Volumetric ConvNets with Mixed Residual Connections for Automated Prostate Segmentation from 3D MR Images

Lequan Yu,[*] **Xin Yang,**[*] **Hao Chen,**[*] **Jing Qin,**[†] **Pheng-Ann Heng**[*,‡]

[*]Department of Computer Science and Engineering, The Chinese University of Hong Kong
[†]Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University
[‡]Guangdong Provincial Key Laboratory of Computer Vision and Virtual Reality Technology,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
{lqyu,xinyang,hchen,pheng}@cse.cuhk.edu.hk{harry.qin}@polyu.edu.hk

Abstract

Automated prostate segmentation from 3D MR images is very challenging due to large variations of prostate shape and indistinct prostate boundaries. We propose a novel volumetric convolutional neural network (ConvNet) with mixed residual connections to cope with this challenging problem. Compared with previous methods, our volumetric ConvNet has two compelling advantages. First, it is implemented in a 3D manner and can fully exploit the 3D spatial contextual information of input data to perform efficient, precise and volume-to-volume prediction. Second and more important, the novel combination of residual connections (i.e., long and short) can greatly improve the training efficiency and discriminative capability of our network by enhancing the information propagation within the ConvNet both locally and globally. While the forward propagation of location information can improve the segmentation accuracy, the smooth backward propagation of gradient flow can accelerate the convergence speed and enhance the discrimination capability. Extensive experiments on the open MICCAI PROMISE12 challenge dataset corroborated the effectiveness of the proposed volumetric ConvNet with mixed residual connections. Our method ranked the first in the challenge, outperforming other competitors by a large margin with respect to most of evaluation metrics. The proposed volumetric ConvNet is general enough and can be easily extended to other medical image analysis tasks, especially ones with limited training data.

Introduction

Prostate diseases (e.g., prostate cancer, prostatitis and enlarged prostate) are very common in men. In particular, prostate cancer is the second leading cause of cancer death in American men. It is estimated to have caused 26,120 deaths in 2016 according to American Cancer Society (Siegel, Miller, and Jemal 2016). Due to a huge increase in screening, prostate cancer is now the most commonly diagnosed cancer in men in American. Accurate segmentation of prostate from 3D Magnetic Resonance (MR) images is very useful for treatment planning and many other diagnostic and therapeutic procedures for prostate cancer as well as other prostate diseases. However, manual segmentation from 3D MR images is time-consuming and subjective with limited reproducibility. It heavily depends on experience and has
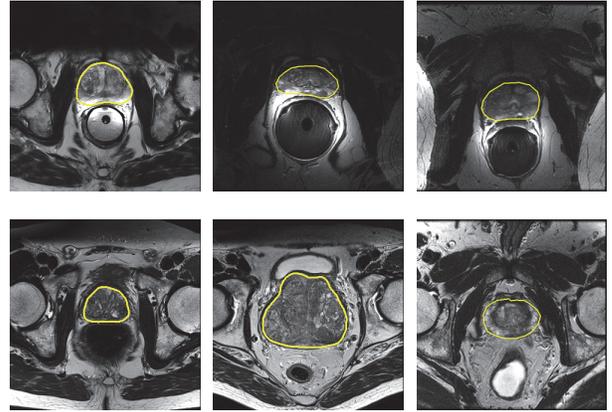
Figure 1: Example of prostate MR images displaying large variations (only show the center slice of prostates in 3D MR images and the yellow contours indicate prostates).

large inter- and intra-observer variations. In this regard, automated segmentation methods are highly demanded in clinical practice.

Automated segmentation of prostate from MR images, however, is very challenging for several reasons (Mahapatra and Buhmann 2014). First, different MR images have global inter-scan variability and intra-scan intensity variation due to different MR scanning protocols, such as with/without endorectal coil (a thin wire placed inside the body to generate detailed MR images). Second, the lack of clear prostate boundaries due to similar appearance of prostate and surrounding tissues (e.g., blood vessels, bladder, rectum and seminal vessels) makes the automated segmentation even harder. Third, prostate has a wide variation in size and shape among different subjects due to pathological changes or different resolutions of images. Figure 1 shows examples of segmented prostate in different MR images and we can see the large variation of prostates.

Over the past few years, lots of automated prostate segmentation methods have been proposed to meet these challenges, such as atlas (registration) based methods, deformable methods and machine learning based methods. Although great progress is achieved, there still exists an ob-

vious gap between the automated segmentation results and manual annotations.

Recently, deep convolutional neural networks (ConvNets) with hierarchical feature learning capability have become the dominant machine learning approach in computer vision field and have achieved promising results in different vision tasks (He et al. 2015; Long, Shelhamer, and Darrell 2015; Xie and Tu 2015). Some researchers have employed Con-vNets in automated prostate segmentation (Cheng et al. 2016). Nevertheless, most of these methods employed 2D ConvNets on 2D MR image slices and hence were incapable of taking full use of the 3D spatial information of the whole volumetric data for more accurate segmentation. Recently, many 3D ConvNets were proposed for both natural video and medical image analysis tasks (Ji et al. 2013; Tran et al. 2015; Kamnitsas et al. 2016). Nevertheless, how to train an efficient volumetric ConvNet under limited training data for medical image analysis applications is still a challenging problem.

Commonly, there are several schemes to dig the potential of the limited training data. The first is harnessing the data augmentation methods (Krizhevsky, Sutskever, and Hinton 2012). Such a scheme usually has performance gains by transforming the training data. However, the information added by these augmented data is limited and we need more efficient methods to further improve the performance. The second scheme is employing skip connections in ConvNet architecture to boost the information propagation within the ConvNet in order to achieve more efficient training with limited training data. Recently, a special skip connection, namely residual connection has been demonstrated as an effective mechanism to train very deep ConvNets and led to a series of breakthrough on some challenge datasets, such as ImageNet and MS COCO dataset (He et al. 2015). It has been demonstrated that the residual connections can improve the information flow within the networks and hence accelerate the convergence speed and improve the performance (He et al. 2016; Szegedy, Ioffe, and Vanhoucke 2016; Drozdzal et al. 2016).

In this paper, we distill the residual connections into a volumetric ConvNet and propose a novel learning architecture with mixed long and short residual connections for automated prostate segmentation from 3D MR images. Our volumetric ConvNet adopts fully convolutional architecture (Long, Shelhamer, and Darrell 2015) and can be trained end-to-end to perform efficient, precise and volume-to-volume prediction. Compared with previous works that just leverage the residual connections within local residual blocks, we extend the residual connections across residual blocks and promote the information exchange between the down-sampling path and up-sampling path in the fully convolutional architecture. By incorporating mixed long and short residual connections into our volumetric network, the information can be smoothly propagated throughout the network, which enhances the discriminative capability of networks and improves the training efficiency. Particularly, the long residual connections can recover the spatial information loss caused by down-sampling operations of the network and leverage the location information propagated from earlier layers of the network to achieve better segmentation results. We evaluated our method on the open MICCAI PROMISE12 challenge dataset; it ranked first in the challenge, outperforming other methods by a large margin.

Related Work

Prostate segmentation Previous automated prostate segmentation methods mainly include multi-atlas based methods (Klein et al. 2008) and deformable methods (Toth and Madabhushi 2012). For example, Klein et al. (2008) proposed an automated segmentation method based on atlas matching. In this method, several template images with corresponding segmentations are registered to the target image using a non-rigid registration method, and then the aligned segmentations are fused to obtain the final results. Toth et al. (2012) proposed an extension of active appearance model (AAM) to capture shape information with a multi-feature landmark-free framework. On the other hand, various graph cut based methods were also proposed to segment the prostate from MR images. Tian et al. (2016) proposed a superpixel-based 3D graph cut algorithm to obtain the prostate surface. Many successful approaches were proposed to use feature-based machine learning methods, such as K-nearest-neighbor, random forest classifier and marginal space learning (Zheng and Comaniciu 2014). Recently, with the impressive performance achieved by deep learning methods, some researchers proposed to utilize deep learning techniques to learn representation features for automated prostate segmentation. For example, Liao et al. (2013) proposed a deep learning framework with an independent subspace analysis network and Cheng et al. (2016) combined CNN and AAM methods for accurate prostate segmentation.

Deep learning for volumetric data processing In the field of medical image computing, many imaging modalities are volumetric, such as 3D Computed Tomography (CT) and MR Images. A lot of effort has been dedicated to employing CNNs to process volumetric data. Some of them employed variants of 2D CNNs to exploit adjacent slices (Chen et al. 2015), orthogonal planes (Prasoon et al. 2013; Roth et al. 2014) or multi-view planes (Setio et al. 2016) to aggregate 3D contextual feature in the model. However, these methods cannot effectively make full use of the 3D spatial information. Some studies started to employ 3D CNN to cope with detection and segmentation problems in medical volumetric data (Li et al. 2014; Brosch et al. 2016; Kamnitsas et al. 2016; Chen et al. 2016a). These networks adopted some relatively simple architectures and hence suffered from limited representation capability. By borrowing the design principles of U-Net (Ronneberger, Fischer, and Brox 2015), some researchers proposed 3D fully convolutional networks for efficient segmentation tasks (Çiçek et al. 2016; Milletari, Navab, and Ahmadi 2016; Merkow et al. 2016). Although these networks improved the segmentation performance, there was still much room to dig the potential of CNNs by effectively training the networks under limited training data. Our work is also related to Chen et al. (2016b), which is a 3D extension of residual networks for the segmentation task. But in our work, we extend the short residual
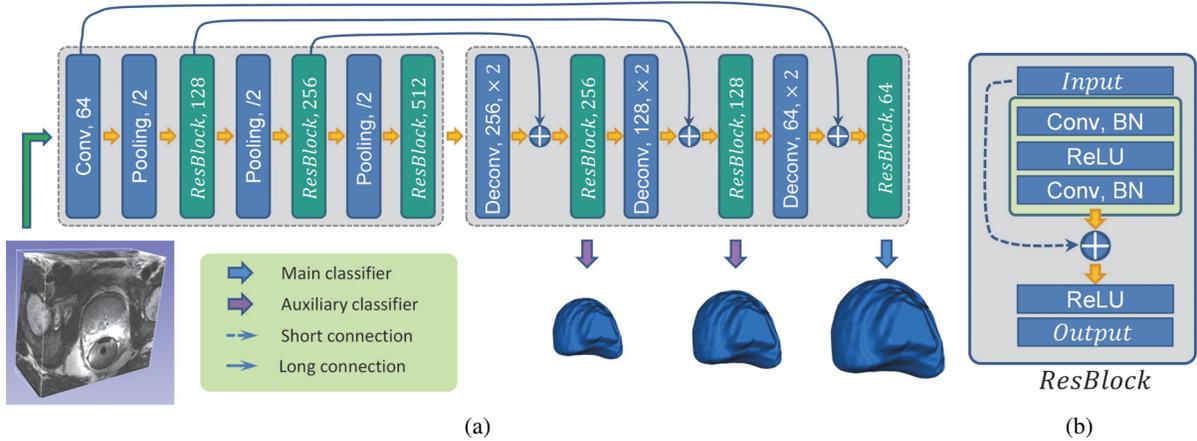
Figure 2: (a) The architecture of the proposed volumetric ConvNet. The number in each box represents the number of feature maps and all convolutional layers contain $3\times3\times3$ filter kernels. (b) The illustration of one residual block.

connections across residual blocks and seamlessly combine the long and short residual connections, which greatly improves the training efficiency and discriminative capability of our network under limited training data.

Method

Residual Connections

Deep ConvNets with residual connections have achieved promising results in many challenging natural image processing tasks. Residual connection, in principle, is a kind of skip connection that bypasses the non-linear transformations with an identity mapping and explicitly reformulates the layers as learning residual functions with reference to the layer inputs (He et al. 2015). Formally, the residual connection can be expressed as:

$$\mathbf{x}_\ell = \mathcal{H}_\ell(\mathbf{x}_{\ell-1}) + \mathbf{x}_{\ell-1}, \tag{1}$$

where $\mathbf{x}_{\ell-1}$ and $\mathbf{x}_\ell$ are input and output of the ℓth unit, and $\mathcal{H}_\ell(\cdot)$ denotes the residual function corresponding to ℓ^{th} unit.

While the residual connection is originally proposed to address the degradation problem when training a very deep ConvNet, a few recent studies have demonstrated that residual connection can also promote information propagation within a ConvNet both forward and backward (He et al. 2016), and hence accelerate its convergence and improve performance (Szegedy, Ioffe, and Vanhoucke 2016; Zagoruyko and Komodakis 2016). In this paper, we extend such a skip connection from local residual unit to the whole ConvNet and propose a learning architecture with mixed residual connections in order to further boost the information propagation during training. This is essential to generate an effective ConvNet under limited training samples commonly occurring in medical image analysis applications.

Our Basic Volumetric ConvNet

In order to fully leverage the 3D spatial contextual information of volumetric data to dig the potential of learning capa-

bility of ConvNet, we first extend a 2D fully ConvNet (FCN) into a volumetric ConvNet to enable volume-to-volume prediction. Our volumetric ConvNet is extended from the 2D architecture reported in (Long, Shelhamer, and Darrell 2015; Ronneberger, Fischer, and Brox 2015), which consists of two parts: a fine-to-coarse down-sampling path and a coarse-to-fine up-sampling path. The down-sampling path, consisting of convolutional and pooling layers, can extract abstract features and increase the receptive field of the ConvNet to enclose more contextual information. However, from the down-sampling path, we can only obtain a coarse prediction, which is sufficient for some detection and classification tasks but is unfit for voxel-wise semantic segmentation. Hence, an up-sampling path, consisting of deconvolutional and convolutional layers, is implemented to generate dense predictions with much higher resolutions (Long, Shelhamer, and Darrell 2015). Note that all layers in our ConvNet, including convolutional, pooling, and deconvolutional layers are implemented in a 3D manner, and thus the ConvNet can fully preserve and exploit the 3D spatial information when extracting features and making predictions.

In addition, we also integrate a deep supervision mechanism (Lee et al. 2015; Dou et al. 2016) in our volumetric ConvNet to accelerate its convergence speed. We exploit additional supervision injected into some hidden layers via auxiliary predictions. Specifically, we add one convolutional layer (kernel size $1\times1\times1$) at the end of the network to generate the main prediction. Besides, we also employ several convolutional layers (kernel size $1\times1\times1$) followed by hidden feature maps in the up-sampling path to obtain auxiliary coarse predictions, and then use deconvolutional layers to get auxiliary dense predictions with the same size of input. We minimize the weighted sum of cross-entropy losses of the main prediction and auxiliary predictions when training the volumetric ConvNet. In principle, the deep supervision mechanism can function as a strong "regularization" during training and thus it is important for training ConvNet with limited training data (Lee et al. 2015).

Our Volumetric ConvNet with Mixed Residual Connections

In order to improve the segmentation performance of our volumetric ConvNet under limited training data, we incorporate residual connections to our basic ConvNet to enhance the propagation of volumetric context information locally and globally, and by this way to achieve more performance gains from the limited training data. Specifically, we introduce two kinds of residual connections into our volumetric ConvNet. The first kind of residual connections are employed to construct the local residual blocks. It is the same connections with those reported in (He et al. 2015) and we refer them as short residual connections in this paper. The second kind of residual connections are applied to connect the residual blocks with the same resolution in the down-sampling and up-sampling paths. These connections can effectively propagate context and gradient information both forward and backward during the end-to-end training process. We refer these connections as long residual connections. Figure 2 (a) shows the architecture of our volumetric ConvNet with mixed residual connections for prostate segmentation from 3D MR images.

Our volumetric ConvNet consists of several residual blocks (ResBlocks). We design the residual blocks with short residual connections based on the design scheme of He et al. (2015). As shown in Figure 2 (b), it is composed of two convolutional layers and two rectified linear units (ReLUs). Previous studies (Simonyan and Zisserman 2014; Tran et al. 2015) have demonstrated that smaller convolutional kernels are more efficient in ConvNet design. Therefore, we adopt small convolution kernels with size of $3\times3\times3$ in convolutional layers. Each convolutional layer is followed by a ReLU as an activation function. We further employ batch normalization (BN) (Ioffe and Szegedy 2015) between each pair of convolution and ReLU, as the BN can reduce the internal covariance shift, and hence accelerate the training process and improve performance.

The down-sampling path contains one convolutional layer and three residual blocks, as shown in Figure 2 (a). Three $2\times2\times2$ max pooling layers with stride of 2 are applied between them. As for the up-sampling path, we also employ three residual blocks. Three deconvolutional layers with stride of 2 are employed to restore the feature map size. Note that in the down-sampling path, the input and output of residual blocks have different numbers of feature maps. We add $1\times1\times1$ convolutional layers in the short residual connections in the down-sampling path to match the dimensions of input and output (He et al. 2015).

To further boost the information exchange among different layers, we implement long residual connections to our volumetric ConvNet by connecting the corresponding residual blocks with the same resolutions in the down-sampling path and the up-sampling path using the same residual learning scheme as the short residual connections. These long residual connections are illustrated as the solid-line arrows in Figure 2 (a). These skip connections can explicitly propagate two kinds of important information within the ConvNet. First, they can propagate the spatial location infor-

mation forward to the up-sampling path in order to recover the spatial information loss caused by down-sampling operations for more accurate segmentation. Second, as we employ summation operations to construct the skip connections, our architecture can more smoothly propagate the gradient flow backward, and hence improve the training efficiency and network performance (He et al. 2016; Wang et al. 2016). In addition, such connections can also implicitly promote the information exchange between feature extraction and prediction and thus offer guidance for the training of earlier layers in the ConvNet, which can be considered as a kind of intermediate supervision to enhance the training. Overall, our volumetric ConvNet with mixed residual connections can be trained end-to-end and achieve efficient volume-to-volume prediction by improving the information propagation within the ConvNet.

Experiments and Results

Dataset and Pre-processing

We performed extensive experiments to evaluate our method on MICCAI Prostate MR Image Segmentation (PROMISE12) challenge dataset (Litjens et al. 2014), an ongoing benchmark for evaluating algorithms for segmentation of the prostate from MR images. The training dataset contains 50 transversal T2-weighted MR images of the prostate and corresponding segmentation ground truth. The testing dataset consists of 30 MR images and the ground truth is held out by the organizer for independent evaluation. These images are acquired in different hospitals, using different equipments and different acquisition protocols, and display the maximum variations of MR images acquired in clinical setting: there are variations in voxel size, dynamic range, position, field of view and anatomic appearance. Different from previous works using complex pre-processing steps, like N4 bias field correction, we simply resized all MR volumes into a fixed resolution of $0.625\times0.625\times1.5$ mm and then normalized them as zero mean and unit variance. We also utilized simple data augmentation strategy to enlarge the training dataset. The augmentation operations included rotation (90, 180 and 270 degrees) and flip in axial plane.

Implementation

Our volumetric ConvNet was implemented based on a modified Caffe library (Jia et al. 2014) supporting 3D operations. All the trainings and experiments were conducted on a workstation equipped with a NVIDIA TITAN X GPU. The networks were trained with Stochastic Gradient Descent (SGD) method with a mini-batch size of 8 due to the limited capacity of GPU memory. The learning rate was set as 0.001 initially and is divided by 10 every 3000 iterations; the models were trained for up to 10000 iterations. We employed a weight decay of 0.0005 and a momentum of 0.9. We utilized two auxiliary predictions in deep supervision scheme and the balancing weights were 0.3 and 0.6, respectively. Due to the limited GPU memory, we randomly cropped $64\times64\times16$ sub-volumes from every sample as input when training the network. In the test phase, we used overlapped sliding windows strategy to crop sub-volumes and then used the av-
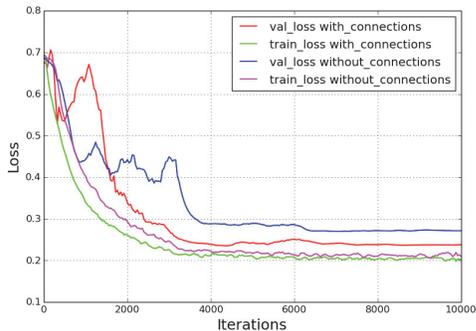
Figure 3: Training and validation loss of networks with and without mixed residual connections.

Table 1: Cross validation performance of our volumetric ConvNets with different configurations.

Method	Dice coefficient [%]
Long-short connections	**86.93**
Only short connections	84.68
Only long connections	84.38
Without residual connections	81.63

erage of the probability maps of these sub-volumes to get the whole volume prediction. The sub-volume size was also $64\times64\times16$ and the stride was $50\times50\times12$. Generally, it took about 4 hours to train the network and about 12 seconds for processing one MR images with size of $320\times320\times60$.

Ablation Analysis of Residual Connections

In order to evaluate the effectiveness of the residual connections in our volumetric ConvNet, we performed a set of ablation experiments on the ConvNet. Because the ground truth of testing data is held out by organizers and the challenge organizers only allow resubmission of substantially different methods, we conducted experiments via standard 10-fold cross validation scheme.

We first analyze the learning behaviors of our volumetric ConvNet with and without mixed residual connections. Figure 3 presents the training and validation losses of different networks with and without mixed residual connections. It is observed that the ConvNet with mixed residual connections converges faster and achieves lower validation loss than the one without mixed residual connections, demonstrating that residual connections can improve the training efficiency of the volumetric ConvNet. Table 1 further shows the performance of our volumetric networks with different residual connections via cross validation. It is observed that adding either long or short residual connections can achieve better Dice performance than networks without residual connections, demonstrating residual connections, as a general skip connection design strategy, can improve the discriminative capability of networks by promoting information propagation. The network with only short residual connections
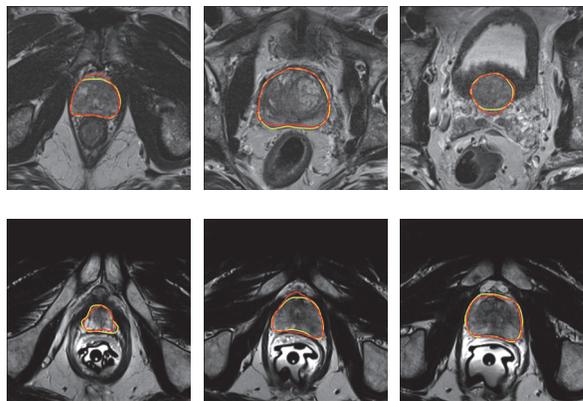

Figure 4: Qualitative segmentation results of case 4 (first row) and case 22 (second row) at the apex(left), center (middle) and base (right) of the prostate in testing dataset. The yellow and red contours indicate the ground truth and our segmentation results. Note that these results are directly obtained from challenge website.

has marginally better performance than that with only long residual connections. The network with mixed residual connections achieves the best performance in the ablation experiments, indicating that long and short residual connections can provide different information propagation locally and globally in the volumetric ConvNet and combining them together can further improve the performance.

Comparison with Other Methods

The evaluation metrics used in the PROMISE12 challenge include the Dice coefficient (DSC), the percentage of the absolute difference between the volumes (aRVD), the average over the shortest distance between the boundary points of the volumes (ABD) and the 95% Hausdorff distance (95HD). Note that all evaluation metrics including both boundary and volume metrics were calculated in 3D not only for the entire prostate segmentation but also specially for the apex and base parts of the prostate, which are the most difficult yet important parts in clinical practice (Litjens et al. 2014). The organizers calculated a total score incorporating above completely different but equally important metrics to rank the submitted methods. Readers can refer to Litjens et al. (2014) for more evaluation details.

Some qualitative results of our method are shown in Figure 4. It is observed that our method can produce accurate segmentation results and delineate the clear contours of prostates in MR images with (case 22) and without (case 4) endorectal coil. The quantitative results of our method and our competitors are shown in Table 2. There were totally 21 teams submitting their results until the paper submission and only top 10 teams are listed in the Table.[1] Note that all the results reported in this section were obtained directly from the organizers.

---

[1] Complete results can be found in https://grand-challenge.org/site/promise12/results/

Table 2: Quantitative comparison between the proposed method and other methods

Method	Type	ABD [mm]			95HD [mm]			DSC [%]			aRVD [%]			Score
		Whole	Base	Apex	Whole	Base	Apex	Whole	Base	Apex	Whole	Base	Apex	
CUMED (ours)	Auto	**1.95**	**2.13**	**1.74**	5.54	**5.41**	**4.29**	**89.43**	**86.42**	**86.81**	**6.95**	**11.04**	**15.18**	**86.65**
Imorphics	Auto	2.10	2.18	1.96	5.94	5.45	4.73	87.99	86.06	84.53	11.65	13.33	20.75	84.36
Emory	Semi	2.14	2.65	2.41	**5.04**	6.03	5.31	87.03	83.52	81.53	8.64	15.70	20.32	83.66
ScrAutoProstate	Auto	2.13	2.23	2.18	5.58	5.60	4.93	87.45	86.30	83.47	13.56	14.46	23.78	83.49
CAMP-TUM2	Auto	2.23	2.46	2.03	5.71	5.84	4.62	86.91	84.31	84.40	14.98	20.84	21.21	82.39
ETHZ	Semi	2.25	2.40	2.52	5.95	5.96	5.99	86.74	84.43	78.16	15.05	19.83	25.14	81.24
SIATMIDS	Auto	2.49	2.58	2.76	6.17	6.21	6.09	84.29	83.20	75.75	12.60	16.16	27.51	80.85
CBA	Interactive	2.33	2.60	2.44	6.57	6.64	5.75	86.56	84.33	80.31	15.49	23.17	23.59	80.66
CAMP-TUM	Auto	2.48	2.96	2.47	5.77	6.48	5.31	84.65	79.49	80.62	15.77	25.59	24.62	79.65
SBIA	Auto	2.85	2.82	2.13	7.73	6.99	4.60	83.55	81.06	83.90	22.78	26.94	24.52	78.34

For ABD, 95HD and aRVD, lower values are better; for DSC, higher values are better.

Seven of the top ten teams employed various hand-crafted features. Besides team (*CUMED*), the other two teams that utilized ConvNet are *SIATMIDS* and *CAMP-TUM2*. The team *SIATMIDS* proposed a multi-stage method that combines the hand-crafted features and features learned from a ConvNet. The team *CAMP-TUM2* harnessed a ConvNet implemented based on V-net (Milletari, Navab, and Ahmadi 2016). However, both of the two teams did not achieve better performance than the algorithms with hand-crafted features. This indicates that, while ConvNet has achieved remarkable succuss in many medical image analysis applications, it is still difficult for it to obtain satisfactory results for some medical applications, where the training data are quite limited. In this regard, it is essential to design more efficient learning algorithms and architectures to dig the potential of the limited training data and improve the performance of ConvNet. Our method ranked the first and achieved the best performance in all metrics except the metric 95HD on the whole prostate, demonstrating the effectiveness of the proposed volumetric ConvNet with mixed residual connections, which improve the performance of ConvNet by enhancing the information propagation through residual connections.

Conclusion

We present a novel volumetric ConvNet with mixed long and short residual connections for automated prostate segmentation from MR images. Our method adopts 3D fully convolutional architecture and is very efficient when handling large MR images. The incorporation of residual connections improves the network performance and indicates the effectiveness of enhancing information propagation through skip connections (especially residual connections) in 3D ConvNet architecture design. We also demonstrate that the combination of long and short residual connections can further improve the segmentation performance. Extensive experiments on an open challenge dataset corroborate the efficacy of our method in dealing with medical 3D segmentation under limited training data. In addition, our network architecture is general enough and can be easily extended to other applications. Future investigations include assessing our method on more 3D volumetric data and further exploring the information propagation mechanism of residual connections.

References

Brosch, T.; Tang, L. Y.; Yoo, Y.; Li, D. K.; Traboulsee, A.; and Tam, R. 2016. Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE transactions on medical imaging* 35(5):1229–1239.

Chen, H.; Yu, L.; Dou, Q.; Shi, L.; Mok, V. C.; and Heng, P. A. 2015. Automatic detection of cerebral microbleeds via deep learning based 3d feature representation. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 764–767. IEEE.

Chen, H.; Dou, Q.; Wang, X.; Qin, J.; Cheng, J. C.; and Heng, P.-A. 2016a. 3d fully convolutional networks for intervertebral disc localization and segmentation. In *International Conference on Medical Imaging and Virtual Reality*, 375–382. Springer.

Chen, H.; Dou, Q.; Yu, L.; and Heng, P.-A. 2016b. Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation. *arXiv preprint arXiv:1608.05895*.

Cheng, R.; Roth, H. R.; Lu, L.; Wang, S.; Turkbey, B.; Gandler, W.; McCreedy, E. S.; Agarwal, H. K.; Choyke, P.; Summers, R. M.; et al. 2016. Active appearance model and deep learning for more accurate prostate segmentation on mri. In *SPIE Medical Imaging*, 97842I–97842I. International Society for Optics and Photonics.

Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S. S.; Brox, T.; and Ronneberger, O. 2016. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In *MICCAI*, 424–432. Springer.

Dou, Q.; Chen, H.; Jin, Y.; Yu, L.; Qin, J.; and Heng, P.-A. 2016. 3d deeply supervised network for automatic liver segmentation from ct volumes. In *MICCAI*, 149–157. Springer.

Drozdzal, M.; Vorontsov, E.; Chartrand, G.; Kadoury, S.; and Pal, C. 2016. The importance of skip connec-

tions in biomedical image segmentation. *arXiv preprint arXiv:1608.04117*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Identity mappings in deep residual networks. *arXiv preprint arXiv:1603.05027*.

Ioffe, S., and Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Ji, S.; Xu, W.; Yang, M.; and Yu, K. 2013. 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence* 35(1):221–231.

Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; and Darrell, T. 2014. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*.

Kamnitsas, K.; Ledig, C.; Newcombe, V. F.; Simpson, J. P.; Kane, A. D.; Menon, D. K.; Rueckert, D.; and Glocker, B. 2016. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *arXiv preprint arXiv:1603.05959*.

Klein, S.; van der Heide, U. A.; Lips, I. M.; van Vulpen, M.; Staring, M.; and Pluim, J. P. 2008. Automatic segmentation of the prostate in 3d mr images by atlas matching using localized mutual information. *Medical physics* 35(4):1407–1417.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097–1105.

Lee, C.-Y.; Xie, S.; Gallagher, P.; Zhang, Z.; and Tu, Z. 2015. Deeply-supervised nets. In *AISTATS*.

Li, R.; Zhang, W.; Suk, H.-I.; Wang, L.; Li, J.; Shen, D.; and Ji, S. 2014. Deep learning based imaging data completion for improved brain disease diagnosis. In *MICCAI*, 305–312. Springer.

Liao, S.; Gao, Y.; Oto, A.; and Shen, D. 2013. Representation learning: a unified deep learning framework for automatic prostate mr segmentation. In *MICCAI*, 254–261. Springer.

Litjens, G.; Toth, R.; van de Ven, W.; Hoeks, C.; Kerkstra, S.; van Ginneken, B.; Vincent, G.; Guillard, G.; Birbeck, N.; Zhang, J.; et al. 2014. Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. *Medical image analysis* 18(2):359–373.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Mahapatra, D., and Buhmann, J. M. 2014. Prostate mri segmentation using learned semantic knowledge and graph cuts. *IEEE Transactions on Biomedical Engineering* 61(3):756–764.

Merkow, J.; Kriegman, D.; Marsden, A.; and Tu, Z. 2016. Dense volume-to-volume vascular boundary detection. *arXiv preprint arXiv:1605.08401*.

Milletari, F.; Navab, N.; and Ahmadi, S.-A. 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *IEEE International Conference on 3DVision (arXiv:1606.04797)*.

Prasoon, A.; Petersen, K.; Igel, C.; Lauze, F.; Dam, E.; and Nielsen, M. 2013. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *MICCAI*, 246–253. Springer.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 234–241. Springer.

Roth, H. R.; Lu, L.; Seff, A.; Cherry, K. M.; Hoffman, J.; Wang, S.; Liu, J.; Turkbey, E.; and Summers, R. M. 2014. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In *MICCAI*, 520–527. Springer.

Setio, A. A. A.; Ciompi, F.; Litjens, G.; Gerke, P.; Jacobs, C.; van Riel, S. J.; Wille, M. M. W.; Naqibullah, M.; Sánchez, C. I.; and van Ginneken, B. 2016. Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks. *IEEE transactions on medical imaging* 35(5):1160–1169.

Siegel, R. L.; Miller, K. D.; and Jemal, A. 2016. Cancer statistics, 2016. *CA: a cancer journal for clinicians* 66(1):7–30.

Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *CoRR* abs/1409.1556.

Szegedy, C.; Ioffe, S.; and Vanhoucke, V. 2016. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261*.

Tian, Z.; Liu, L.; Zhang, Z.; and Fei, B. 2016. Superpixel-based segmentation for 3d prostate mr images. *IEEE transactions on medical imaging* 35(3):791–801.

Toth, R., and Madabhushi, A. 2012. Multifeature landmark-free active appearance models: application to prostate mri segmentation. *IEEE Transactions on Medical Imaging* 31(8):1638–1650.

Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; and Paluri, M. 2015. Learning spatiotemporal features with 3d convolutional networks. In *2015 IEEE International Conference on Computer Vision (ICCV)*, 4489–4497. IEEE.

Wang, J.; Wei, Z.; Zhang, T.; and Zeng, W. 2016. Deeply-fused nets. *arXiv preprint arXiv:1605.07716*.

Xie, S., and Tu, Z. 2015. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, 1395–1403.

Zagoruyko, S., and Komodakis, N. 2016. Wide residual networks. *arXiv preprint arXiv:1605.07146*.

Zheng, Y., and Comaniciu, D. 2014. *Marginal Space Learning for Medical Image Analysis*, volume 662. Springer.