Multi-scale and Modality Dropout Learning for Intervertebral Disc Localization and Segmentation

Xiaomeng Li[✉], Qi Dou, Hao Chen, Chi-Wing Fu, and Pheng-Ann Heng

Department of Computer Science and Engineering,
The Chinese University of Hong Kong, Hong Kong, China
`xmli@cse.cuhk.edu.hk`

Abstract. Automatic localization and segmentation of intervertebral discs (IVDs) from volumetric magnetic resonance (MR) images is important for spine disease diagnosis. It dramatically alleviates the workload of radiologists given that the traditional manual annotation is time-consuming and error-prone with limited reproducibility. Compared with single modality data, multi-modality MR images are able to provide complementary information. However, how to effectively integrate them to generate more accurate segmentation results still remains open for studies. In this paper, we introduce a multi-scale and modality dropout learning framework to segment IVDs from four-modality MR images. Specifically, we design a 3D fully convolutional network which takes multiple scales of images as input and merges these pathways at higher layers to jointly integrate multi-scale information. Furthermore, in order to harness the complementary information from different modalities, we propose a modality dropout strategy to alleviate the co-adaption issue during the training. We evaluated our method on the *MICCAI 2016 Challenge on Automatic Intervertebral Disc Localization and Segmentation from 3D Multi-modality MR Images*. Our method achieved the best overall performance with the mean segmentation Dice as 91.2% and localization error as 0.62 mm, which demonstrated the superiority of our proposed framework.

1 Introduction

Accurate localization and segmentation of intervertebral discs (IVDs) from volumetric magnetic resonance (MR) images plays an important role for spine disease related diagnosis. Automatic localization and segmentation of IVDs are quite challenging due to the large intra-class variations and similar appearance among different IVDs.

Previous methods segmented the IVDs by employing hand-crafted features which were derived based on intensity and shape information [2,8,12]. However, these hand-crafted features tend to suffer from limited representation capability compared with the automatically learned features. Furthermore, these methods were usually performed based on 2D slices which might neglect the volumetric
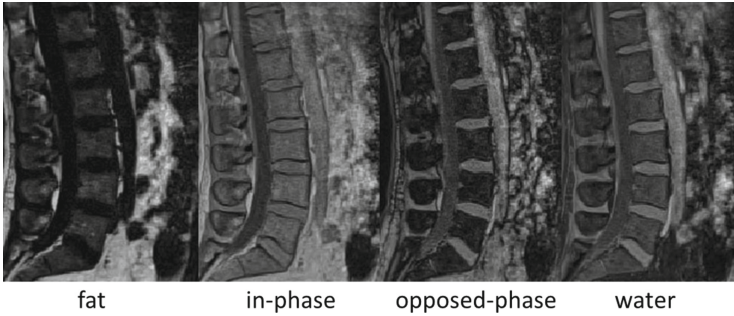
fat   in-phase   opposed-phase   water

Fig. 1. Illustration of IVD appearance in multi-modality MR images.

spatial contexts, thus degrading the performance. Recently, deep learning based methods have been proposed to directly localize and segment IVDs or vertebrae from volumetric data [4,7,10,14]. For example, Jamaludin *et al.* [10] proposed a convolutional neural network (CNN) framework to automatically label each disc and the surrounding vertebrae with a number of radiological scores. Chen *et al.* [5] 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.

Those previous works employed single modality MR data instead of taking multi-modality information into consideration, which would limit the localization and segmentation accuracy. Multi-modality MR images (see Fig. 1) collected by setting different scanning configurations can provide comprehensive information for robust diagnosis and treatment. Previous studies on brain segmentation indicated that multi-modality data could help to improve the segmentation performance significantly [3,6,15]. Meanwhile, incorporating multi-scale information into the learning process can further improve the performance [6,11].

In these regards, we propose a 3D multi-scale and modality dropout learning framework for localizing and segmenting IVDs from multi-modality MR images. Our contribution in this paper is twofold. First, we propose a novel multi-scale 3D fully convolutional network which consists of three pathways to integrate multiple scales of spatial information. Second, we propose a modality drop strategy for harnessing the complementary information from multi-modality MR data. Experimental results on the *MICCAI 2016 Challenge on Automatic Intervertebral Disc Localization and Segmentation from 3D Multi-modality MR Images* have demonstrated the superiority of our proposed framework.

2   Method

Figure 2 presents an overview of our proposed multi-scale and modality dropout learning framework based on multi-modality MR images. Our multi-scale fully convolutional network consists of three pathways with each inputting a different scale of volumetric image. In each training iteration, modality dropout strategy is
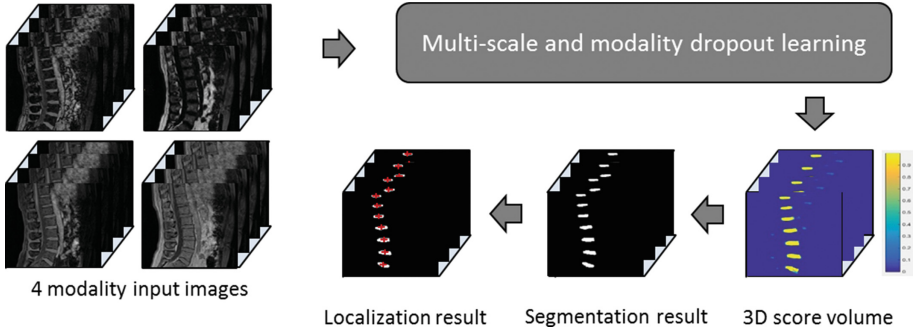
Fig. 2. An overview of our proposed multi-scale and modality dropout learning framework for IVDs segmentation and localization from multi-modality MR images

used on the input multi-modality data in order to reduce the feature co-adaption and encourage each single modality image to provide discriminative information.

2.1 Multi-scale FCN Architecture

One limitation of previous methods for IVDs segmentation is that they usually considered a single scale of spatial information surrounding the discs. However, multi-scale contextual information can contribute to better recognition performance. With this consideration, we employ multi-scale fully convolutional neural network with different scales of input data volumes. Figure 3 shows details of our proposed architecture, indicating input patch sizes, construction of layers, and kernel size and numbers. This multi-scale architecture consists of three pathways corresponding to different input volume sizes. During the training phase, three selected modality volumes (with one modality being randomly dropped) are input to the architecture. A 3D probability map with voxelwise predictions is generated as the output of the network. The final segmentation results can be determined from the score volume while the localization results can be generated as the centroids of the segmentation masks. In the experiments, we observe that the number of IVD voxels is much less than background voxels. To deal with the problem of imbalanced training samples, we employed weighted loss function during the training process, as shown in the following:

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

where w is the weight for strengthening the importance of foreground voxels. N denotes the total number of voxels in each training process, t_i denotes the label at voxel i and $p(x_i)$ denotes the corresponding prediction for voxel x_i.

2.2 Dropout Modality Learning

Dropout technique was proposed in [9,13] and it has been recognized as an effective way to prevent co-adaption of feature detectors and alleviate the overfitting
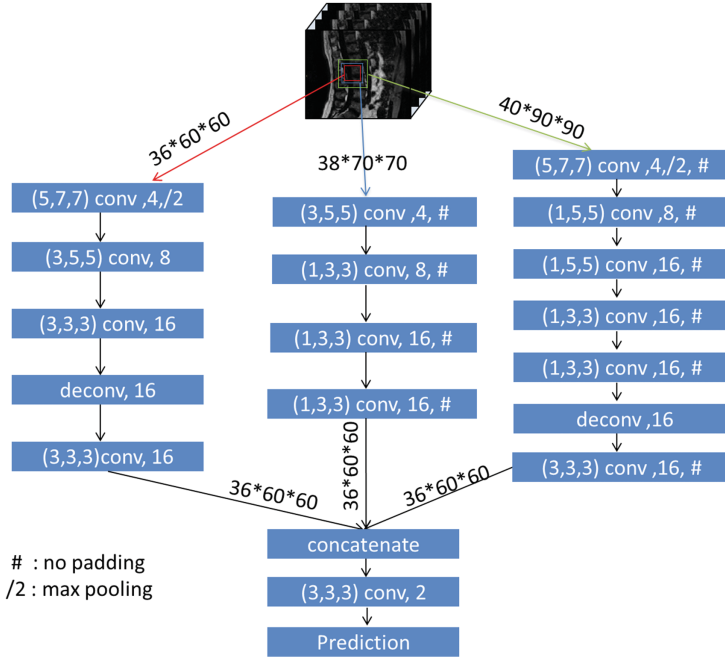
Fig. 3. The architecture of our proposed 3D multi-scale FCN. The red, blue and green boxes represent different scales of input to three different pathways. We only include one modality in this figure for clear illustration of the multi-scale framework. In experiments, the inputs are actually multi-modality images. (Color figure online)

problem. In our task of IVD localization and segmentation from multi-modality MR images, an intuitive approach is to input all modality data into the network for training. However, training four modality volumes all together may cause too much dependency among modalities, which leads to feature co-adaption and thus degrades the performance. Therefore, in order to fully take advantage of the complementary information from different modalities, we randomly dropped one modality during each training iteration to break the co-adaption and encourage harnessing discriminative information from remaining modalities. This can be regarded as a regularization on the optimization of neural networks. In the testing phase, we took all the four modality images as the input and generated the final segmentation and localization results.

3   Experiment

3.1   Dataset and Preprocessing

We evaluated our method on the dataset from *2016 MICCAI Challenge on Automatic Intervertebral Disc Localization and Segmentation from 3D Multi-modality*

MR Images [1]. The dataset was collected from a study investigating the effects of prolonged bed rest on lumbar intervertebral discs. The training data contains volumetric images from 8 patients and each subject consists of four modality MR datasets, i.e., in-phase, opposed-phase, fat and water. There are at least 7 IVDs in each image (size $36 \times 256 \times 256$). These multi-modality images of each subject are well registered and one binary mask is provided by manual annotation from radiologists. The testing data includes 6 subjects with ground truth held out by the organizers for independent evaluation.

3.2   On-site Competition Results

The evaluation metric for IVDs localization is mean localization distance (MLD) with standard deviation (SD), where MLD measures the accuracy of localization and SD quantifies the degree of variation. For IVDs segmentation evaluation, Mean Dice Overlap Coefficients (MDOC) and standard deviation (SDDOC) are used to measure the accuracy and variation of segmentation results. Mean Average Absolute Distance (MASD) with standard deviation (SDASD) is another measurement for evaluating segmentation accuracy. More details can be found on the challenge website [1]. Table 1 and Fig. 4 show the on-site challenge results. Our method achieved the performance of MDOC as 91.2% and MLD as 0.62 mm, which demonstrated the superiority of our proposed framework. We achieved the first place out of 3 teams during the on-site challenge according to the overall performance of these measurements.

Table 1. IVDs localization and segmentation results of our method in on-site challenge.

	MLD (mm)	SD	MDOC (%)	SDDOC	MASD (mm)	SDASD	Rank
Ours	0.62	0.38	91.2	1.8	1.26	1.22	1

fat   In-phase   opposed-phase   water   segmentation result
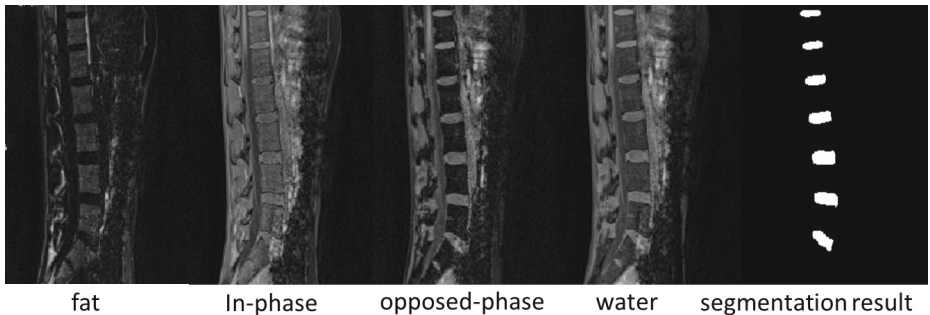
Fig. 4. Example of on-site challenge results from one testing patient. We show one slice of the 3D volumetric data for clear visualization.

4   Conclusion

In this paper, we proposed a novel 3D multi-scale and dropout modality learning method for IVDs localization and segmentation from multi-modality images. Experimental results on the challenge demonstrated the advantage of our proposed method, which is inherently general and can be applied in other multi-modality image segmentation tasks. Future work includes shape regression based methods to further improve the performance and applying our method on larger dataset.

References

1. http://ivdm3seg.weebly.com
2. Ben Ayed, I., Punithakumar, K., Garvin, G., Romano, W., Li, S.: Graph cuts with invariant object-interaction priors: application to intervertebral disc segmentation. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 221–232. Springer, Heidelberg (2011). doi:10.1007/978-3-642-22092-0_19
3. Cai, Y., Landis, M., Laidley, D.T., Kornecki, A., Lum, A., Li, S.: Multi-modal vertebrae recognition using transformed deep convolution network. Comput. Med. Imaging Graph. **51**, 11–19 (2016)
4. Chen, C., Belavy, D., Yu, W., Chu, C., Armbrecht, G., Bansmann, M., Felsenberg, D., Zheng, G.: Localization and segmentation of 3D intervertebral discs in mr images by data driven estimation. IEEE Trans. Med. Imaging **34**(8), 1719–1729 (2015)
5. Chen, H., Dou, Q., Wang, X., Qin, J., Cheng, J.C.Y., Heng, P.-A.: 3D fully convolutional networks for intervertebral disc localization and segmentation. In: Zheng, G., Liao, H., Jannin, P., Cattin, P., Lee, S.-L. (eds.) MIAR 2016. LNCS, vol. 9805, pp. 375–382. Springer, Cham (2016). doi:10.1007/978-3-319-43775-0_34
6. Chen, H., Dou, Q., Yu, L., Heng, P.A.: VoxResNet: deep voxelwise residual networks for volumetric brain segmentation. arXiv preprint arXiv:1608.05895 (2016)
7. Chen, H., Shen, C., Qin, J., Ni, D., Shi, L., Cheng, J.C.Y., Heng, P.-A.: Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 515–522. Springer, Cham (2015). doi:10.1007/978-3-319-24553-9_63
8. Chevrefils, C., Chériet, F., Grimard, G., Aubin, C.-E.: Watershed segmentation of intervertebral disk and spinal canal from MRI images. In: Kamel, M., Campilho, A. (eds.) ICIAR 2007. LNCS, vol. 4633, pp. 1017–1027. Springer, Heidelberg (2007). doi:10.1007/978-3-540-74260-9_90
9. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012)
10. Jamaludin, A., Kadir, T., Zisserman, A.: SpineNet: automatically pinpointing classification evidence in spinal MRIs. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 166–175. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_20
11. Kamnitsas, K., Chen, L., Ledig, C., Rueckert, D., Glocker, B.: Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI. Ischemic Stroke Lesion Segmen. 13 (2015)

12. Law, M.W., Tay, K., Leung, A., Garvin, G.J., Li, S.: Intervertebral disc segmentation in MR images using anisotropic oriented flux. Med. Image Anal. **17**(1), 43–61 (2013)
13. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
14. Wang, Z., Zhen, X., Tay, K., Osman, S., Romano, W., Li, S.: Regression segmentation for spinal images. IEEE Trans. Med. Imaging **34**(8), 1640–1648 (2015)
15. Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., Shen, D.: Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. NeuroImage **108**, 214–224 (2015)