



HHS Public Access

Author manuscript

Med Image Comput Assist Interv. Author manuscript; available in PMC 2018 July 20.

Published in final edited form as:

Med Image Comput Assist Interv. 2017 September ; 10435: 692–700. doi:
10.1007/978-3-319-66179-7_79.

Does Manual Delineation only Provide the Side Information in CT Prostate Segmentation?

Yinghuan Shi¹, Wanqi Yang^{1,2}, Yang Gao¹, and Dinggang Shen³

¹State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

²School of Computer Science, Nanjing Normal University, Nanjing, China

³Department of Radiology and BRIC, UNC Chapel Hill, Chapel Hill, USA

Abstract

Prostate segmentation, for accurate prostate localization in CT images, is regarded as a crucial yet challenging task. Nevertheless, due to the inevitable factors (*e.g.*, low contrast, large appearance and shape changes), the most important problem is how to learn the informative feature representation to distinguish the prostate from non-prostate regions. We address this challenging feature learning by leveraging the manual delineation as guidance: the manual delineation does not only indicate the category of patches, but also helps enhance the appearance of prostate. This is realized by the proposed cascaded deep domain adaptation (CDDA) model. Specifically, CDDA constructs several consecutive source domains by employing a mask of manual delineation to overlay on the original CT images with different mask ratios. Upon these source domains, convnet will guide better transferrable feature learning until to the target domain. Particularly, we implement two typical methods: *patch-to-scalar* (CDDA-CNN) and *patch-to-patch* (CDDA-FCN). Also, we theoretically analyze the generalization error bound of CDDA. Experimental results show the promising results of our method.

1 Introduction

Prostate cancer is believed as one of the most leading diseases for cancer-related deaths for male population in the world. According to the report from the National Cancer Institute, about 1 out of 5 will be diagnosed with prostate cancer in the US¹. Recently, the CT image guided radio therapy (IGRT) is commonly adopted, which has shown its ability in better guiding the delivery of radiation to prostate cancer tissues. From clinical consideration, the prostate cancer tissues should be precisely killed by high energy X-rays, and meanwhile the other normal tissues should not be excessively hurt by the delivered X-rays. Therefore, precisely segmenting the prostate in CT images for localization is clinically significant. Traditional segmentation is manually done by the physician with slice-by-slice delineation. However, this process is very time-consuming and laborious, *i.e.*, up to 20min to segment a CT image.

Correspondence to: Yinghuan Shi.

¹<http://seer.cancer.gov/statfacts/html/prost.html>.

In recent years, several efforts have been contributed to segment prostate in treatment images [12]. In this setting, for each patient, first several images (*i.e.*, a planning image and first 2–4 treatment images) are collected and delineated as ground-truth training images. Then, prostate likelihood estimation is performed as a classification/regression task to generate a likelihood map. With the obtained likelihood map, the refinement methods by borrowing the patient-specific shape prior [12] are finally employed to segment the prostate.

However, compared with segmentation in the treatment images (only performing the segmentation on a same patient), segmenting prostate in the planning images for different patients, which is normally without any shape prior, is usually believed as a much more challenging task [2]. Therefore, we focus on segmentation in the planning images in this paper. Recently, several contributions have been advocated for segmenting prostate in the planning images. Lay et al. [4] combined the global and local context information to segment the prostate. Lu et al. [8] adopted the deformable model by using Jensen-Shannon divergence guided boundary detection. [2] proposed to segment the pelvic organs by regression-based deformable model and multi-task random forests. Martnez et al. [9] developed a geometrical shape model for prostate segmentation. Most of these previous methods utilized the deformable model, with the performance largely influenced by the model initialization.

The major challenges from segmenting the planning images consist of (1) large shape and appearance changes of prostate regions across different patients (even different CT slices from a same patient), and (2) prostate boundary is hard to distinguish due to the low contrast in CT images. Thus, extracting the discriminative features or learning the informative feature representation on patch-level largely influences subsequent segmentation performance. Different from segmenting the treatment images in which the low-level hand-crafted features (*e.g.*, HoG, LBP) could be adopted (the images of a same patient from different day have relatively similar appearance), the performance of these hand-crafted features cannot be well guaranteed in representing prostate/non-prostate regions across different patients, during the segmentation of the planning images.

Recently, convolutional neural net (convnet) has shown its powerful ability on automatical feature representation. Although there are few attempts for CT prostate segmentation, recently in other modalities (*i.e.*, MR), a new trend is to use convnet or deep features for prostate segmentation [3,10]. Unfortunately, according to our observation, for our task of segmenting CT planning images, directly applying the convnet for feature learning is not desirable, since these aforementioned factors across different patients make convnet very difficult to train.

To guide better feature learning for convnet, we propose to employ an *easy-to-hard* transfer sequence (see Fig. 1), and devise a novel model namely cascaded deep domain adaptation (CDDA). In CDDA, manual delineation could not only be used as the side information to indicate the category of corresponding patch, but also be employed as a mask image to enhance the prostate appearance upon the constructed transfer sequence. Typically, beyond original CT images, we generate several consecutive source domains by employing a mask of manual delineation to overlay on the original CT images with different mask ratios (*i.e.*,

0.2, 0.1, 0.05 and 0.01) in Fig. 1. Basically, higher mask ratio indicates the prostate can be easier to distinguish while lower mask ratio indicates more difficult to segment.

Through CDDA, our goal is thus becoming sequentially transferring the convnet from the first source domain to the last source domain, until to the target domain without the mask (*i.e.*, original CT images). For domain transfer, the convnet offers us a feasible way by fine-tuning on the obtained convnet, using the patches sampled from current source domain (or target domain). In particular, two specific methods from *patch-to-scalar* and *patch-to-patch* perspectives, *i.e.*, CDDA-CNN (convolutional neural networks) and CDDA-FCN (fully convolutional networks), are proposed to support our motivation. We also theoretically analyze that, our CDDA could not only help reduce the sample difference in two consecutive domains, but also ensure relative low training error on *easy* task (*i.e.*, the highest mask ratio). These two factors are helpful to mathematically derive a low generalization error.

2 Our Method

The whole framework mainly consists of (1) the image preprocessing, (2) the prostate likelihood estimation, and (3) the result refinement. Please see Fig. 2 for the whole framework of our method.

Image Preprocessing

Formally, by denoting the n -th training images in target domain (*i.e.*, original CT image) as \mathbf{I}_n and the corresponding *binary* manual delineation as \mathbf{G}_n , we can generate the corresponding n -th training image in k -th ($k = 1, 2, 3, 4$) source domain, \mathbf{I}_n^k , as

$\mathbf{I}_n^k = \mathbf{I}_n + \alpha_k \mathbf{G}_n$, where $\alpha_1 = 0.2$, $\alpha_2 = 0.1$, $\alpha_3 = 0.05$, $\alpha_4 = 0.01$. Please note that, these values of mask ratio (*i.e.*, α_k) are empirically selected. Setting too high mask ratio will lead the source and target domains existing large sample difference especially on prostate regions, while setting too low mask ratio will lose the meaning of domain adaptation.

Regarding the fact that different local regions of prostate and non-prostate regions are not uniformly distributed or similarly appeared, we sample the prostate and non-prostate patches from training images for both the source and target domains, to reflect various appearance inside and outside the prostate region. Basically, to prevent the possible class imbalance issue if uniformly sampling training patches from CT slices (*i.e.*, prostate only belongs to a small proportion of CT scans), we normally sample training patches from three aspects: (1) sampling (almost) all patches whose centers are located inside the prostate region, (2) densely sampling the patches which are centered closely to prostate boundary, and (3) sparsely sampling patches which are centered far from prostate boundary. It is noteworthy that the number of non-prostate patches nearly equals to that of prostate patches. Specifically, the number of training patches varies from 1000 – 3000 in each CT slice according to the size of prostate region, and totally around $10^6 \sim 3 \times 10^6$ training patches are used for training convnet in each source domain. We set the patch size as 32×32 to fully capture local appearance context.

CDDA-CNN & CDDA-FCN

For training CDDA-CNN, considering that our task is dealing with small-size patch, we employ the *LeNet* [5] as the base CNN learner. For the label (*i.e.*, scalar) of each patch, we simply use the label of central voxel as that of the whole patch, which follows a *patch-to-scalar* manner. Typically for training CDDA-CNN, the patches sampled from the images in the first source domain (with highest mask) as well as the corresponding label (*i.e.*, 0 indicates the non-prostate, while 1 indicates the prostate) will be first used to train *LeNet*. Then, the learned convnet will be fine-tuned by the patches from the images in the second source domain, and the process continues until we use the original patches in the target domain to fine-tune the convnet. Finally, the learned convnet will be directly used to estimate the prostate likelihood for testing patches from the images to be segmented.

For training CDDA-FCN, we modified the traditional FCN [7] designed to deal with relative large-size natural image for annotation and segmentation task, in order to fit the requirement of the small-size patch problem. The whole training process is similar with that of CDDA-CNN. The only difference is that as FCN [7], instead of using discrete values as the labels, we directly use the response patches (*i.e.*, manual delineation patches) as the labels, which follows a *patch-to-patch* manner. The max-pooling in *Layer 3* is used for image de-noising, and an up-sampling operator is thus required in CDDA-FCN. The network setups of CDDA-CNN and CDDA-FCN can be found in Tables 1 and 2, respectively.

Also, please note that there are two major fine-tuning paradigms for domain adaptation in convnet. The first paradigm is that parameters in all layer should be tuned [15], while the second paradigm is that parameters in the first several convolutional layers should be fixed [14]. In this paper, we adopt the first paradigm and analyze this issue in our experiment. We implement CDDA-CNN and CDDA-FCN with MatConvNet toolbox [13], on a PC with 3.7 GHz CPU, 32GB RAM, and GeForce GTX 745 GPU. The number of epoch is set to 50 for each domain both for CDDA-CNN and CDDA-FCN. The running time for training convnet of each domain requires 5 to 8 h. Currently, our method takes 5–10 min for segmenting one image.

Refinement

The learned prostate likelihood map by performing CDDA-CNN or CDDA-FCN cannot be directly used as final segmentation. Since the shape prior is not available, we adopt the level-set model [6] for final segmentation, which does not require any shape prior. In particular, the center of the prostate likelihood map will be set as the initialized point when performing level-set model during refinement.

Theoretical Explanation

To perceptively analyze our motivation, we first introduce the theoretical study in [1] to reveal the generalization error bound of domain adaptation. Formally, Shai Ben-David *et al.* [1] analyzed that generalization error bound is up-bounded by the training error in source domain and the difference of sample distribution between source and target domains. Inspired by [1], we can obtain the following *Theorem 1* for our proposed CDDA.

Theorem 1—Let \mathcal{R} be a fixed representation function from \mathcal{X} to \mathcal{Z} , and \mathcal{H} be a hypothesis space. If applying \mathcal{R} to $D_S^i - i.i.d.$ samples and denoting $d_{\mathcal{H}}$ as the distance according to \mathcal{H} , then with probability at least $1 - \delta$, for every $\theta \in \mathcal{H}$:

$$\varepsilon_T(\theta) \leq \varepsilon_S^1(\theta) + \left(\sum_{i=1}^{l-1} d_{\mathcal{H}}(D_S^i, D_S^{i+1}) + d_{\mathcal{H}}(D_S^l, D_T) \right) + \hat{\lambda} + \hat{\mathcal{C}} \quad (1)$$

where $\varepsilon_S^i(\theta)$ is the expected risk of the i -th ($i = 1, \dots, l$) source domain D_S^i . $\varepsilon_T(\theta)$ is the expected risk of target domain D_T . $\hat{\lambda}$ (i.e., the error calculated by the optimal hypothesis) and $\hat{\mathcal{C}}$ (i.e., a constant for the complexity of hypothesis space and the risk of an ideal hypothesis for both domains) can be regarded as the constants.

Theorem 1 can be mathematically verified by [1] using *Induction*. Please note that, *Theorem 1* reveals in CDDA, a low generalization error bound can be achieved if (1) training error in the first source domain should be small, and (2) two consecutive domains should have smaller sample differences. Normally, both these two major factors could be satisfied in our CDDA, since (1) in the first source domain, the prostate can be easily distinguished, due to the high mask value is used, and (2) the cascaded transfer could ensure that the patches from two consecutive domains have relatively small sample difference. Thus, the low generalization error can be theoretically expected.

3 Experimental Results

We evaluated our method on a CT image set including 22 different patients, in which each patient has only 1 planning image. The resolution of each CT image after preprocessing is $193 \times 152 \times 60$, with in-plane voxel size as $1 \times 1 \text{mm}^2$ and the inter-slice thickness as 1 mm. Each image is manually delineated by physician as the ground-truth for evaluation. Normally, we perform 2-fold cross validation for segmenting 22 planning images. For evaluation metric, we employ the Dice ratio, average surface distance (ASD), and centroid distance (CD) along 3 directions (i.e., lateral x-axis, anterior-posterior y-axis, and superior-inferior z-axis), which are widely used in previous studies.

Hand-Crafted Features or Deep Features

We compare the performance of hand-crafted method (short for HCM) and our CDDA-CNN. For implementation of HCM, three types of low-level features, i.e., HoG, LBP, and Haar wavelet, are employed. Specifically, we extract 9-dimensional HoG features by calculating within 3×3 cell blocks, 30-dimensional LBP features by calculating with neighboring voxel-number as 8 and radius value as 2, 14-dimensional Haar wavelet features by convolving the 14 multi-resolution wavelet basis functions. The window size for extracting these hand-crafted features is set to 21×21 as [12] suggested. Thus, each patch will be represented by a 53-dimensional hand-crafted feature. After feature extraction, a RBF kernel-based SVR is trained on all the training patches, and the prostate likelihood map for testing images will then be predicted. Also, for the refinement, a same level-set model [6] is adopted on respective prostate-likelihood maps for HCM. Table 3 reports the

comparison of Dice, ASD, CD, showing that CDDA-CNN is significantly better than HCM. Moreover, we provide several typical likelihood maps to visually show the comparison in Fig. 3. HCM performs poor when large appearance variation happens among training and testing images, due to its fixed feature representation way (several top/bottom part of prostate is totally mis-classified).

Comparison of Different Transfer Paradigms

To fully investigate the *easy-to-hard* transfer sequence, we compare our method with four different baselines: (1) WT-CNN (*without transfer*): Training the convnet with only the images from target domain, and then performing subsequent prediction; (2) ST-CNN (*selective transfer*): The first two convolutional layers will not involve in the fine-tuning with the fixed parameters [14]; (3) DTH-CNN (*direct transfer*): Directly transferring the first source domain to target domain; (4) DTL-CNN (*direct transfer*): Directly transferring the last source domain to target domain.

Table 3 lists the results among these competitive transfer paradigms, respectively. We can observe that CDDA-CNN could obtain the best performance among these baseline methods. Please note that the convnet setup in WT-CNN, ST-CNN, DTH-CNN, DTL-CNN, and our CDDA-CNN is consistent for fair comparison. The results show that by sequentially transferring the convnet with continuously reducing the mask ratio in *easy-to-hard* transfer sequence, the convnet is expected to achieve good generalization ability across different patients, which might exist large sample difference. Moreover, we illustrate the typical examples for visual comparison in Fig. 3. Red curves denote the manual delineations. e_i indicates the i -th example.

Furthermore, for ST-CNN, different from natural images which may share common low-level representation (*e.g.*, different cars share same tire), the selective transfer strategy could not ensure promising results in our small-size patches. For DTH-CNN, the two domains exist large difference in prostate appearance, making the direct transfer hard to work, while for DTL-CNN, both domains are difficult to segment, thus leading very limited improvements.

CDDA-CNN or CDDA-FCN

Moreover, we compare our CDDA-CNN and CDDA-FCN in Table 3. It can be observed that these two methods are overall comparable, however, CDDA-FCN is with more stable results (*i.e.*, small standard deviation) compared to CDDA-CNN. The major reason is that the response patch could provide more smoothness information than the discrete values (*i.e.*, 0 or 1) used in CDDA-CNN. Visual comparison can be referred to Fig. 4.

Finally, although the adopted CT image sets are different, we still illustrate our results with related works developed for segmenting the planning images recently in Table 4. Please note that all the reported results are only for reference to see current level of state-of-the-art methods.

Conclusion

We propose to construct an *easy-to-hard* transfer sequence for CT prostate segmentation, in which the manual delineation could not only provide side information but also be used to enhance the appearance of prostate to guide better feature learning. Upon this constructed sequence, the cascaded deep domain adaptation (CDDA) is originally proposed, by sequentially transferring the convnet across different domains. The results demonstrate the promising results of our method.

Acknowledgments

This work was supported by NSFC (61673203, 61432008, 61603193), NIH Grant (CA206100), and Young Elite Scientists Sponsorship Program by CAST (YESS 20160035).

References

- 1 Ben-David S, Blitzer J, Crammer K, Pereira F. Analysis of representations for domain adaptation. NIPS. 2006:137–144.
- 2 Gao Y, Shao Y, Lian J, Wang AZ, Chen RC, Shen D. Accurate segmentation of CT male pelvic organs via regression-based deformable models and multi-task random forests. IEEE TMI. 2016; 35(6):1532–1543.
- 3 Guo Y, Gao Y, Shen D. Deformable MR prostate segmentation via deep feature learning and sparse patch matching. IEEE TMI. 2016; 35:1077–1089.
- 4 Lay N, , Birkbeck N, , Zhang J, , Zhou SK. Rapid multi-organ segmentation using context integration and discriminative models. In: Gee JC, Joshi S, Pohl KM, Wells WM, , Zöllei L, editors IPMI 2013. LNCS Vol. 7917. Springer; Heidelberg: 2013 450462
- 5 LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998; 86(11):2278–2324.
- 6 Li C, Xu C, Gui C, Fox MD. Distance regularized level set evolution and its application to image segmentation. IEEE TIP. 2010; 19(12):3243–3254.
- 7 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. CVPR. 2015:3431–3440.
- 8 Lu C, , et al. Precise segmentation of multiple organs in CT volumes using learning-based approach and information theory. In: Ayache N, Delingette H, Golland P, , Mori K, editors MICCAI 2012. LNCS Vol. 7511. Springer; Heidelberg: 2012 462469
- 9 Martínez F, Romero E, Dréan G, Simon A, Haigron P, De Crevoisier R, Acosta O. Segmentation of pelvic structures for planning CT using a geometrical shape model tuned by a multi-scale edge detector. Phys Med Biol. 2014; 59(6):1471. [PubMed: 24594798]
- 10 Milletari F, Navab N, Ahmadi S. V-Net: fully convolutional neural networks for volumetric medical image segmentation. 2016 arXiv:1606.04797.
- 11 Shao Y, Gao Y, Wang Q, Yang X, Shen D. Locally-constrained boundary regression for segmentation of prostate and rectum in the planning CT images. Med Image Anal. 2015; 26(1): 345–356. [PubMed: 26439938]
- 12 Shi Y, Gao Y, Liao S, Zhang D, Gao Y, Shen D. Semi-automatic segmentation of prostate in CT images via coupled feature representation and spatial-constrained transductive Lasso. IEEE TPAMI. 2015; 37(11):2286–2303.
- 13 Vedaldi A, , Lenc K. Matconvnet - convolutional neural networks for matlab. Proceeding of the ACM International Conference on Multimedia; 2015
- 14 Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? NIPS. 2014:3320–3328.
- 15 Zhang W, , Li R, , Zeng T, , Sun Q, , Kumar S, , Ye J, , Ji S. KDD ACM; 2015 Deep model based transfer and multi-task learning for biological image analysis; 14751484

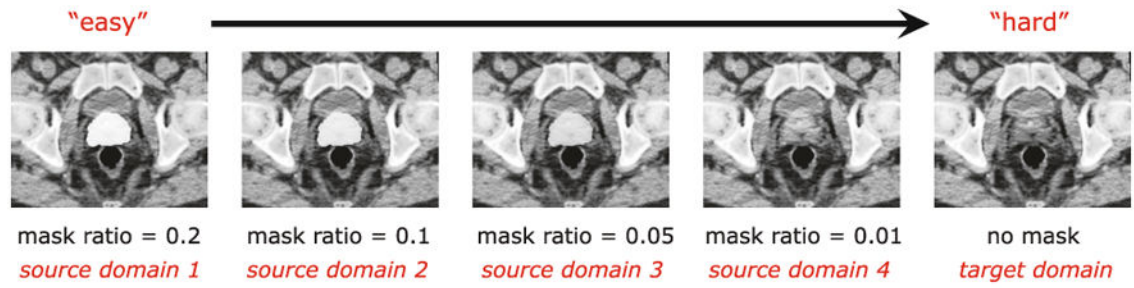


Fig. 1.
Easy-to-hard transfer sequence.

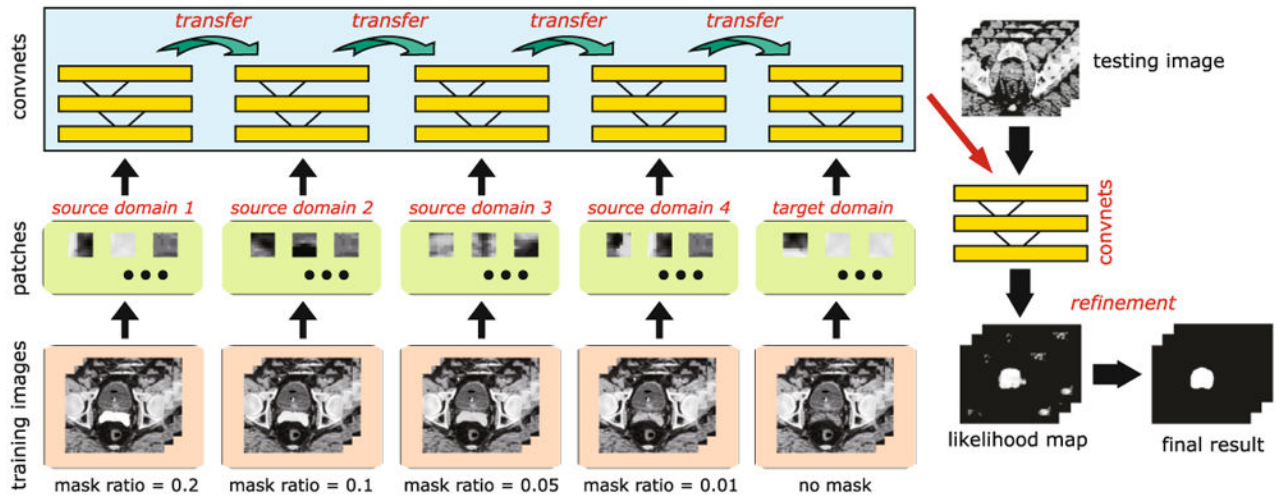


Fig. 2.
The whole framework of our method.

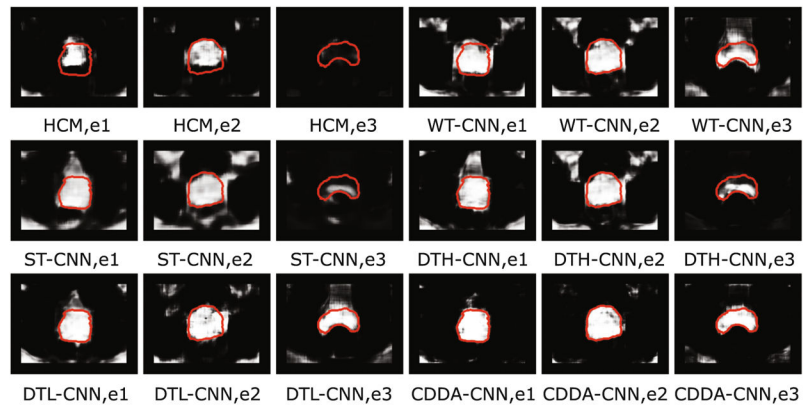


Fig. 3. Typical likelihood maps of HCM, WT-CNN, ST-CNN, DTH-CNN, DTL-CNN, and CDDA-CNN. e_i indicates the i -th example.

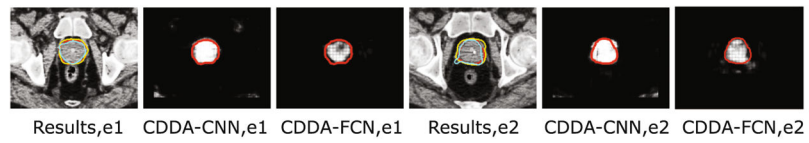


Fig. 4.

Typical likelihood maps and results of CDDA-CNN and CDDA-FCN. Red, yellow, light blue curves denote the manual delineation, CDDA-CNN and CDDA-FCN, respectively.

Table 1

Network setup of CDDA-FCN.

Layer	Kernel	Pooling	Activation
Conv	$5 \times 5 \times 1 \times 32$	Max	Relu
Conv	$5 \times 5 \times 32 \times 32$	Max	Relu
Conv	$5 \times 5 \times 32 \times 64$	Max	Relu
Conv	$4 \times 4 \times 64 \times 64$	n.a	Relu
Conv	$4 \times 4 \times 64 \times 2$	n.a	Softmax
Anonymous	1270 (39.5%)		
Unclassifiable	103 (3.2%)		

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Network setup of CDDA-FCN.

Layer	Kernel	Pooling	Activation
Conv	$5 \times 5 \times 1 \times 32$	n.a	Relu
Conv	$5 \times 5 \times 32 \times 32$	n.a	Relu
Conv	$5 \times 5 \times 32 \times 64$	Max	Relu
Conv	$3 \times 3 \times 64 \times 64$	n.a	Relu
Conv	$3 \times 3 \times 64 \times 2$	Upsample	Softmax

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Comparison of different methods. Best results are in bold.

	HCM	WT-CNN	ST-CNN	DTH-CNN	DTL-CNN	CDDA-CNN	CDDA-FCN
Dice	0.745 ± 0.008	0.801 ± 0.009	0.729 ± 0.106	0.829 ± 0.010	0.834 ± 0.009	0.886 ± 0.006	0.888 ± 0.005
ASD (mm)	5.65 ± 2.33	5.13 ± 2.43	8.39 ± 7.70	3.72 ± 2.08	3.74 ± 1.95	2.15 ± 1.17	1.64 ± 0.96
CD-x (mm)	3.73 ± 1.95	3.27 ± 1.86	4.30 ± 2.11	1.96 ± 1.37	1.78 ± 1.23	1.55 ± 1.07	1.03 ± 0.87
CD-y (mm)	2.33 ± 1.88	2.39 ± 1.77	3.25 ± 1.92	1.97 ± 1.28	2.07 ± 0.99	1.32 ± 1.14	0.97 ± 0.93
CD-z (mm)	1.95 ± 1.06	2.00 ± 1.01	3.31 ± 1.87	1.54 ± 0.66	1.63 ± 0.52	0.67 ± 0.31	0.65 ± 0.30

Table 4

Comparison with previous methods.

	[9]	[11]	[2]	CDDA-CNN	CDDA-FCN
Dice	0.87	0.88	0.87	0.89	0.89
ASD (mm)	N.A	1.86	1.77	2.15	1.64