3D Fully Convolutional Networks for Intervertebral Disc Localization and Segmentation

Hao Chen[1(✉)], Qi Dou[1], Xi Wang[2], Jing Qin[3], Jack C.Y. Cheng[4], and Pheng-Ann Heng[1]

[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China
hchen@cse.cuhk.edu.hk
[2] College of Computer Science, Sichuan University, Chengdu, China
[3] School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China
[4] Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong, China

Abstract. Accurate localization and segmentation of intervertebral discs (IVDs) from volumetric data is a pre-requisite for clinical diagnosis and treatment planning. With the advance of deep learning, 2D fully convolutional networks (FCN) have achieved state-of-the-art performance on 2D image segmentation related tasks. However, how to segment objects such as IVDs from volumetric data hasn't been well addressed so far. In order to resolve above problem, we extend the 2D FCN into a 3D variant with end-to-end learning and inference, where voxel-wise predictions are generated. In order to compare the performance of 2D and 3D deep learning methods on volumetric segmentation, two different frameworks are studied: one is a 2D FCN with deep feature representations by making use of adjacent slices, the other one is a 3D FCN with flexible 3D convolutional kernels. We evaluated our methods on the 3D MRI data of *MICCAI 2015 Challenge on Automatic Intervertebral Disc Localization and Segmentation*. Extensive experimental results corroborated that 3D FCN can achieve a higher localization and segmentation accuracy than 2D FCN, which demonstrates the significance of volumetric information when confronting 3D localization and segmentation tasks.

1 Introduction

Accurate localization of intervertebral discs (IVDs) is a pre-requisite for quantitative diagnosis and treatment planning of various spinal pathologies [1]. The aim of the localization task is to identify the center of each IVD, while delineating the regions of IVD for the segmentation task. In clinical practice, the IVDs are manually segmented by radiologists thus suffers from the drawbacks of considerable efforts, time-consuming and error-prone [2]. Therefore, automatic approaches are highly demanded to alleviate the workload and improve the efficiency as well

H. Chen and Q. Dou—Authors contributed equally.

G. Zheng et al. (Eds.): MIAR 2016, LNCS 9805, pp. 375–382, 2016.
DOI: 10.1007/978-3-319-43775-0_34

as reliability. Nevertheless, the automatic localization and segmentation of IVDs from volumetric images is quite challenging for several reasons. First, IVDs often carry similar morphological appearance due to the repetitive nature of spine, which makes the labeling of individual IVD difficult. Second, the existence of similar anatomical structures or image artifacts would impede the localization and segmentation process. Third, the large shape variations of IVDs among different subjects make the robust localization and segmentation more challenging.

Many researchers have devoted their efforts on this challenging problem. Previous methods commonly utilized hand-crafted features (such as Haar features [14] and HOG [13]) for localizing the IVDs or vertebrae by different classifiers including random forests [6] and Adaboost classifier with geometric constraints [14]. Although considerable progress has been achieved, the employed low-level features are over-specified with a limited representation capability. A joint 2D learning model leveraging feature representations learned from deep convolutional networks was proposed in [3] to localize and identify the centroid of vertebrae from Computed Tomography (CT) data. Recently, 2D fully convolutional networks (FCN) have achieved the state-of-the-art performance on image segmentation tasks [9]. However, 2D FCN may be not optimal for 3D object localization and segmentation tasks from volumetric data, which are common in the field of medical image computing, since limited spatial information is considered.

Inspired by the success of 2D FCN on natural image segmentation tasks and aim to tackle the challenges of 3D object localization and segmentation problems, we propose a novel 3D FCN model for localization and segmentation tasks from high-dimensional volumetric data. We comprehensively studied and compared the 2D and 3D deep learning with end-to-end learning and inference on a challenging medical application. Extensive experiments on the task of IVD localization and segmentation demonstrated that exploring flexible 3D information can achieve more promising performance. In addition, the 3D FCN is overall general and can be readily adapted to other volumetric localization and segmentation tasks.

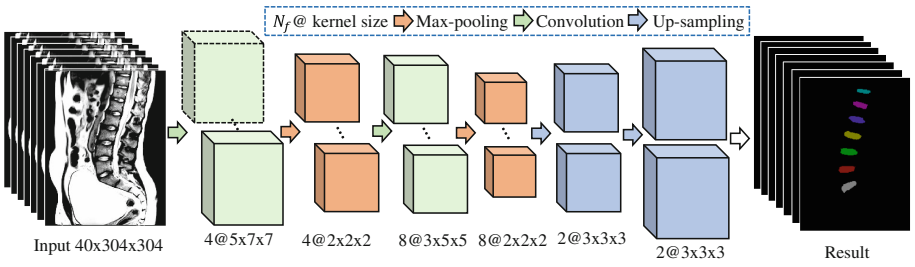

Fig. 1. The illustration of 3D fully convolutional networks.

2 Method

In this section, we present the design and implementation of the proposed end-to-end 3D FCN and explain its advantages over 2D versions. Figure 1 illustrates the overview of the 3D FCN.

2.1 Architecture Design

With the development of imaging technology, volumetric data have been more and more popular in clinical practice, such as the 3D MR images employed in diagnosis and treatment of spinal pathologies. Therefore, researchers attempted to deal with the 3D data by using several adaptations of 2D convolutional neural networks (CNNs) [4,10,11]. However, 2D CNN cannot sufficiently leverage the volumetric information, which is crucial for 3D detection and segmentation tasks. To the best of our knowledge, 3D CNN models have been rarely presented in medical image processing community so far [5,7,12]. Although these models were not trained in an end-to-end way on segmentation tasks, preliminary studies have demonstrated the effectiveness of 3D CNN on volumetric tasks. In this paper, we, for the first time, present an effective and efficient end-to-end trained 3D FCN framework for volumetric data processing. As shown in Fig. 1, our model includes three kinds of layers.

3D Convolutional Layer. In the 3D convolutional layer, a 3D feature volume is produced by convolving the input with 3D kernels, adding a bias term and finally applying a non-linear activation function. Thus, the output is 3D feature volumes (N_f denotes the number of feature volumes) instead of 2D feature maps for 2D CNNs. The 3D feature volumes can be presented as:

$$h_i^l = \sigma\left(b_i^l + \sum_k h_k^{l-1} * W_{ki}^l\right) \tag{1}$$

where h_i^l represents the ith 3D feature volume in the lth layer, W_{ki}^l denotes the 3D convolution kernel connecting the successive feature volumes and b_i^l is the bias term. The $\sigma(\cdot)$ is a non-linear rectifier function [8]. The above operation can be expanded in all three dimensions as:

$$(h_k^{l-1} * W_{ki}^l)[x,y,z] = \sum_m \sum_q \sum_r W_{ki}^l[m,q,r]\ h_k^{l-1}[x-m, y-q, z-r] \tag{2}$$

where $W_{ki}^l[m,q,r]$ and $h_k^{l-1}[x-m, y-q, z-r]$ represent the element-wise values within the 3D convolution kernel and 3D feature volume, respectively. Thus, the 3D kernel is shared within the same feature volume, and the spatial information can be effectively exploited.

3D Max-Pooling Layer. In-between 3D convolutional layers, 3D max-pooling layers are periodically inserted for endowing local invariance. Specifically, max-pooling partitions the input volume into a set of non-overlapping cubes, for each

such sub-volume, outputs the maximum value. In this way, the max-pooling operation is performed in a 3D fashion, where activations within a cubic neighborhood are abstracted and promoted to higher layers.

3D Up-Sampling Layer. Due to the utilization of successive down-sampling layers, the output dimensions are typically reduced compared to the original input size. In this regard, we propose a 3D up-sampling layer to bridge the coarse feature volumes into dense predictions. In the up-sampling layer, the dimensions of down-sampled feature volumes are gradually up-sampled to the original input size. Specifically, up-sampling with a factor of d (d was typically set as 2 in our experiments) is achieved by increasing d times convolutions on the coarse feature volumes. While reshaping the neurons into higher resolution feature volumes, a neighboring-like interpolation of cubic neurons was preserved. Note that the up-sampling kernels are not fixed (thus it doesn't have to be bilinear interpolation), but are learned in an end-to-end way. In this way, the network can take a whole volume as input and output the result within a single forward propagation.

2.2 End-to-End Learning and Voxel-Wise Inference

Previous 3D CNN based models were trained on *sub-volume* samples and employed in a sliding window way to generate the segmentation result [12]. Specifically, fixed-sized 3D training sub-volumes were extracted from the volumetric data and utilized to train a classification model. However, these methods suffer from the inefficient testing inference due to the redundant overlapping computations. Compared with previous studies, our method integrated with up-sampling layers can perform end-to-end learning and voxel-wise inference, which significantly improve the efficiency. The network takes the whole volume as input and generates the 3D segmentation mask (the same size of original input) within single forward propagation. Finally, the training of whole network is formulated as a per-voxel classification problem with respect to the ground-truth segmentation mask. By denoting the parameters in the network by $\theta = \{W, b\}$, the optimization objective is to minimize the following negative log likelihood function via standard back-propagation:

$$\mathcal{L}(\mathcal{X}; \theta) = \frac{\lambda}{2}||W||_2^2 - \sum_{x \in \mathcal{X}} \sum_{c=1}^{C} y_c^x \log p_c(x; W, b) \tag{3}$$

where the first part is the regularization term and latter one is the fidelity term. The tradeoff of these two terms is controlled by the hyperparameter λ. $p_c(x; W, b)$ denotes the predicted probability of cth class (total C classes) after softmax classification layer for voxel x in volume space $\mathcal{X}$, and $y_c^x \in \{0, 1\}$ is the corresponding label. The parameters $\theta = \{W, b\}$ of our 3D FCN are jointly optimized in an end-to-end way by minimizing the loss function $\mathcal{L}$. In our 3D FCN, each voxel in the 3D image is taken as a training sample to the network. Therefore, the equivalent training database is dramatically enlarged and the risk of over-fitting with the limited medical dataset is effectively alleviated. In addition, with no need to crop overlapped sub-volumes, the learning process is quite efficient.

Finally, we utilized simple post-processing steps to generate local smooth segmentation results. First, we binarized the probability maps by a given threshold after filtering with a small disk. Then the segmentation mask can be obtained by finding the connected component after removing small areas. Furthermore, the center of IVD can be determined as the centroid of the connected component. To this end, the centers and segmentation masks of IVDs (type from $S1$ to $T11$) can be localized and segmented sequentially.

3 Experimental Results

3.1 Dataset and Pre-processing

We evaluated our method on the MICCAI 2015 challenge dataset on *Automatic Intervertebral Disc Localization and Segmentation from MR Images*. The dataset consisted of 25 3D T2-weighted turbo spin echo MR images, which were acquired with the 1.5 Tesla MRI scanner of Siemens Magnetom Sonata. The images were resampled into the resolution of $2 \times 1.25 \times 1.25$ mm^3. A total of 15 3D images with ground-truth annotations were released for training, while testing data was divided into two sections (5 images in test1 for offline evaluation and 5 images in test2 for on-site competition) with ground-truths held out by the challenge organizers for independent evaluation. We pre-processed the data by subtracting the mean value before inputting into the network.

3.2 Implementation Details

The kernels of network were randomly initialized from the Gaussian distribution ($\mu = 0, \sigma = 0.01$). The proposed 3D FCN was implemented with Python based on the Theano library and it took about 0.3 s to process one test image with size $40 \times 304 \times 304$, which was much faster than 2D FCN and methods utilizing the sliding window way [12], which caused a large amount of redundant computations on neighboring voxels.

For comparison, we also implemented a 2D FCN for the processing of volumetric data, where the input is the adjacent slices (3 slices in our implementation and the output is the binary mask of the middle slice.) The 2D FCN was implemented with Matlab and C++ based on the study of [9]. Generally, it took less than 1 min to process one test image with the same size using a standard PC with a 2.50 GHz Intel(R) Xeon(R) E5-1620 CPU and a NVIDIA GeForce GTX X GPU.

3.3 Qualitative Evaluation

Two examples of qualitative localization and segmentation results from different methods can be seen in Fig. 2. We can see that methods including both 2D FCN and 3D FCN can generate visually smooth and accurate segmentation results. As the green crosses shown in the figure, they can successfully localize the centers

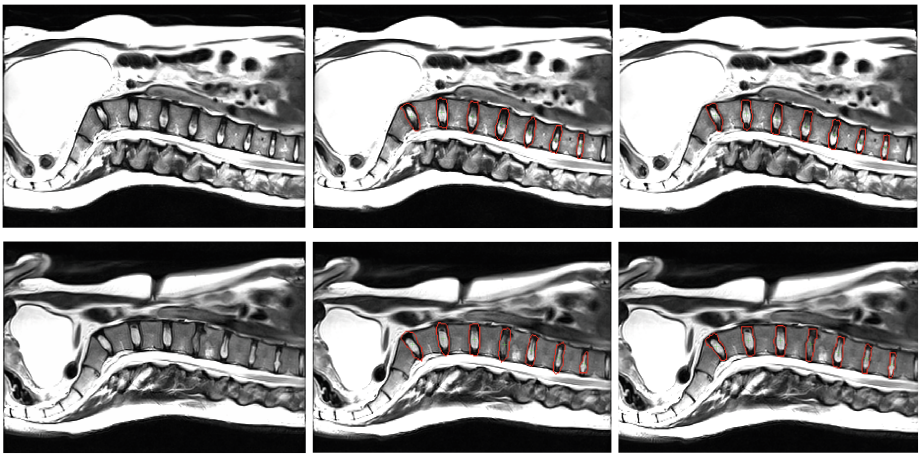

Fig. 2. Examples of localization (the centers of IVDs are marked by green crosses) and segmentation results (the boundaries of segmentation masks are delineated by red lines) of different methods: original images, 2D FCN and 3D FCN (from left to right). (Color figure online)

of IVDs. When comparing the results of 2D FCN and the proposed 3D FCN, it is observed that the 3D FCN can achieve more accurate and smooth results, which is attributed to the advantage of proposed 3D FCN by exploiting large volumetric contextual information.

3.4 Quantitative Evaluation and Comparison

Evaluation Metrics. The evaluation metrics on IVD localization include mean localization distance (MLD) with standard deviation (SD) and successful detection rate P. If the absolute difference between the localized IVD center and the ground truth center is no greater than 2 mm, the localization of this IVD is considered as a correct detection; otherwise, it is considered as a false detection. The evaluation metrics on IVD segmentation include mean dice overlap coefficients (MD) with SD and mean average absolute distance (MAAD) with SD. Larger MD means better segmentation accuracy. MAAD is a metric measuring the average absolute distance between the ground truth disc surface and the segmented surface, hence smaller MAAD means better segmentation accuracy.

Table 1. Results of IVD localization and segmentation on test1 dataset

Method	MLD ± SD (mm)	P (2.0 mm)	MD ± SD	MAAD± SD (mm)
2D FCN	1.07 ± 0.62	91.4 %	83.2 % ± 4.6 %	1.58 ± 0.28
3D FCN	**0.91 ± 0.58**	**94.3 %**	**88.4 % ± 5.3 %**	**1.27 ± 0.26**

Table 2. Results of IVD localization and segmentation on test2 dataset

Method	MLD ± SD (mm)	P (2.0 mm)	MD ± SD	MAAD± SD (mm)
2D FCN	0.89 ± 0.48	**94.3%**	82.2 % ± 6.8 %	1.77 ± 0.29
3D FCN	**0.85 ± 0.52**	**94.3 %**	**89.0 % ± 3.4 %**	**1.22 ± 0.15**

Results of IVD Localization. The quantitative localization results of different methods on test1 and test2 datasets can be seen in Tables 1 and 2, respectively. It is observed that both 3D FCN and 2D FCN can localize the centers of IVD with more than 90 % detection rate, while 3D FCN achieved a higher detection rate (94.3 %) than that of 2D FCN (91.4 %) within range of 2 mm on test1 dataset. In addition, the 3D FCN achieved a smaller MLD with a smaller SD than 2D FCN. The comparison between 2D FCN and 3D FCN demonstrates the efficacy of taking full advantage of 3D volumetric information consistently. The results of our method achieved the best localization results on the onsite competition, outperforming all the other methods.

Results of IVD Segmentation. From the segmentation results of different methods on test1 and test2 datasets, we can see that the 3D FCN achieved much better performance than 2D FCN on different segmentation measurements, highlighting the utility of volumetric information on 3D object segmentation problems. Although without any sophisticated post-processing steps or incorporating explicit shape regression methods (e.g., active shape model), our methods with 3D FCN achieved competitive performance during the challenge on the segmentation task of IVD. To sum up, in comparison of 2D and 3D FCN, we corroborated the significance of volumetric feature representation in 3D object localization and segmentation tasks.

4 Conclusions

In this paper, we propose a novel 3D FCN model with end-to-end learning and inference (i.e., voxel-wise predictions) for intervertebral disc localization and segmentation. We compare the performance of 2D and 3D FCN to validate the efficacy of exploiting volumetric contextual information. Extensive experiments on the 3D T2 MRI data of MICCAI 2015 challenge dataset corroborated that our method achieved the best results on the localization task and competitive performance on the segmentation task. In addition, our approach is general and can be easily extended to other 3D localization and segmentation applications. Future work will include incorporating shape regression methods to further improve the performance and testing our method on a larger dataset with pathological cases included.

Acknowledgements. The work described in this paper was supported by Research Grants Council of the Hong Kong Special Administrative Region (Nos. CUHK 412513 and CUHK 14202514).

References

1. 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., et al.: Introduction: disc degeneration: summary. Spine **29**(23), 2677–2678 (2004)
2. 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)
3. 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, Heidelberg (2015)
4. Chen, H., Yu, L., Dou, Q., Shi, L., Mok, V.C., Heng, P.A.: Automatic detection of cerebral microbleeds via deep learning based 3D feature representation. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 764–767. IEEE (2015)
5. Dou, Q., Chen, H., Yu, L., Zhao, L., Qin, J., Wang, D., Mok, V.C., Shi, L., Heng, P.A.: Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. IEEE Trans. Med. Imaging **35**(5), 1182–1195 (2016)
6. Glocker, B., Zikic, D., Konukoglu, E., Haynor, D.R., Criminisi, A.: Vertebrae localization in pathological Spine CT via dense classification from sparse annotations. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 262–270. Springer, Heidelberg (2013)
7. Kamnitsas, K., Chen, L., Ledig, C., Rueckert, D., Glocker, B.: Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI. In: Ischemic Stroke Lesion Segmentation, p. 13 (2015)
8. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)
9. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
10. Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 246–253. Springer, Heidelberg (2013)
11. Roth, H.R., Lu, L., Liu, J., Yao, J., Seff, A., Kevin, C., Kim, L., Summers, R.M.: Improving computer-aided detection using convolutional neural networks and random view aggregation. (2015). arXiv preprint arXiv:1505.03046
12. Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J.: Multi-modal brain tumor segmentation using deep convolutional neural networks. In: Proceedings in MICCAI BraTS (Brain Tumor Segmentation) Challenge, pp. 31–35 (2014)
13. Wang, Z., Zhen, X., Tay, K., Osman, S., Romano, W., Li, S.: Regression segmentation for M3 spinal images. IEEE Trans. Med. Imaging **34**(8), 1640–1648 (2015)
14. Zhan, Y., Maneesh, D., Harder, M., Zhou, X.S.: Robust MR spine detection using hierarchical learning and local articulated model. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 141–148. Springer, Heidelberg (2012)