Residual Semantic Segmentation of the Prostate from Magnetic Resonance Images

Md Sazzad Hossain(✉), Andrew P. Paplinski, and John M. Betts

Faculty of Information Technology, Monash University, Melbourne, Australia
{sazzad.hossain,andrew.paplinski,
john.betts}@monash.edu

Abstract. The diagnosis and treatment of prostate cancer requires the accurate segmentation of the prostate in Magnetic Resonance Images (MRI). Manual segmentation is currently the most accurate method of performing this task. However, this requires specialist knowledge, and is time consuming. To overcome these limitations, we demonstrate an automatic segmentation of the prostate region in MRI images using a VGG19-based fully convolutional neural network. This new network, VGG19RSeg, identifies a region of interest in the image using semantic segmentation, that is, a pixel-wise classification of the content of the input image. Although several studies have applied fully convolutional neural networks to medical image segmentation tasks, our study introduces two new forms of residual connections (remote and neighbouring) which increases the accuracy of segmentation over the basic architecture. Our results, using this new architecture, show that the proposed VGG19RSeg can achieve a mean Dice Similarity Coefficient of 94.57%, making it more accurate than comparable methods reported in the literature.

Keywords: Prostate · MRI images · Semantic segmentation
Deep convolutional neural networks

1 Introduction

Prostate cancer is one of the most common and lethal cancers among men, with over 221,000 cases diagnosed in 2015 in the United States [1]. In Australia, approximately 17,000 males were diagnosed with prostate cancer, and it was the cause of death for approximately 3,500 men in 2017 [2]. Medical imaging is used in many cases to aid the diagnosis and treatment of prostate cancer. MRI is used for the detection of tumours and for treatment planning whereas Trans-Rectal Ultrasound (TRUS) is often used in a clinical setting, to guide needle placement for biopsies and treatments such as brachytherapy [3]. A current goal in medical imaging is to fuse images from these two modalities (MRI and TRUS) by the creation of a 3D model of the patient's prostate. Automatic image segmentation (the identification of organ boundaries) is a necessary task in achieving this goal.

L. Cheng et al. (Eds.): ICONIP 2018, LNCS 11307, pp. 510–521, 2018.
https://doi.org/10.1007/978-3-030-04239-4_46

Despite significant progress in prostate segmentation, fully automatic segmentation still remains challenging due to the deformability of the organ and the typically low resolution of MRI and TRUS images. Therefore, in many situations, segmentation is usually manually performed, which requires skilled clinicians, and is time-consuming. Because of the low accuracy of automatic segmentation, much recent work on MRI-TRUS registration still relies on manual segmentation by experts, for example [4, 5]. To address this challenge, we present a novel method of fully automatic segmentation of the MRI images of prostate using deep convolutional neural networks with residual connections.

Convolutional Neural Networks (CNNs) were designed and implemented primarily to classify objects in images [6]. These networks typically have a structure consisting of kernels or filters, activation functions and subsampling arranged in 16–200 layers. This enables a very detailed feature representation of an image to be made, in order to identify given objects. This gives rise to the terms 'convolutional' and 'deep' in describing these networks. Besides classification, CNNs have been adopted in segmentation tasks as well. Many recent medical imaging studies have used CNNs for automatic segmentation of regions of interest.

Non-CNN based medical image segmentation procedures have previously mainly been based on feature engineering. For example, Liao et al. [7] proposed an automatic feature extraction procedure using representation learning and a deep learning framework to identify the most significant parts of the extracted features. Toth et al. [8] proposed an advanced feature selection algorithm by modifying the typical Active Appearance Model (AAM) utilizing level-set representation. Yan et al. [9] presented a partial contour based segmentation method utilizing an a priori shape. These methods rely greatly on identifying suitable feature information. Another popular segmentation method is based on the concept of the nearest neighbours as in [10]. The CNN based procedures here seems to be more straightforward and more efficient, because CNNs combine both feature extraction and classification.

Feature extraction by deep learning is more effective in the sense that the algorithm 'learns' how to extract the features by utilizing real-world samples. As a consequence, many researchers are now investigating how CNNs may be used for segmentation tasks on medical images. Pereira et al. [11] used a CNN to segment a brain tumour. Havaei et al. [12] segmented brain tumours in MRI slices using two concatenated CNNs where the output probabilities from one CNN was fed to another CNN. Jia et al. implemented a pre-trained CNN, 'VGG-19' [13], to identify the prostate boundary in MRI images. Instead of segmenting at once, they first performed a coarse segmentation using atlas registration to obtain a rough boundary around the prostate, after which the VGG-19 CNN was applied to refine the segmentation. Zhang et al. segmented infant brain tissue by applying CNNs on multi-modal MRI images of brain. Other researchers, [14–16], have achieved segmentation by dividing the images into subregions known as 'patches' and by training a convolutional neural network on those patches. However, according to [17], training with whole image is more efficient and effective than patch-wise training.

More recently, researchers have developed methods to implement deep neural networks for segmentation of images automatically into particular classes using pixel-wise classification of the whole image known as semantic segmentation [17, 18]. For example, when provided an image of a cat and a dog together, semantic segmentation will identify the regions containing cats or dogs in that image. Using this method, Tian et al. [19]

segmented the prostate region from MRI images. Ahmad et al. [20] segmented human thigh quadriceps and Tran [21] presented an automated cardiac segmentation procedure using CNN based semantic segmentation. However, the network models used in these studies for semantic segmentation do not possess neighbouring residual connections, which can notably improve accuracy. The reason is that residual connection bypasses information, i.e. input to a deeper layer rather than just to the adjacent layer. He et al. [22] proved that residual connections are important to effectively train deep neural network models, because residual mapping is easier to optimize than unreferenced mapping of a stacked layer series. In another study [23], performance of several CNN were examined with and without residual connections and it was found that residual connections accelerate training speed and improve accuracy notably.

Therefore, this study uses residual connection on a VGG19 based fully convolutional network (FCN) to deliver more efficient network for automatic identification and localization of the prostate gland. Although remote residual connection between the convolution and the deconvolution part of an FCN has been implemented previously such as U-net [24] and V-net [25], residual connections between neighbouring layers in combination with residual connections between convolution and deconvolution layers in an FCN is still unreported to this date.

To extend the work by previous researchers we introduce a novel semantic segmentation of MRI prostate image slices using an adaptation of the popular deep neural network model, VGG19 [26]. Our new model, VGG19RSeg, modifies the original network by adding residual/skip connections between neighbouring and distant layers, thereby creating a semantic segmentation structure as explained in Sect. 2.3. Hence, this study first introduces an FCN with residual connections between stacked convolution layers and inter-convolution-deconvolution layers.

The following section describes details of the methodology applied in the research reported in this paper, including details of our VGG19RSeg network starting with its predecessor, the VGG19 network. Section 3 then presents computational experiments and results.

2 Methodology

2.1 Convolutional Neural Networks

A significant number of papers and tutorials describe convolutional neural networks in a great detail. To establish our notation, we say that each convolutional layer performs the following feedforward operations:

$$U = convT(X, W) + b; Z = \max(U, 0) \tag{1}$$

where X is a $M \times N \times D_x$ input tensor (a colour image, in particular), W is a $P \times Q \times D$ filter/kernel, $convT(\cdot)$ denotes a tensor, i.e., multidimensional convolution that generates a tensor U typically of dimensions $M \times N \times D_w$. The bias b has a matching dimension. The output of the convolutional layer, Z, represents the output of the ReLU activation function applied on U. The output Z becomes the input tensor for the next

layer. It is typically referred to as a feature map. Convolutional layers are typically separated by a subsampling operation that reduces the dimension of the feature maps. The most common method for subsampling operation is 'maxpooling', which replaces a square subregion of a feature map of size $c \times c$ with a single value being equal to the largest value in the subregion. A composition of the convolutional layers is followed by one to three fully connected layers as in "traditional" neural networks. For completeness of the description, we say that during the learning procedure all the weight filters W are modified using a stochastic gradient error-backpropagation algorithm that minimizes the specific loss function L.

2.2 The VGG19 Model

In 2014, Simonyan and Zisserman introduced the very deep network model, VGG16 [26]. This network was composed of 3 sequential convolutional layers, each having a max pooling layer to reduce the volume size. Followed by these convolutional layers, there are two fully connected layers containing 4096 neurons each. The final layer – a softmax layer, follows the fully connected layer, and delivers the ultimate probabilistic output of classification. In the same study, they modified the network by adding 3 extra weight layers thus creating a VGG19 network which outperformed VGG16. Although other deep learning models such as AlexNet, ResNet and InceptionNet have been developed [23], VGG models have proven to be more accurate image classifiers due to their simpler yet very deep network architecture compared to those other models [26]. Furthermore, VGG models appear to have a greater accuracy for semantic segmentation tasks [17, 18].

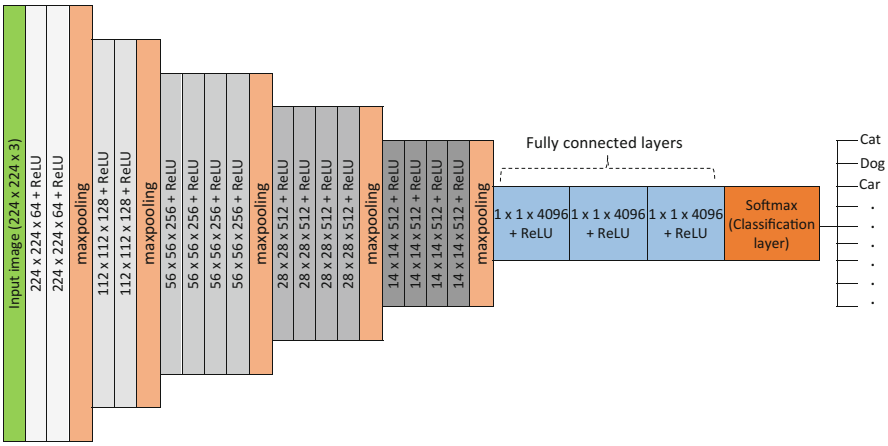

Fig. 1. VGG19 deep neural network model.

As Fig. 1 shows, the VGG19 structure beginning with an RGB image of the size $224 \times 224 \times 3$. There are total 16 convolution layers, where most of them are connected consecutively having 5 maxpooling stages/layers. Each convolution layer output goes

through a non-linear ReLU activation function. The whole convolution part of the network can be divided into 5 subregions, where each subregion is followed by a maxpooling layer to reduce the learnable parameters. The first two subregions consist of 2 consecutive convolution layers and the remaining 3 subregions comprise 4 consecutive convolution layers each. Convolution layers in same subregion each produce the same number of feature maps D_w. In the first region layers $D_w = 64$, in the second region $D_w = 128$, in the third region layers $D_w = 256$, and in the fourth and fifth regions $D_w = 512$. The product of the final maxpooling layer is flattened and passes through 3 fully connected layers comprising 4096 neurons each. The output of the final fully connected layer is then processed through classification layer which produces a probabilistic output classifying the input image, for example, as 97% cat, 2% dog, 1% car etc.

2.3 Semantic Segmentation

The structure of the semantic segmentation network consists of two asymmetric parts identified as a convolutional and the de-convolutional parts, respectively, as shown in Fig. 2. Deep neural network models have been designed to classify visual objects by giving a probability of the object belonging to each of the classes for which the neural network has been designed to classify. For example, a picture of a dog may yield a 90% probability associated with "dog", and smaller probabilities associated with the other outputs in order that the total equals to 1.

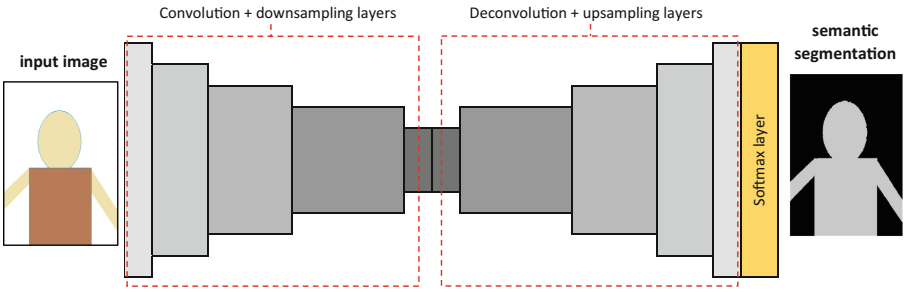

Fig. 2. An example of a structure of a semantic segmentation network.

For semantic segmentation, the input and output are both images, where the input image is a regular image and output is a 'pixel-class' version of the input image. That is, each pixel of the output image belongs to a certain class defined by the user for the corresponding input image. Thus, semantic segmentation is a pixel-wise classification of a given image. Since convolution and downsampling layers take an input image and break it down hierarchically, a reverse procedure is employed in place of fully connected layer (Fig. 2), by which deconvolution, upsampling and finally a softmax classifier, will produce a pixel-classified, i.e. semantically segmented image instead of just one object label. For instance, Fig. 2 shows an image of a human body that is applied to a trained convolutional neural network to identify the regions of the human in the image by a distinguishable colormap. Here, the grey region indicates the human and the black region is the background.

2.4 The Structure of the Residual VGGRSeg Network for Semantic Segmentation of Prostate

Because VGG19 was designed for object classification in images, adaptation for pixel classification requires that its fully-connected layers and the softmax layer are replaced with deconvolution and upsampling layers whose architecture is like a reflection of the convolution and downsampling layers shown in Fig. 2.

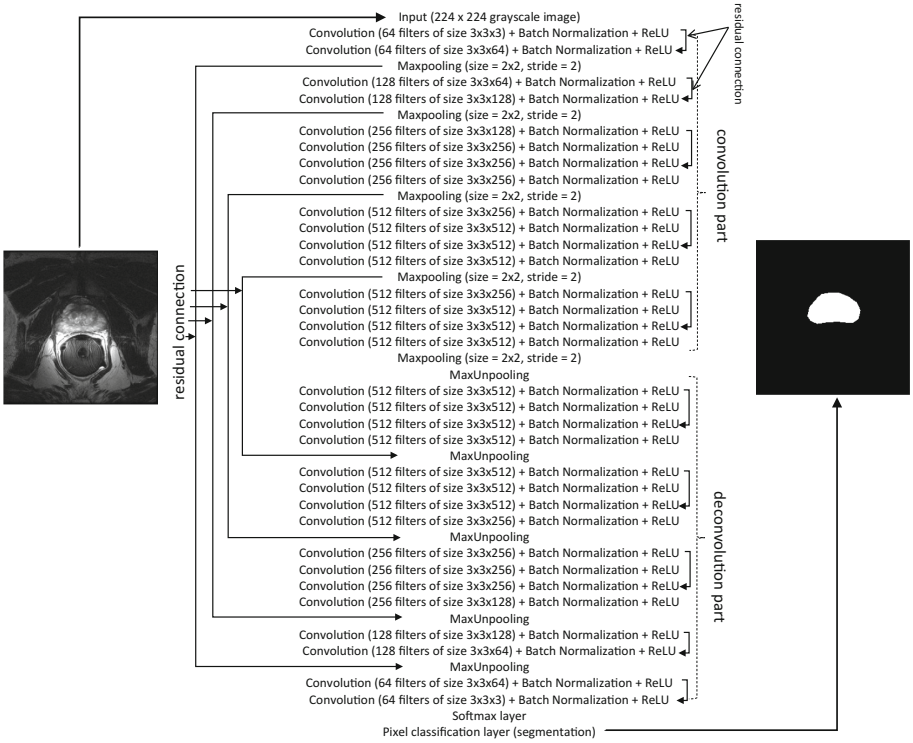

Fig. 3. Configuration of the VGG19RSeg network for semantic segmentation of prostate.

Each slice within the 3D MRI image stack was processed individually as a 2D monochromatic image. Therefore, the initial number of channels of VGG19 was changed from 3 to 1. Since VGG19 was designed to take input images of 224 × 224 pixels, we also resampled each MRI slice to this size to maintain consistency. The primary part of the proposed network is the residual connections between layers as shown in Fig. 3. Firstly, 'Maxpooling' layers from the convolution part and their corresponding mirror reflection layers in the deconvolution part, i.e. 'MaxUnpooling' were residually connected. Networks with such distantly connected layers are also known as Directed Acyclic Graph (DAG) networks. As shown by previous researchers [27, 28], a layer in a DAG network can connect, i.e. receive or provide inputs to a remote layer which makes the network more complex and more effective for

classification. Secondly, during convolution, at each stack of convolution layers before 'maxpooling', the first and the last convolution layers were residually connected. Such connections between neighbouring stacked layers were inspired from ResNet structure [22], which won first place in ILSVRC 2015 image classification competition. Therefore, due to residual connections and VGG19 architecture based FCN, the proposed network has been named as VGG19RSeg.

2.5 Training Procedure

To train the segmentation network, the ground truth segmentation of input images was provided as the output of the network for supervised learning. The dataset used in this research was PROMISE12 [29] grand challenge datasets for prostate segmentation from MRI. MRI images from 49 patients, consisting of total 1377 slices, were used in this study. 90% of slices were used for training and the remainder for testing. Images were pre-processed by taking the square root of each pixel value. These were then normalized to be between 0 and 255. As explained in [19], because the prostate is much smaller than the surrounding region in each image, a class-weighted cross entropy loss function was used as Eq. 2.

$$L = -\frac{1}{n}\sum_{i=1}^{N} w_i^{cl}[\hat{P}_i log P_i + (1-\hat{P}_i)\log(1-P_i)] \tag{2}$$

where, $w_i^{cl} = \frac{1}{pixels\ of\ class\ x_i}$

Here, L is the loss function, $\hat{P}_i$ and P_i represent the ground truth version and probabilistic version of a pixel i belonging to a given class, respectively.

The network was built and trained in MATLAB. It was executed on Monash University High Performance Computing (HPC) grid, Massive3. Training time was slightly over 5 h. The machine configuration and training parameters are as follows:

Machine configuration:

- No. of processors – 18
- Memory – 120GB
- GPU – Nvidia Tesla K80

Training parameters:

- Initial learning rate – 0.005
- L2 regularization – 0.0005
- Momentum – 0.9
- Mini batch size – 4
- No. of epochs – 50
- Iterations per epoch – 309

3 Results

3.1 Visual Inspection

As we performed segmentation slice-by-slice, some qualitative results, i.e., visualization of deep learning based semantic segmentations on different MRI slices have been given compared, in addition, with the ground truth segmentation in Fig. 4. In this figure, the green and blue overlays on the MRI slices indicate the ground truth segmentations and the proposed deep learning model based segmentations respectively. Visual representation shows that proposed model was able to perform segmentation almost as accurate as the ground truth version.

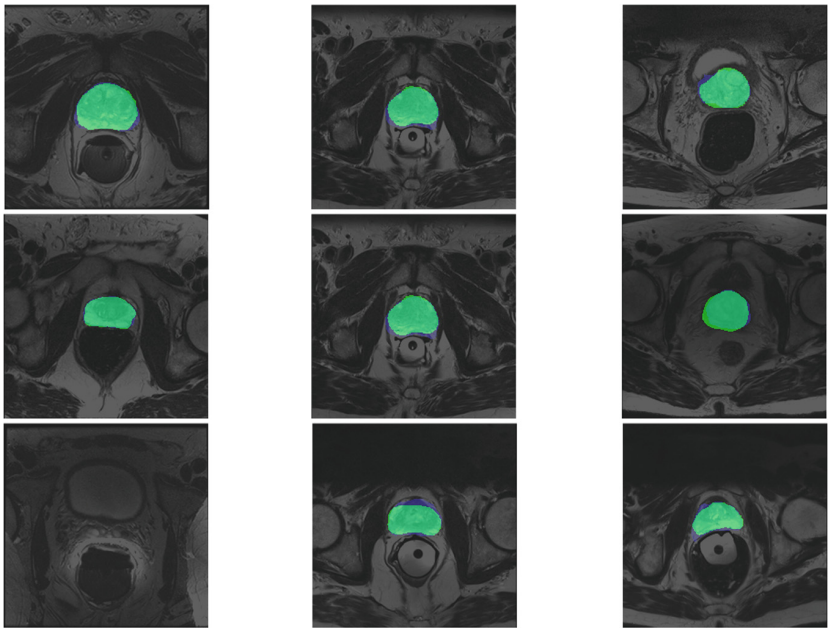

Fig. 4. Prostate segmentation by proposed model (blue) vs. ground truth (green) (Color figure online)

3.2 Quantitative Comparisons

To quantify the accuracy, we follow two common metrics: Dice Similarity Coefficient (DSC) [30] and Intersection-over-Union (IU) as given in Eqs. (3) and (4).

$$DSC = 2\frac{|X \cap Y|}{|X| + |Y|} \tag{3}$$

$$IU = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \tag{4}$$

where X and Y are two different regions, i.e. sets of pixels in the image which is 'prostate region' and 'background' in our case. The modulus sign '| |' defines the cardinal of the corresponding sets.

Table 1 shows the accuracy measures of the proposed method, which signifies achievement of a high accuracy in the proposed segmentation task. Table 2 compares our result with some other studies.

Table 1. Segmentation accuracy using VGG19RSeg.

	DSC (%)	IU (%)
Average	94.57	91.48
Max	99.97	99.94
Min	80.88	80.18

Table 2. Comparison of VGG19RSeg with other studies.

	DSC (%)	Method
Proposed method of this study	94.57% avg., 99.97% max	VGG19RSeg
Tian et al. [19]	85.3% average, 91.5% max	Long's FCN [17]
Ghasab et al. [31]	87% average, 94% max	Active Appearance Model (AAM)
Cho et al. [32]	78% average	CNN + topological derivative
Other method	87.7% average	VGG19-equipped FCN

Table 2 shows that our proposed method has increased accuracy over other similar methods. The next best case is a non-CNN method obtained by Ghasab et al. [31] using AAM. The table also shows that using a FCN having the original VGG19 structure obtains a lower accuracy than the proposed VGG19RSeg method, illustrating the effectiveness of the residual connections in the network over conventional series connections.

A convex hull-based 3D model of the prostate was created for two different patients from MRI slices segmented using the proposed new method. This model is presented alongside a 3D model created from ground truth segmentations and shown in Fig. 5. Visual inspection shows that the using proposed convolutional neural network leads to the creation an almost identical 3D model to that created using human segmentation.

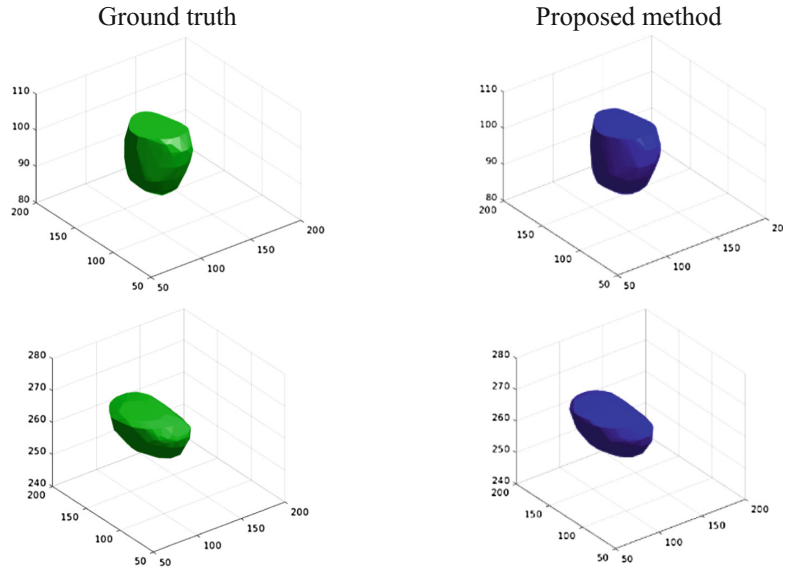

Fig. 5. Ground truth vs. the proposed method for creating a 3D model of the prostate.

4 Conclusion

This study has introduced a highly accurate automatic method for segmenting the prostate in MRI images. Our model has taken a popular deep neural network model, VGG19, which was originally designed for classification, and modified it to build a fully convolutional network with residual connections namely 'VGG19RSeg' for semantic image segmentation. Our results using this proposed deep learning method obtained a mean IU accuracy of 91.48% and a 94.57% DSC accuracy. This is greater than the accuracy of any other comparable method we are aware of. The accuracy of the proposed model is also greater than that of a typical VGG19 based fully convolutional network. Therefore, this study demonstrates the contribution of residual connections in FCN to obtain a greater accuracy in semantic segmentation. Future research will focus on achieving even higher accuracy by using effective pre-processing techniques, as well as adapting this method to other imaging modalities for the purpose of automatically creating and registering 3D models of the patient prostate in real time.

Acknowledgements. Datasets used in this study were part of the PROMISE12 grand challenge for prostate segmentation data sets. The authors wish to thank the Monash University Massive-HPC facility for the provision of high performance computing resources.

References

1. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2016. CA Cancer J. Clin. **66**(1), 7–30 (2016)
2. Prostate Cancer Statistics. https://prostate-cancer.canceraustralia.gov.au/statistics
3. Moore, C.M., et al.: Image-guided prostate biopsy using magnetic resonance imaging-derived targets: a systematic review. Eur. Urol. **63**(1), 125–140 (2013)
4. Khallaghi, S., et al.: Biomechanically constrained surface registration: application to MR-TRUS fusion for prostate interventions. IEEE Trans. Med. Imaging **34**(11), 2404–2414 (2015)
5. Fedorov, A., et al.: Open-source image registration for MRI–TRUS fusion-guided prostate interventions. Int. J. Comput. Assist. Radiol. Surg. **10**(6), 925–934 (2015)
6. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
7. Liao, S., Gao, Y., Oto, A., Shen, D.: Representation learning: a unified deep learning framework for automatic prostate MR segmentation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 254–261. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40763-5_32
8. Toth, R., Madabhushi, A.: Multifeature landmark-free active appearance models: application to prostate MRI segmentation. IEEE Trans. Med. Imaging **31**(8), 1638–1650 (2012)
9. Yan, P., Xu, S., Turkbey, B., Kruecker, J.: Discrete deformable model guided by partial active shape model for TRUS image segmentation. IEEE Trans. Biomed. Eng. **57**(5), 1158–1166 (2010)
10. Abdullah, S., Tischer, P., Wijewickrema, S., Paplinski, A.: Parameter-free hierarchical image segmentation. In: Visual Communications and Image Processing (VCIP), pp. 1–4. IEEE, St. Petersburg (2017)
11. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain Tumor segmentation using convolutional neural networks in MRI images. IEEE Trans. Med. Imaging **35**(5), 1240–1251 (2016)
12. Havaei, M., et al.: Brain tumor segmentation with deep neural networks. Med. Image Anal. **35**, 18–31 (2017)
13. Jia, H., Xia, Y., Song, Y., Cai, W., Fulham, M., Feng, D.D.: Atlas registration and ensemble deep convolutional neural network-based prostate segmentation using magnetic resonance imaging. Neurocomputing **275**, 1358–1369 (2018)
14. Tajbakhsh, N., et al.: Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging **35**(5), 1299–1312 (2016)
15. Zhang, W., et al.: Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. Neuroimage **108**, 214–224 (2015)
16. Milletari, F., et al.: Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. Comput. Vis. Image Underst. **164**, 92–102 (2017)
17. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440. IEEE, Boston (2015)
18. Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1520–1528. IEEE, Santiago (2015)
19. Tian, Z., Liu, L., Fei, B.: Deep convolutional neural network for prostate MR segmentation. In: Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling: International Society for Optics and Photonics, vol. 10135, p. 101351L (2017)

20. Ahmad, E., Goyal, M., McPhee, J.S., Degens, H., Yap, M.H.: Semantic Segmentation of Human Thigh Quadriceps Muscle in Magnetic Resonance Images. arXiv preprint arXiv:1801.00415 (2018)
21. Tran, P.V.: A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI. arXiv preprint arXiv:1604.00494 (2016)
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
23. Canziani, A., Paszke, A., Culurciello, E.: An Analysis of Deep Neural Network Models for Practical Applications. arXiv preprint arXiv:1605.07678 (2016)
24. Norman, B., Pedoia, V., Majumdar, S.: Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. Radiology, 172322 (2018)
25. Milletari, F., Navab, N., Ahmadi, S.-A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE, Stanford (2016)
26. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556 (2014)
27. Yang, S., Ramanan, D.: Multi-Scale Recognition with DAG-CNNs. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1215–1223. IEEE, Santiago (2015)
28. Shuai, B., Zuo, Z., Wang, B., Wang, G.: Dag-recurrent neural networks for scene labeling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
29. MICCAI Grand Challenge: Prostate MR Image Segmentation 2012. https://promise12.grand-challenge.org/
30. Zou, K.H., et al.: Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. Acad. Radiol. **11**(2), 178–189 (2004)
31. Ghasab, M.A.J., Paplinski, A.P., Betts, J.M., Reynolds, H.M., Haworth, A.: Automatic 3D modelling for prostate cancer brachytherapy. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 4452–4456. IEEE, Beijing (2017)
32. Cho, C., Lee, Y.H., Lee, S.: Prostate detection and segmentation based on convolutional neural network and topological derivative. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 4452–4456. IEEE, Beijing (2017)