. Image Anal.

Fraser, R., Osti, O., Vernon-Roberts, B., 1993. Intervertebral disc degeneration. Eur. Spine J. 1 (4), 205–213.

Gao, Y., Shen, D., 2014. Context-aware anatomical landmark detection: application to deformable model initialization in prostate ct images. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 165–173.

Greenspan, H., van Ginneken, B., Summers, R.M., 2016. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans. Med. Imaging 35 (5), 1153–1159.

Hamanishi, C., Matukura, N., Fujita, M., Tomihara, M., Tanaka, S., 1994. Cross-sectional area of the stenotic lumbar dural tube measured from the transverse views of magnetic resonance imaging. J. Spinal Disord. Tech. 7 (5), 388–393.

Haq, R., Besachio, D.A., Borgie, R.C., Audette, M.A., 2014. Using shape-aware models for lumbar spine intervertebral disc segmentation. In: Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, pp. 3191–3196.

Havaei, M., Guizard, N., Chapados, N., Bengio, Y., 2016. Hemis: hetero-modal image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 469–477.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Heinrich, M.P., Oktay, O., 2016. Accurate intervertebral disc localisation and segmentation in mri using vantage point hough forests and multi-atlas fusion. In: International Workshop on Computational Methods and Clinical Applications for Spine Imaging. Springer, pp. 77–84.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv:1207.0580.

Huang, S.-H., Chu, Y.-H., Lai, S.-H., Novak, C.L., 2009. Learning-based vertebra detection and iterative normalized-cut segmentation for spinal mri. IEEE Trans. Med. Imaging 28 (10), 1595–1605.

Jamaludin, A., Kadir, T., Zisserman, A., 2016. Spinenet: Automatically pinpointing classification evidence in spinal mris. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 166–175.

Ji, X., Zheng, G., Belavy, D., Ni, D., 2016a. Automated intervertebral disc segmentation using deep convolutional neural networks. In: International Workshop on Computational Methods and Clinical Applications for Spine Imaging. Springer, pp. 38–48.

Ji, X., Zheng, G., Liu, L., Ni, D., 2016b. Fully automatic localization and segmentation of intervertebral disc from 3d multi-modality mr images by regression forest and cnn. In: International Workshop on Computational Methods and Clinical Applications for Spine Imaging. Springer, pp. 92–101.

Kamnitsas, K., Chen, L., Ledig, C., Rueckert, D., Glocker, B., 2015. Multi-scale 3d convolutional neural networks for lesion segmentation in brain mri. Ischemic Stroke Lesion Segmentation 13.

Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78.

Kelm, B.M., Wels, M., Zhou, S.K., Seifert, S., Suehling, M., Zheng, Y., Comaniciu, D., 2013. Spine detection in ct and mr using iterated marginal space learning. Med. Image Anal. 17 (8), 1283–1292.

Kingma, D., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv:1412.6980.

Kjaer, P., Leboeuf-Yde, C., Korsholm, L., Sorensen, J.S., Bendix, T., 2005. Magnetic resonance imaging and low back pain in adults: a diagnostic imaging study of 40-year-old men and women. Spine 30 (10), 1173–1180.

Kong, B., Zhan, Y., Shin, M., Denny, T., Zhang, S., 2016. Recognizing end-diastole and end-systole frames via deep temporal regression network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 264–272.

Korez, R., Ibragimov, B., Likar, B., Pernuš, F., Vrtovec, T., 2015. Deformable model-based segmentation of intervertebral discs from mr spine images by using the ssc descriptor. In: International Workshop on Computational Methods and Clinical Applications for Spine Imaging. Springer, pp. 117–124.

Law, M.W., Tay, K., Leung, A., Garvin, G.J., Li, S., 2013. Intervertebral disc segmentation in mr images using anisotropic oriented flux. Med. Image Anal. 17 (1), 43–61.

Li, X., Dou, Q., Chen, H., Fu, C.-W., Heng, P.A., 2016a. Multi-scale and modality dropout learning for intervertebral disc localization and segmentation. In: International Workshop on Computational Methods and Clinical Applications for Spine Imaging. Springer, pp. 85–91.

Li, Z., Gong, B., Yang, T., 2016b. Improved dropout for shallow and deep learning. In: Advances In Neural Information Processing Systems, pp. 2523–2531.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.

Luoma, K., Riihimäki, H., Luukkonen, R., Raininko, R., Viikari-Juntura, E., Lamminen, A., 2000. Low back pain in relation to lumbar disc degeneration. Spine 25 (4), 487–492.

Misri, R., 2013. Multimodality imaging. Future Med. 162–176.

Neubert, A., Fripp, J., Shen, K., Salvado, O., Schwarz, R., Lauer, L., Engstrom, C., Crozier, S., 2011. Automated 3d segmentation of vertebral bodies and intervertebral discs from mri. In: Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on. IEEE, pp. 19–24.

Niemeläinen, R., Videman, T., Dhillon, S., Battié, M., 2008. Quantitative measurement of intervertebral disc signal using mri. Clin. Radiol. 63 (3), 252–255.

Raja'S, A., Corso, J.J., Chaudhary, V., 2011. Labeling of lumbar discs using both pixel-and object-level features with a two-level probabilistic model. IEEE Trans. Med. Imaging 30 (1), 1–10.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Roth, H.R., Lu, L., Seff, A., Cherry, K.M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., Summers, R.M., 2014. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 520–527.

Schmidt, S., Kappes, J., Bergtholdt, M., Pekar, V., Dries, S., Bystrov, D., Schnörr, C., 2007. Spine detection and labeling using a parts-based graphical model. In: Biennial International Conference on Information Processing in Medical Imaging. Springer, pp. 122–133.

Schneiderman, G., Flannigan, B., Kingston, S., Thomas, J., Dillin, W.H., Watkins, R.G., 1987. Magnetic resonance imaging in the diagnosis of disc degeneration: correlation with discography. Spine 12 (3), 276–281.

Shen, D., Wu, G., Suk, H.-I., 2017. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. (0).

Shi, R., Sun, D., Qiu, Z., Weiss, K.L., 2007. An efficient method for segmentation of mri spine images. In: Complex Medical Engineering, 2007. CME 2007. IEEE/ICME International Conference on. IEEE, pp. 713–717.

Shin, H.-C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging 35 (5), 1285–1298.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.

Sirinukunwattana, K., Raza, S.E.A., Tsang, Y.-W., Snead, D., Cree, I., Rajpoot, N., 2015. A spatially constrained deep learning framework for detection of epithelial tumor nuclei in cancer histology images. In: International Workshop on Patch-based Techniques in Medical Imaging. Springer, pp. 154–162.

Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Sun, C., Guo, S., Zhang, H., Li, J., Chen, M., Ma, S., Jin, L., Liu, X., Li, X., Qian, X., 2017. Automatic segmentation of liver tumors from multiphase contrast-enhanced ct images based on fcns. Artif. Intell. Med..

Suzani, A., Seitel, A., Liu, Y., Fels, S., Rohling, R.N., Abolmaesumi, P., 2015. Fast automatic vertebrae detection and localization in pathological ct scans-a deep learning approach. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 678–686.

Tertti, M., Paajanen, H., Laato, M., Aho, H., Komu, M., Kormano, M., 1991. Disc degeneration in magnetic resonance imaging: a comparative biochemical, histologic, and radiologic study in cadaver spines. Spine 16 (6), 629–634.

Urban, J.P., Roberts, S., 2003. Degeneration of the intervertebral disc. Arthritis Res. Ther. 5.

Violas, P., Estivalezes, E., Briot, J., de Gauzy, J.S., Swider, P., 2007. Objective quantification of intervertebral disc volume properties using mri in idiopathic scoliosis surgery. Magn. Reson. Imaging 25 (3), 386–391.

Wan, L., Zeiler, M., Zhang, S., Cun, Y.L., Fergus, R., 2013. Regularization of neural networks using dropconnect. In: Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp. 1058–1066.

Wang, Z., Zhen, X., Tay, K., Osman, S., Romano, W., Li, S., 2015. Regression segmentation for spinal images. IEEE Trans. Med. Imaging 34 (8), 1640–1648.

Yao, J., Vrtovec, T., Zheng, G., Frangi, A., Glocker, B., Li, S., 2017. Computational Methods and Clinical Applications for Spine Imaging: 4th International Workshop and Challenge, CSI 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers, 10182. Springer.

Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., Shen, D., 2015. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. NeuroImage 108, 214–224.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2016. Pyramid scene parsing network. arXiv:1612.01105.

Zheng, G., Chu, C., Belavỳ, D.L., Ibragimov, B., Korez, R., Vrtovec, T., Hutt, H., Everson, R., Meakin, J., Andrade, I.L., et al., 2017. Evaluation and comparison of 3d intervertebral disc localization and segmentation methods for 3d t2 mr data: a grand challenge. Med. Image Anal. 35, 327–344.

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D., 2008. Four-chamber heart modeling and automatic segmentation for 3-d cardiac ct volumes using marginal space learning and steerable features. IEEE Trans. Med. Imaging 27 (11), 1668–1681.