International Journal of Computer Assisted Radiology and Surgery (2018) 13:967–975
https://doi.org/10.1007/s11548-018-1733-7

ORIGINAL ARTICLE

Automatic bladder segmentation from CT images using deep CNN and 3D fully connected CRF-RNN

Xuanang Xu[1] · Fugen Zhou[1,2] · Bo Liu[1,2]

Received: 5 December 2017 / Accepted: 8 March 2018 / Published online: 19 March 2018

Abstract

Purpose Automatic approach for bladder segmentation from computed tomography (CT) images is highly desirable in clinical practice. It is a challenging task since the bladder usually suffers large variations of appearance and low soft-tissue contrast in CT images. In this study, we present a deep learning-based approach which involves a convolutional neural network (CNN) and a 3D fully connected conditional random fields recurrent neural network (CRF-RNN) to perform accurate bladder segmentation. We also propose a novel preprocessing method, called dual-channel preprocessing, to further advance the segmentation performance of our approach.

Methods The presented approach works as following: first, we apply our proposed preprocessing method on the input CT image and obtain a dual-channel image which consists of the CT image and an enhanced bladder density map. Second, we exploit a CNN to predict a coarse voxel-wise bladder score map on this dual-channel image. Finally, a 3D fully connected CRF-RNN refines the coarse bladder score map and produce final fine-localized segmentation result.

Results We compare our approach to the state-of-the-art V-net on a clinical dataset. Results show that our approach achieves superior segmentation accuracy, outperforming the V-net by a significant margin. The Dice Similarity Coefficient of our approach (92.24%) is 8.12% higher than that of the V-net. Moreover, the bladder probability maps performed by our approach present sharper boundaries and more accurate localizations compared with that of the V-net.

Conclusion Our approach achieves higher segmentation accuracy than the state-of-the-art method on clinical data. Both the dual-channel processing and the 3D fully connected CRF-RNN contribute to this improvement. The united deep network composed of the CNN and 3D CRF-RNN also outperforms a system where the CRF model acts as a post-processing method disconnected from the CNN.

Keywords Segmentation · Bladder · CNN · CRF-RNN

Introduction

Bladder segmentation from CT images is an essential preorder procedure of pelvic radiotherapy treatment planning [1–3]. Accurate and reliable delineation of bladder is required to reduce the risk of irradiation of normal tissues. In clinical practice, this task is usually carried out by radiation oncologists [1,2]. Because manual segmentation of bladder from CT images is laborious and prone to inter- and intra-observer variation, automatic bladder segmentation methods are highly desirable. This is a challenging task because the bladder usually suffers large variations of size and shape across different patients. The neighboring low-contrast soft-tissues also cloud the definition of bladder boundaries in CT images. Traditional gray level-based methods have shown limited success in dealing with this problem.

In recent years, due to the great development of deep CNN, deep learning-based methods have become a popular methodology for medical image segmentation. Many of these methods achieve state-of-the-art performance compared to the traditional methods which relies on the handcrafted features. However, the CNN's very invariance property usually makes it lose the sight of details and produce coarse pixel-/voxel-wise label predictions when applied to

✉ Bo Liu
bo.liu@buaa.edu.cn

[1] Image Processing Center, School of Astronautics, Beihang University, Xueyuan Road No.37, Beijing, China

[2] Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100083, China

organ segmentation tasks, especially for the organ with variable size and shape, such as liver, stomach and bladder. This is mainly caused by the large receptive fields of the last few convolutional layers in the deep networks. A lack of smoothness constraints also results in poor organ delineation and isolated regions in the final segmentation results [4]. These problems restrict the performance of deep learning-based methods for accurate organ segmentation.

In this study, we propose an automatic deep learning-based approach which combines CNN with 3D CRF-RNN for accurate bladder segmentation from CT images. Compared to other deep learning-based methods for organ segmentation, our approach has two compelling advantages. First of all, we exploit a 3D fully connected CRF-RNN to refine the coarse voxel-wise label score maps output from CNN and produce fine-localized bladder probability maps with sharp boundaries. This 3D CRF-RNN is extended from a previous work of Zheng et al. [4] where the CRF-RNN is used to process 2D RGB natural images. All parameters in the 3D CRF-RNN are retuned to adjust monochrome CT images. The 3D CRF-RNN and CNN compose a united deep network, which can be trained end-to-end and outperforms a system where the CRF model acts as a post-processing method disconnected from the CNN. To the best of our knowledge, it is the first application of 3D fully connected CRF-RNN in medical image processing. Secondly, we present a novel preprocessing method, called dual-channel preprocessing, to further improve segmentation accuracy and convergence ability of our deep network. In this method, we extend the CT images to a dual-channel format. The first image channel stores the original CT images and the second image channel stores an enhanced density map of bladder. This dual-channel design can emphasize structural features of bladder as well as preserve original image information. The deep network learns more valuable bladder features from this dual-channel data than the original single-channel CT images within the same training epochs. We build a dataset on clinical data for training and evaluation since there is no available public dataset for bladder segmentation. A set of experiments on this dataset show that our approach outperforms the state-of-the-art method by a large margin on segmentation accuracy, at a computational cost of a few seconds.

The remaining sections are organized as following: "Related works" section briefly introduces related works. "Method" section presents major components and implement details of our approach. In "Experiments" section, we conduct a set of experiments to evaluate the performance of our proposed approach. Finally, we conclude this study in "Discussions" and "Conclusions" sections.

Related works

CNN for medical volumetric image segmentation

In the past few years, many prominent CNN methods [5–7] are presented to accomplish the segmentation tasks on 2D medical images. However, many medical imaging modalities produce volumetric data, such as CT and Magnetic Resonance Imaging (MRI), and they are more complex and memory-consuming than 2D imaging modalities. A lot of effort is devoted to utilizing CNN to segment objects from medical volumetric images. A straightforward idea for addressing this problem is to cut the volumetric image into slices and apply 2D CNN models on these slices [8–13]. This slice-wise manner enriches the training samples since the training data in terms of slices is much larger than the number of volumetric images. However, these methods cannot effectively make full use of the context information in 3D images to perform accurate object segmentation. Some other studies [14–20] exploit 3D CNN models to cope with the segmentation problems on medical volumetric images and several elegant 3D CNN architectures are proposed [14,15,17]. For example, Çiçek et al. [15] present a 3D version of the U-net [5], which firstly introduces Fully Convolutional Network (FCN) [21] into medical image processing, to accomplish the segmentation tasks on medical volumetric images; Milletari et al. [14] propose an alternative architecture to the 3D U-net, called V-net, utilizing the residual connections [22] to improve the fitting ability of models and achieves state-of-the-art performance in public challenge. In this study, we build our segmentation CNN based on the V-net architecture.

Fully connected CRF

Fully connected CRF models are widely used as a kind of post-processing method for various segmentation tasks [4,16,19,23,24]. Differing from basic CRF model which is computed on neighboring pixels, the fully connected CRF model is defined on all pixel pairs in the image. Benefiting from long-range connections between remote pixels, the fully connected CRF model can produce more accurate segmentation results than the basic CRF model. However, the enormous size of the fully connected CRF model increases the computational complexity dramatically and restricts its application in practice. To deal with this problem, Krähenbühl and Koltun [23] propose a mean-field approximation to the CRF distribution and greatly simplify the computation. Following this work, Chen et al. [24] utilize a fully connected CRF model as a post-processing filter to refine segmentation results output from CNN and achieve state-of-the-art performance on public dataset. Kamnitsas et al. [16] firstly employ a 3D fully connected CRF model to post-process brain lesion segmentations on medical volumetric images. In order to

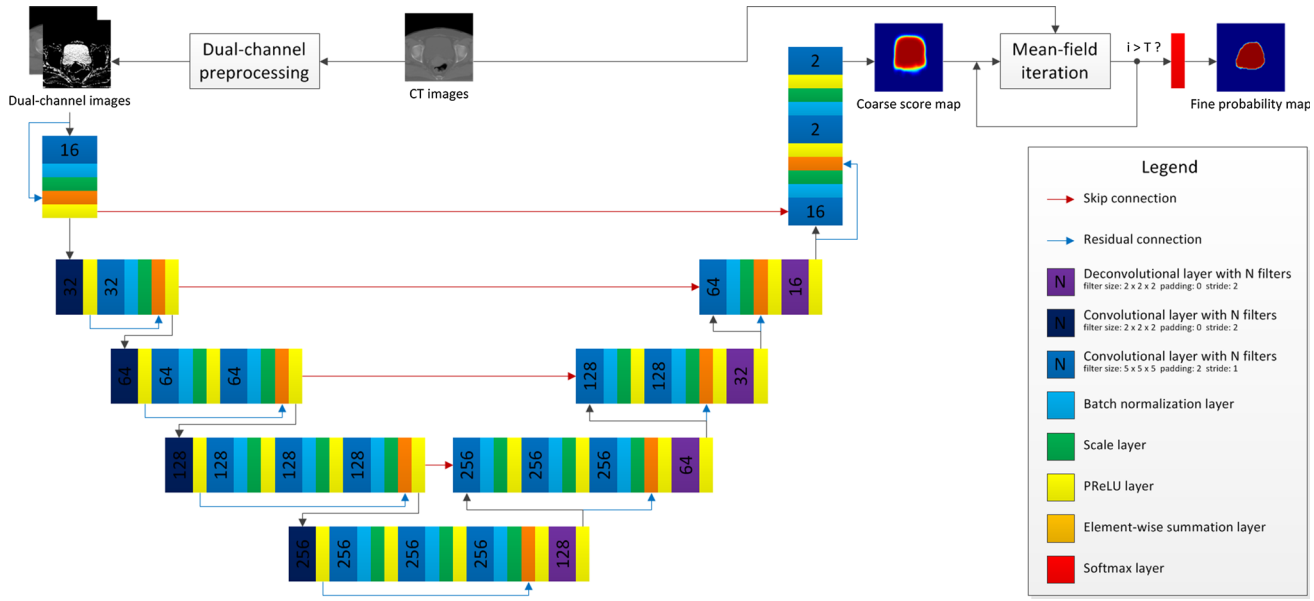

Fig. 1 Schematic representation of our approach. Best viewed in color

obtain a deep network that has desirable properties of both CNN and CRF, Zheng et al. [4] reformulate the mean-field CRF inference as a RNN architecture, called CRF-RNN, and plug in it as a part of the CNN. In this study, we extend this CRF-RNN to a 3D version and exploit it to cooperate with CNN to produce fine-localized segmentation results.

Method

Figure 1 provides a schematic representation of our proposed approach. It works as following: Firstly, we compute an enhanced density map based on CT image and combine this density map with the CT image to compose a dual-channel image. Next, we send this dual-channel image to a CNN to produce a coarse voxel-wise score map of bladder. Finally, a 3D fully connected CRF-RNN refines the score map according to the original CT image and presents a fine-localized bladder probability map.

Dual-channel preprocessing

Since the intensity value of CT voxels reflects the density of materials, we can filter a part of irrelevant organs and soft-tissues whose density differs from bladder obviously according to their voxel intensities. For this purpose, we design a density filter defined as following:

$$D_i = \exp\left[-\left(\frac{I_i - w_l}{w_w}\right)^2\right]$$

where I_i denotes the intensity of CT voxels and D_i represents the corresponding enhanced density value. w_l and w_w are two constants which, respectively, denote the mean and variance of the enhanced organ's intensity. In this study, we use the values of $w_l = 1030$ and $w_w = 20$ to enhance the bladder voxels.

As shown in Fig. 2, the enhanced density maps exclude a number of the irrelevant organs and soft-tissues around the bladder and the bladder-to-background contrast is improved. However, it is also obviously that there is a loss of detail information in the enhanced density maps. In order to preserve the detail information, we keep the original CT image with the enhanced density map to compose a dual-channel image. Note that, after density filtering, the voxel intensities I_i are rescaled to a range of [0, 1].

Segmentation CNN

The CNN used for segmentation in our approach is shown in Fig. 1. We build it based on the V-net architecture proposed in [14]: a contracting path extracts spatially coarse but semantically strong features; an expansive path upsamples these features to hallucinate high resolution feature maps in the same size as input images; skip connections are established between the contracting path and expansive path to improve segmentation details; residual connections are exploited to improve the fitting ability of the network. Differing from the V-net, we introduce batch normalization operation [25] after each convolutional layer to normalize the batch data with its mean and variance. This operation can accelerate the learning convergence of the network during the training stage. All layers in this segmentation CNN are implemented in a 3D manner.

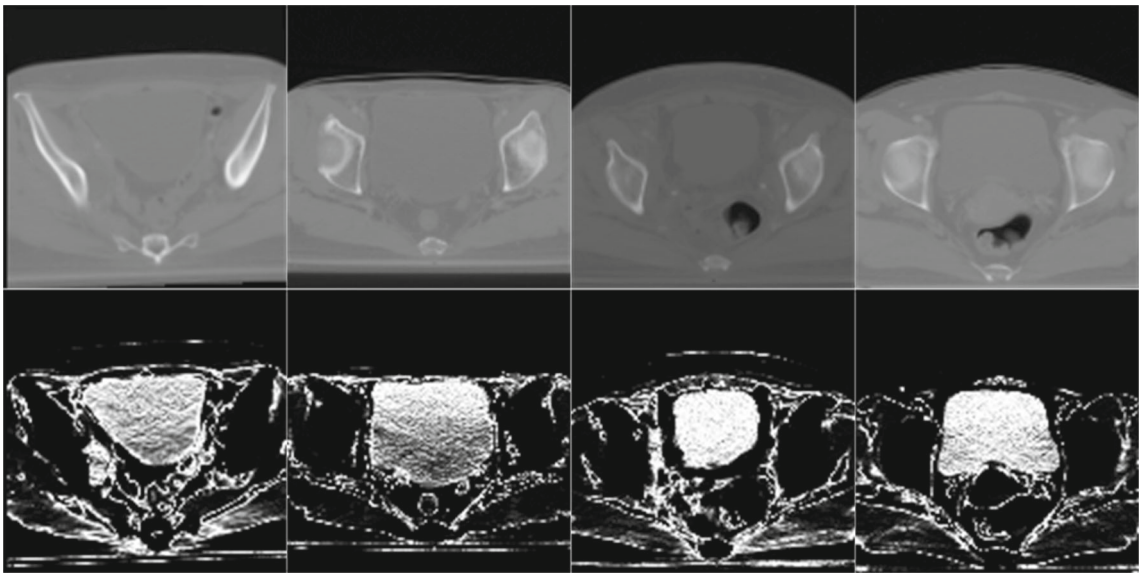

Fig. 2 Original CT images (first row) and enhanced bladder density maps (second row)

3D fully connected CRF-RNN

Due to the large receptive fields of the last few convolutional layers, there are usually non-sharp boundaries and blob-like shapes [4], which affect the final segmentation accuracy, in the bladder score maps output from the segmentation CNN. To deal with this problem, we exploit a 3D fully connected CRF-RNN to refine the outputs of the segmentation CNN as shown in Fig. 1. Next, we provide a brief overview of the CRF model and CRF-RNN.

For a CT image $\boldsymbol{I} = \{I_1, I_2, \ldots, I_N\}$, the voxel-wise label prediction can be modeled as a random variable vector $\boldsymbol{X} = \{X_1, X_2, \ldots, X_N\}$ which is conditioned upon $\boldsymbol{I}$. The pair $(\boldsymbol{X}, \boldsymbol{I})$ compose a CRF which can be characterized by a Gibbs distribution:

$$\mathrm{P}(X = x) = \frac{1}{Z(I)}\exp[-E(x)]$$

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i<j} \psi_p(x_i, x_j)$$

where $E(\boldsymbol{x})$ denotes the energy of a label assignment $\boldsymbol{x}= \{x_1, x_2, \ldots, x_N\}$ $(x_i \in \{l_1, l_2, \ldots, l_L\})$ and $Z(\boldsymbol{I})$ is the partition function. We could get the most probable label assignment $\boldsymbol{x}$ by minimizing the energy $E(\boldsymbol{x})$. For the CRF model, $E(\boldsymbol{x})$ is composed of unary potentials ψ_u and pairwise potentials ψ_p. As introduced in [4,24], the unary potentials can be obtained from the label score map output from CNN and the pairwise potentials can be expressed by two weighted Gaussian kernels:

$$\psi_p(x_i, x_j) = \mu(x_i, x_j)\left[\omega_1 \exp\left(-\frac{p_i - p_j^2}{2\theta_\alpha^2} - \frac{I_i - I_j^2}{2\theta_\beta^2}\right) + \omega_2 \exp\left(-\frac{p_i - p_j^2}{2\theta_\gamma^2}\right)\right]$$

The first appearance kernel indicates the likelihood of two voxels labeled in the same class, according to their positions and intensities with respect to the kernel size $\theta\alpha$ and $\theta\beta$. The second smoothness kernel, which is only related to the distance between voxels with respect to the kernel size $\theta\gamma$, is used to remove small isolated regions. ω_1 and ω_2 controls the weights of these two kernels. Label compatibility function μ, which is initialized by the Potts model, $\mu(x_i, x_j) = [x_i \neq x_j]$, describes the compatibility between different classes.

Since computing the exact distribution $P(\boldsymbol{X})$ above is intractable, Krähenbühl and Koltun [23] propose a mean-field approximation $Q(\boldsymbol{X})$ to the CRF distribution and an iterative algorithm to compute it. Based on this algorithm, Zheng et al. [4] reformulate the mean-field CRF inference as a RNN which can be plugged in as a part of a CNN and trained end-to-end. We extend this work to a 3D fully connected CRF-RNN and exploit it to refine the bladder score maps output from our segmentation CNN. For this 3D CRF-RNN, the initial values of the appearance kernel parameters $\theta\alpha$, $\theta\beta$ and ω_1, are determined by a grid search procedure on validation set. The parameters of the smoothness kernel, which we find do not directly impact segmentation accuracy, are initialized using the values of $\theta\gamma = 1$ and $\omega_2 = 1$ as recommended in [23]. By utilizing the back-propagation algorithm [26], the kernel weights ω_1, ω_2 and the label compatibility function μ can be optimized during the training stage.

Implement details

Our approach is implemented in Caffe framework [27] with Insight Toolkits 4.10 on Windows 7 (64 Bit) platform. All training procedures and experiments are conducted on a workstation equipped with an Intel(R) Core$^{\mathrm{TM}}$ i7-5960X

CPU working at 3.00 GHz and a NVIDIA GTX 1080 Ti graphic card with 11 GB of memory. Because of the limited graphic memory, the CT images processed by our approach are resampled to an identical size of $128 \times 128 \times 64$ voxels with a spatial resolution of $2.0 \times 2.0 \times 3.0\,\text{mm}^3$. To prevent over-fitting, we apply random elastic deformations to each CT image with a probability of fifty percent. All the deformations are performed in CT slice plane. For the 3D fully connected CRF-RNN, the number of mean-field iterations is set to 3 in both training and validation phase. All layers in this deep network are initialized by drawing weights from a zero-mean Gaussian distribution with standard deviation 0.01. We optimize these weights by minimizing the pixel-wise cross-entropy loss of bladder predictions using back-propagation [26] and stochastic gradient descent (SGD) algorithm. For a better reproducibility, the full implementation of our approach has been made publicly available.[1]

The training procedure of our approach consists of two stages. During the first stage, we train the segmentation CNN separately with a fixed learning rate of 10^{-4} for 400 epochs to make sure that the model converges. Restricted by the graphic memory size, each training batch contains two CT images. In the second stage, we initialize the segmentation CNN using the model trained in the first stage and fine-tune it with the 3D fully connected CRF-RNN using a fixed learning rate of 10^{-7} for 10 epochs. We use these extreme values of learning parameters since the 3D CRF-RNN have already been initialized with optimum values as mentioned in "3D fully connected CRF-RNN" section. Due to the extra memory cost of the 3D CRF-RNN, we only feed one CT image to the deep network in each training batch. The weight decay and momentum parameter are set to 0.0005 and 0.99, respectively, in both two training stages. Because there is no clear guide about how to select the hyper-parameters of deep networks for a given application, we determine the parameters above by experimentation, where the best parameters are chosen from a reasonable range based on the segmentation results on the validation set.

Experiments

Dataset

The dataset used in this study consists of 124 CT image volumes from several different CT scanners in one hospital. All CT slices are in two sizes: 256×256 pixels and 512×512 pixels. The number of slices in one volume is ranged from 65 to 156. The in-plane spatial resolution of slice ranges from 0.78 to 2.24 mm and the slice thickness varies from 3.0 to 5.0 mm. 100 image volumes chosen randomly are used for training and the remaining 24 volumes compose the validation set. Each CT image contains a complete anatomic structure of bladder and a segmentation mask corresponding to it. The segmentation masks are delineated by two radiation oncologists manually and verified/modified by a third radiation oncologist who is more experienced. We use these masks as ground truth for training and evaluation.

Quality metrics

The evaluation metrics of segmentation accuracy used in our experiments include the Dice Similarity Coefficient (DSC), the Volume Overlap Error (VOE), the Relative Volume Difference (RVD), the Average symmetric Surface Distance (ASD), and the Maximum Surface Distance (MSD). The DSC is used as major metric of segmentation accuracy. All these metrics are calculated on the validation set. In addition, we evaluate the approach efficiency by counting the average time of bladder segmentation for one CT image volume.

Comparison with state-of-the-art method

In this section, we compare the performance of our approach with that of the state-of-the-art V-net. To make this comparison fair, we implement and train the V-net model along the lines of Milletari et al. [14]

The results of this comparison are summarized in Table 1. It is observed that our approach outperforms the V-net on segmentation accuracy. The DSC of our approach is 8.12% higher than that of the V-net. Although the average segmentation time of our approach (5.4 s) is longer than that of the V-net (0.9 s), it is still acceptable and worth for the considerable improvement of segmentation accuracy. This extra time cost is mainly caused by the 3D fully connected CRF-RNN, which requires time to refine the responses of the preceding segmentation CNN.

Visualization

In Fig. 3, we can visualize the segmentation results of V-net and our approach. It is obviously that the bladder probability maps of our approach present sharper boundaries and more accurate localizations compared with that of the V-net, although the jagged edges, which is caused by the smooth constraints in the 3D fully connected CRF-RNN, lower the visual effect. This improvement makes our approach produce more conformal bladder contours and outperform the V-net on segmentation accuracy. It's remarkable that, in some local regions, the contours performed by our approach are closer to the real bladder boundaries than the manually delineated outlines, although this actually reduces the accuracy metrics. We believe that both the dual-channel preprocessing and the 3D fully connected CRF-RNN contribute to the improvement

[1] https://github.com/superxuang/caffe_3d_crf_rnn.

Table 1 Segmentation performance compared with state-of-the-art method

Approaches	DSC (%)	VOE (%)	RVD (%)	ASD (mm)	MSD (mm)	Average segmentation time (s)
V-net	84.12	27.40	− 3.91	3.39	28.37	0.9
Our approach	92.24	14.41	− 3.40	2.02	67.68	5.4

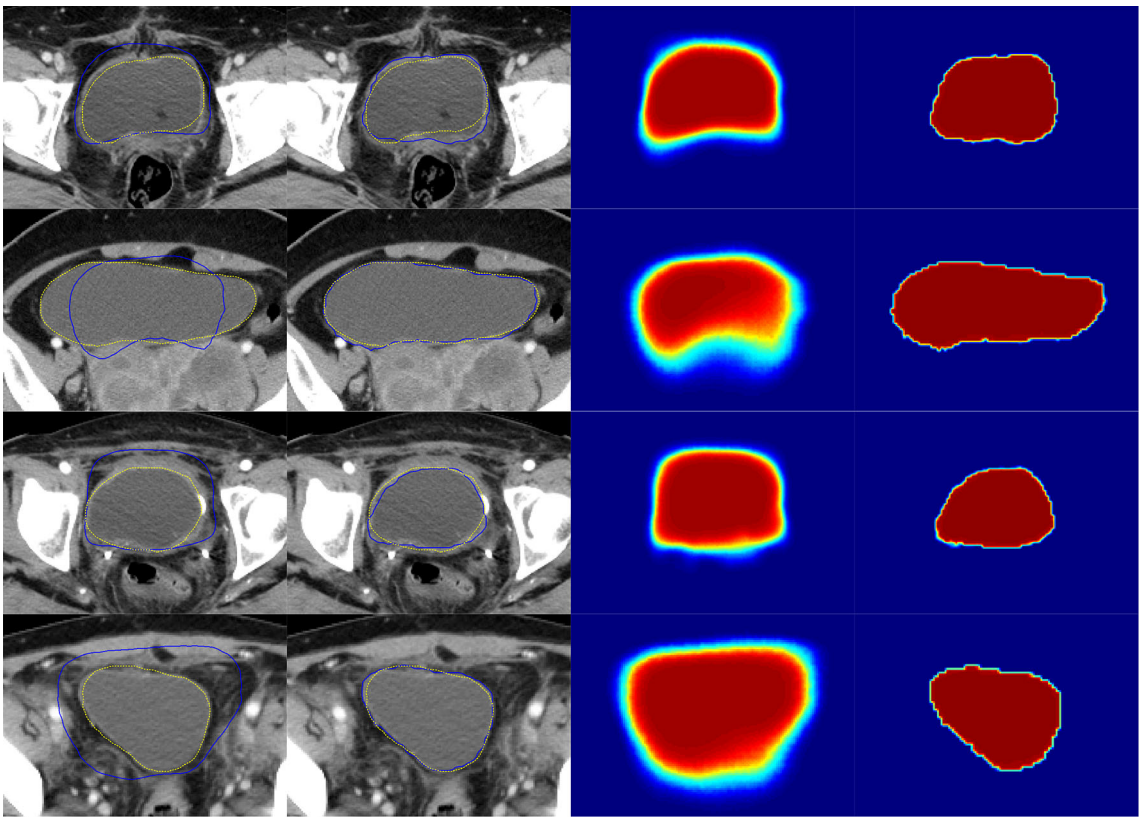

Fig. 3 Visualization of segmentation results. From left to right: segmentation results of V-net and our approach (the yellow and blue contours indicate the ground truth and segmentation result, respectively), bladder probability maps of V-net and our approach. Best viewed in color

of segmentation accuracy. Quantitative analysis is presented in "Ablation experiments" section.

Ablation experiments

In this section, we conduct a set of ablation experiments to investigate the behaviors of major components in our approach and justify the choices made in the design of our approach. The experiment results are summarized in Table 2 and discussed in detail next.

Segmentation CNN

As mentioned before, we add the batch normalization operation after each convolutional layer in the V-net architecture to build our segmentation CNN. In order to evaluate the impact of this modification, we perform a comparison of our segmentation CNN to the original V-net. Figure 4 presents the training losses and DSC curves of the V-net and our segmentation CNN. It is observed that our segmentation CNN converges faster than the V-net, demonstrating that the batch normalization operation can improve the learning convergence of the network. Table 2 further shows the performance of the V-net and our segmentation CNN. We can see that, although the batch normalization operation accelerates the network's convergence, it plays a very limited role in the improvement of segmentation accuracy. In this experiment, we do not use the dual-channel preprocessing and the result of our segmentation CNN will serve as a baseline for the following comparisons.

Dual-channel preprocessing

We then proceed to analyze the performance of our proposed dual-channel preprocessing. Figure 5 presents the training losses and DSC curves of V-net and our segmentation CNN with and without the dual-channel preprocessing. We can see that the dual-channel preprocessing can improve segmen-

Table 2 Results of ablation experiments

Approaches	DSC (%)	VOE (%)	RVD (%)	ASD (mm)	MSD (mm)
V-net	83.57	28.23	−3.25	3.48	28.37
Dual-channel preprocessing + V-net	91.35	15.92	−7.96	1.93	43.60
V-net + Batch normalization (our CNN)	84.13	27.40	−7.70	3.45	32.56
Dual-channel preprocessing + our CNN	91.90	14.99	−2.98	2.19	67.68
Our CNN + 3D CRF-RNN	88.39	20.80	−2.68	2.72	30.00
Our CNN + 3D CRF post-processing	87.81	21.72	−3.81	2.86	34.99
Our approach	92.24	14.41	−3.40	2.02	67.68

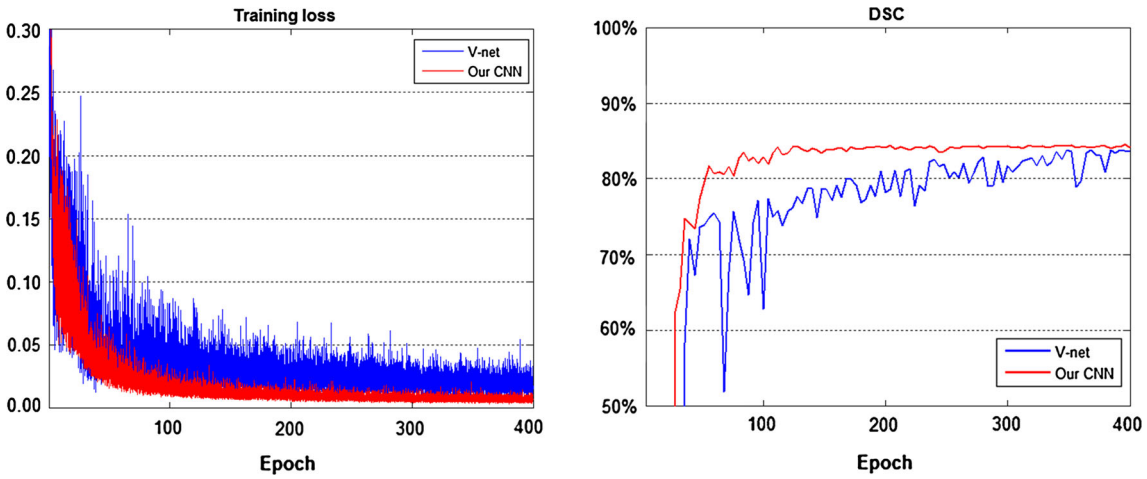

Fig. 4 Training losses and DSC curves of V-net and our CNN

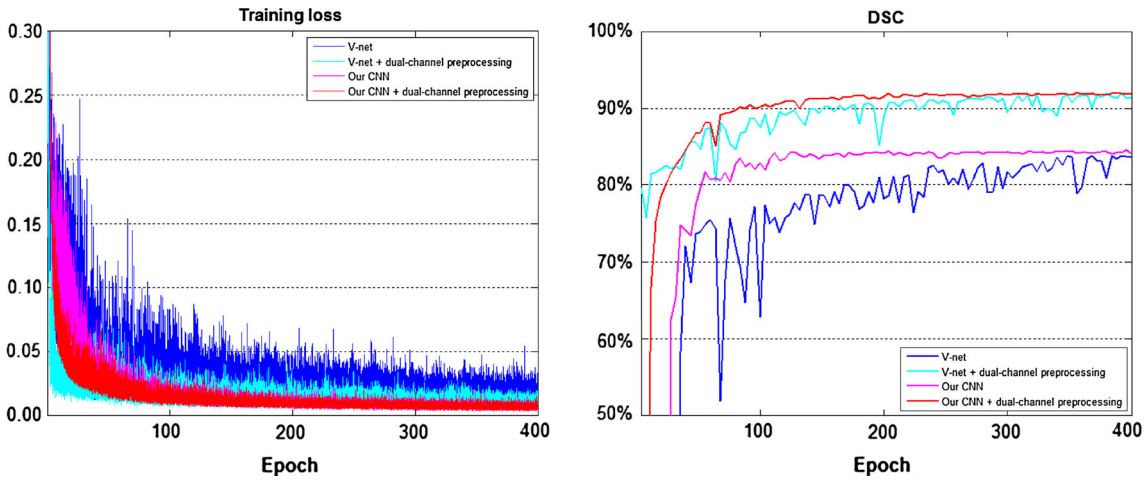

Fig. 5 Training losses and DSC curves of V-net and our CNN with and without dual-channel preprocessing

tation performance and convergence of both V-net and our CNN. Table 2 further shows the quantitative results. The network with dual-channel preprocessed data achieves higher segmentation accuracy than the one using single-channel CT images. This can be attributed to the fact that the additional channel, which stores the enhanced density map, emphasizes the structural information of bladder and makes the CNN learn more valuable features from CT images.

3D fully connected CRF-RNN

Next, we conduct an experiment to evaluate the influence of the 3D fully connected CRF-RNN in our approach. In this experiment, we connect the 3D fully connected CRF-RNN to the segmentation CNN and fine-tune them from the model trained in "Segmentation CNN" section. Note that, we do not use the dual-channel preprocessing in order to eliminate its

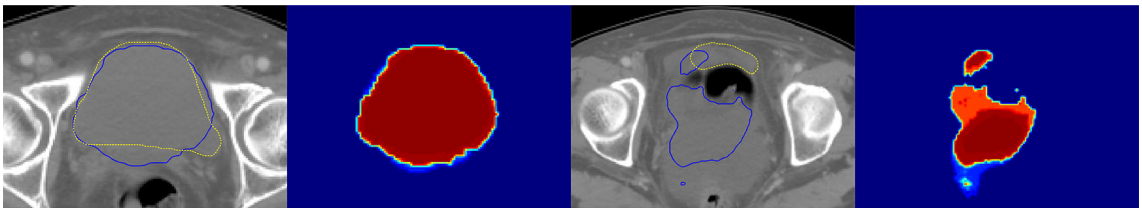

Fig. 6 Segmentation results (the yellow and blue contours indicate the ground truth and segmentation result, respectively) and bladder probability maps of our approach with the lowest DSC. Best viewed in color

influence on the final results. According to the quantitative results shown in Table 2, the 3D fully connected CRF-RNN improves the segmentation accuracy of our CNN. It demonstrates that the 3D fully connected CRF-RNN does refine the coarse bladder score maps output from the segmentation CNN by utilizing the smoothness constraints on voxel positions and appearances.

Additionally, we investigate the potential advantages of training the CRF and CNN jointly with respect to the offline application of CRF model. We post-process the segmentation results of our CNN in "Segmentation CNN" section by a 3D fully connected CRF model. This 3D CRF model is initialized using the same configurations as our 3D CRF-RNN. The results are reported in Table 2. Using the 3D fully connected CRF model as a post-processing filter achieves slightly lower segmentation accuracy than the application of 3D fully connected CRF-RNN. It demonstrates that, during the end-to-end training procedure, the parameters of the segmentation CNN and 3D CRF-RNN are fine-tuned. This makes these two networks cooperate better and produce higher segmentation accuracy than the case where they work separately.

Discussions

We further analyze the segmentation results of our approach with the lowest DSC as shown in Fig. 6. It can be seen that there are some false positive regions around the bladder. This is mainly due to the fact that soft-tissues/tumors whose density approximates to that of bladder have similar appearances to the bladder and the long-range connections in CRF propagate this misleading information to the segmentation CNN. The dual-channel preprocessing also facilitates the spread of misinformation. How to distinguish the soft-tissues/tumors which has a similar appearance to the bladder is left as a future work.

There are a number of hyper-parameters (e.g., learning rate, weight decay and momentum parameter) chosen based on the segmentation results on the validation set. Since we also use the validation set to evaluate the final performance of our approach, the reported results may be over-fitted and better than the performance on unseen test cases. Typically, the real performance is evaluated on a test set, which is kept "unseen" during the development of the approach. As a future work, we expect to collect more images and build a dedicated test set for the performance evaluation.

In the next stage of our work, we plan to exploit our approach to segment other organs, such as liver and pancreas. We believe that the principle and architecture of our approach is generalizable enough to be extended to other segmentation tasks. We are also interested in investigating the performance of our approach on other medical imaging modalities such as MRI.

Conclusions

We present an automatic deep learning-based approach for accurate bladder segmentation from CT images. In this approach, we exploit a 3D fully connected CRF-RNN cooperating with a CNN to produce fine-localized bladder probability maps and accurate segmentation results. Furthermore, a dual-channel preprocessing method is proposed to further boost the segmentation performance of our deep network. By introducing the batch normalization operation, we also accelerate the convergence of the deep network. Extensive experiments on clinical data corroborate the efficacy of our approach. It outperforms the state-of-the-art V-net on segmentation accuracy. Moreover, the architecture of our approach is generalizable and can be easily extend to other medical segmentation tasks.

Compliance with ethical standards

Funding This work is funded by the National Key R&D Program of China (2017YFC0113100) and the National Natural Science Foundation of China (61601012).

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval For this type of study formal consent is not required.

Informed consent Statement of informed consent is not applicable since the manuscript does not contain any participants' data.

References

1. Men K, Dai J, Li Y (2017) Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. Med Phys 44(12):6377–6389. https://doi.org/10.1002/mp.12602
2. Costa MJ, Delingette H, Novellas S, Ayache N (2007) Automatic segmentation of bladder and prostate using coupled 3D deformable models. In: International conference on medical image computing and computer-assisted intervention, pp 252–260. https://doi.org/10.1007/978-3-540-75757-3_31
3. Haas B, Coradi T, Scholz M, Kunz P, Huber M, Oppitz U, André L, Lengkeek V, Huyskens D, van Esch A, Reddick R (2008) Automatic segmentation of thoracic and pelvic CT images for radiotherapy planning using implicit anatomic knowledge and organ-specific segmentation strategies. Phys Med Biol 53(6):1751–1771. https://doi.org/10.1088/0031-9155/53/6/017
4. Zheng S , Jayasumana S, Romera-Paredes B, Vineet V, Su Z, Du D, Huang C, Torr PHS (2015) Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE international conference on computer vision, pp 1529–1537. https://doi.org/10.1109/ICCV.2015.179
5. 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, pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
6. Cirean DC, Giusti A, Gambardella LM, Schmidhuber J (2012) Deep Neural networks segment neuronal membranes in electron microscopy images. In: Advances in neural information processing systems, pp 2852–2860
7. Roth HR, Le L, Amal F, Shin H-C, Liu J, Turkbey E, Summers RM (2015) DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation. In: International conference on medical image computing and computer-assisted intervention, pp 556–564. https://doi.org/10.1007/978-3-319-24553-9_68
8. Han X (2017) Automatic liver lesion segmentation using a deep convolutional neural network method. arXiv preprint. arXiv:1704.07239. Accessed 24 Apr 2017
9. Zhou Y, Xie L, Shen W, Fishman E, Yuille A (2016) Pancreas segmentation in abdominal CT scan: a coarse-to-fine approach. arXiv preprint. arXiv:1612.08230. Accessed 25 Dec 2016
10. Chen H, Yu L, Dou Q, Shi L, Mok VCT, Heng PA (2015) Automatic detection of cerebral microbleeds via deep learning based 3D feature representation. In: International symposium on biomedical imaging, pp 764–767. https://doi.org/10.1109/isbi.2015.7163984
11. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M (2013) Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: International conference on medical image computing and computer-assisted intervention, pp 246–253. https://doi.org/10.1007/978-3-642-40763-5_31
12. Roth HR, Lu L, Seff A, Cherry KM, Hoffman J, Wang S, Liu J, Turkbey E, Summers RM (2014) A new 2.5D 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, pp 520–527. https://doi.org/10.1007/978-3-319-10404-1_65
13. Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C, van Riel SJ, Wille MMW, Naqibullah M, Sanchez CI, van Ginneken B (2016) Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imaging 35(5):1160–1169. https://doi.org/10.1109/tmi.2016.2536809
14. Milletari F, Navab N, Ahmadi S-A (2016) V-net: fully convolutional neural networks for volumetric medical image segmentation. In: Fourth international conference on 3d Vision, pp 565–571. https://doi.org/10.1109/3dv.2016.79
15. Çiçek Ö, Abdulkadir A, Lienkamp SS, 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, pp 424–432. https://doi.org/10.1007/978-3-319-46723-8_49
16. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, 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. https://doi.org/10.1016/j.media.2016.10.004
17. Chen H, Dou Q, Lequan Y, Heng P-A (2017) Voxresnet: deep voxelwise residual networks for brain segmentation from 3D MR images. Neuroimage. https://doi.org/10.1016/j.neuroimage.2017.04.041
18. Yu L, Yang X, Chen H, Qin J, Heng P-A (2017) Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images. In: Thirty-first AAAI conference on artificial intelligence
19. Christ PF, Ettlinger F, Grün F, Elshaer MEA, Lipkova J, Schlecht S, Ahmaddy F, Tatavarty S, Bickel M, Bilic P, Remper M, Hofmann F, D'Anastasi M, Ahmadi S-A, Kaissis G, Holch J, Sommer W, Braren R, Heinemann V, Menze B (2017) Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv preprint. arxiv:1702.05970. Accessed 20 Feb 2017
20. Roth HR, Oda H, Hayashi Y, Oda M, Shimizu N, Fujiwara M, Misawa K, Mori K (2017) Hierarchical 3D fully convolutional networks for multi-organ segmentation. arXiv preprint. arXiv:1704.06382. Accessed 21 Apr 2017
21. Shelhamer E, Long J, Darrell T (2017) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 39(4):640–651. https://doi.org/10.1109/TPAMI.2016.2572683
22. 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. https://doi.org/10.1109/cvpr.2016.90
23. Krähenbühl P, Koltun V (2012) Efficient inference in fully connected CRFs with Gaussian edge potentials. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 109–117. https://doi.org/10.1109/cvpr.2012.6247724
24. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017) DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell PP(99):1–1. https://doi.org/10.1109/tpami.2017.2699184
25. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint. arXiv:1502.03167. Accessed 11 Feb 2015
26. Lécun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324. https://doi.org/10.1109/5.726791
27. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the ACM international conference on multimedia, pp 675–678. https://doi.org/10.1145/2647868.2654889