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

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

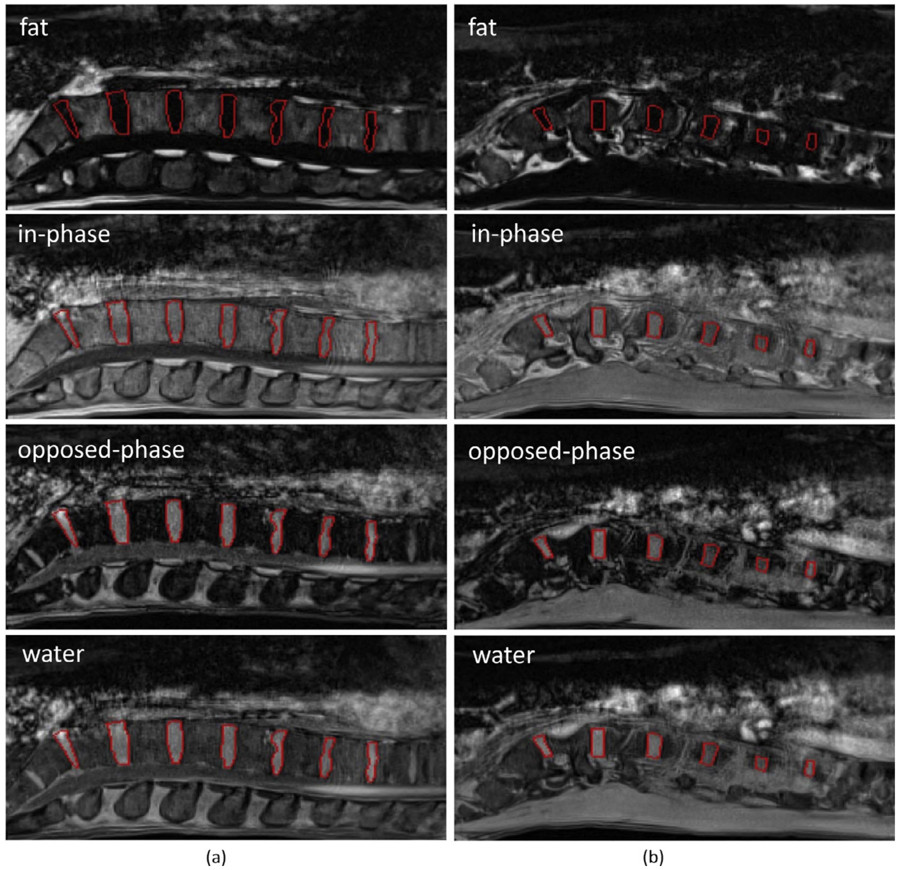

Fig. 1. Examples of 3D multi-modality input data. (a) and (b) show two data sets scanned from two different patients, each including four 3D modalities: fat, in-phase, opposed-phase, and water (top to bottom). In these figures, we show the 18th slice in the 3D images; red contours indicate the boundary 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