



# Spatial Probabilistic Distribution Map Based 3D FCN for Visual Pathway Segmentation

Zhiqi Zhao<sup>1</sup>, Danni Ai<sup>1</sup>(✉), Wenjie Li<sup>1</sup>, Jingfan Fan<sup>1</sup>, Hong Song<sup>2</sup>, Yongtian Wang<sup>1,3</sup>, and Jian Yang<sup>1</sup>

<sup>1</sup> School of Optics and Photonics, Beijing Engineering Research Center of Mixed Reality and Advanced Display, Beijing Institute of Technology, Beijing 100081, China  
danni@bit.edu.cn

<sup>2</sup> School of Computer Science and Technology,  
Beijing Institute of Technology, Beijing 100081, China

<sup>3</sup> AICFVE of Beijing Film Academy, 4, Xitucheng Road, Haidian, Beijing 100088, China

**Abstract.** Image-guided surgery has become an important aid in sinus and skull base surgery. In the preoperative planning stage, vital structures, such as the visual pathway, must be segmented to guide the surgeon during surgery. However, owing to the elongated structure and low contrast in medical images, automatic segmentation of the visual pathway is challenging. This study proposed a novel method based on 3D fully convolutional network (FCN) combined with a spatial probabilistic distribution map (SPDM) for visual pathway segmentation in magnetic resonance imaging. Experimental results indicated that compared with the FCN that relied only on image intensity information, the introduction of an SPDM effectively overcame the problem of low contrast and blurry boundary and achieved better segmentation performance.

**Keywords:** Visual pathway segmentation · Optic nerve · 3D fully convolutional network · Spatial probabilistic distribution map

## 1 Introduction

At present, image-guided surgery (IGS) is important in sinus and skull base surgery. The anatomy of the sinus, skull base, and adjacent orbital area is complex and contains important neurological and vascular structures (e.g., visual pathway, internal carotid artery, etc.), and surgeons may inadvertently penetrate these critical structures during surgery and even cause severe complications. Therefore, the successful implementation of IGS requires the accurate delineation of vital structures in preoperative planning. The segmentation of these vital structures is often conducted by manual delineation, which is time-consuming and tedious for the operators. High inter-operator and intra-operator variability may also affect the reproducibility of the treatment plans. Hence, the automatic and precise segmentation of vital structures is highly desired.

The visual pathway is responsible for transmitting visual signals from the retina to the brain. It consists of (1) a paired optic nerve, also known as the cranial nerve II; (2)

the optic chiasm, which is the part of the visual pathway where nearly half the fibers from each optic nerve pair cross to the contralateral side; and (3) an optic tract pair. The optic tract is a continuation of the optic nerve that relays the information from the optic chiasm to the ipsilateral lateral geniculate body. Automatic segmentation of the visual pathway is challenging because of the elongated structure and low contrast in medical images. Some methods have been proposed in this area. These segmentation methods are mainly divided into three categories [1].

1. Atlas-based methods: For atlas-based methods, a transformation is computed between a reference image volume (i.e., atlas), where the structures of interest have been previously delineated, and the target image. Image registration is often used to compute such a transformation. This transformation is then used to project labels in the atlas onto the image volume to be segmented, thus identifying the structures of interest [2]. Multi-atlas-based methods use a database of atlas images and then fuse the projected labels with a specific label fusion strategy to produce the final segmentation result [3]. Atlas-based methods can segment the optic nerve and chiasm [4–6], with Dice similarity coefficients (DSCs) that range from 0.39 to 0.78, which proves that the methods are not robust.
2. Model-based methods: Based on either statistical shape or statistical appearance models, these methods typically produce closed surfaces that can effectively preserve anatomical topologies because the final segmentation results are constrained by the statistical models. Bekes et al. [7] proposed a semi-automated approach using a geometrical model for eyeballs, lenses, and optic nerve segmentation. However, the reproducibility of their approach was found to be less than 50%. Nobel et al. [8] proposed an atlas-navigated optimal medial axis and deformable model algorithm in which a statistical model and image registration are used to incorporate a priori local intensity and tubular shape information. Yang et al. [9] proposed a weighted partitioned active shape model (ASM) for visual pathway segmentation in magnetic resonance imaging (MRI) to improve the shape flexibility and robustness of ASMs to capture the visual pathway's local shape variability. Mansoor et al. [10] proposed a deep learning guided partitioned shape model for anterior visual pathway segmentation. Initially, this method exploited a marginal space deep learning concept-based stacked autoencoder to locate the anterior visual pathway. Then it combined a novel partitioned shape model with an appearance model to guide the anterior visual pathway segmentation.
3. Classification-based methods: Classification-based methods usually train the classifiers by using the features extracted from the neighborhood of each individual voxel. The trained classifiers are then used for voxel-wise tissue classification. The distinctive features and superior performance of the classifiers are integral to these methods. In 2015, Dolz et al. [2] proposed an approach based on support vector machine. The approach extracted features in 2D slices, including image intensity of the voxel and its neighborhood, gradient pixel value, voxel probability, and spatial information. In 2017, Dolz et al. [11] proposed a deep learning classification scheme based on augmented enhanced features to segment organs at risk on the optic region in patients with brain cancer. This method composed novel augmented enhanced

feature vectors that incorporate additional information about a voxel and its environment, such as contextual features, first-order statistics, and spectral measures. It also used stacked denoised auto-encoders as classifiers.

With the development of deep learning, convolutional neural networks (CNNs) in particular have been widely used in medical image segmentation and have achieved superior performance compared with previous methods. Moreover, the fully convolutional networks (FCNs) proposed by Long et al. [12] designed a dense training strategy to train a network on multiple or all voxels of a single volume per optimization step. Compared with classical machine learning algorithms, CNN does not require hand-crafted features for classification. Instead, the network is capable of learning the best features during the training process [1].

However, low tissue contrast or some artifacts in medical images may corrupt the true boundaries of the target tissues and adversely influence the precision of segmentation. Under these circumstances, CNNs cannot effectively extract discriminative features, leading to poor segmentation results. Unfortunately, this problem also exists in the visual pathway segmentation in MRI. Given the location and shape similarity of the target tissue between individuals, image intensity information can be combined with shape and position prior information to learn the neural network. On the basis of this hypothesis, we proposed a novel method based on 3D FCN combined with spatial probabilistic distribution map (SPDM) for visual pathway segmentation in MRI. SPDM reflects the probability that a voxel belongs to a given tissue, and it is obtained by summing all the manual labels contained in the training data set. The experimental results show that compared with the FCN relying only on image intensity information, the introduction of SPDM effectively overcomes the problem of low contrast and blurry boundary and achieves better segmentation performance. To the best of our knowledge, this deep learning method is the first to incorporate an SPDM into an FCN for visual pathway segmentation.

## 2 Method

Given the size of visual pathway accounts for a small proportion of the entire brain MRI, the task was divided into two stages, location of the region of interest (ROI) and target segmentation, to save computing resources and speed up the calculation. Figure 1 illustrates the entire scheme of the proposed method.

### 2.1 The Creation of SPDM and Extraction of ROI

SPDM was obtained by summing all the manual labels contained in the training data set. Specifically, each image in the training set was first registered to a selected reference image. Registration was implemented using the well-known and publicly available tool elastix [13]. Then, the transformation obtained by registration was used to project the attached label to the reference space. Finally, the registered binary labels were summed up to create an SPDM (Fig. 2). SPDM reflected the probability that a voxel belonged to a visual pathway, which was used to integrate shape and locality information into the

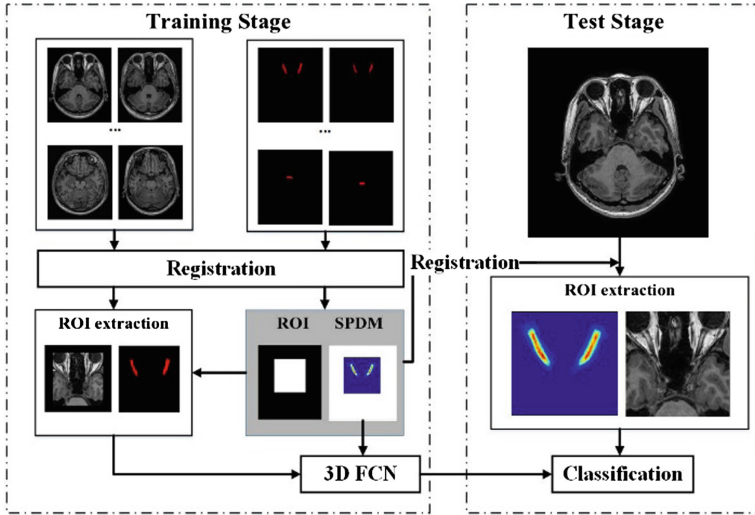


Fig. 1. The entire scheme of the proposed method

3D FCN training process, as described in the next section. SPDM also played another important role, which was to achieve the extraction of ROI. The starting and ending slice positions in the axial, coronal, and sagittal planes were calculated and expanded outward to form a bounding box  $90 \times 54 \times 90$  in size. The morphological dilation operation was performed on the registered label to fill up the holes caused by registration. To eliminate the influence of very different images, we experimentally observed that the value of SPDM less than  $2/n$  was set to 0, where  $n$  is the number of images in the training set.

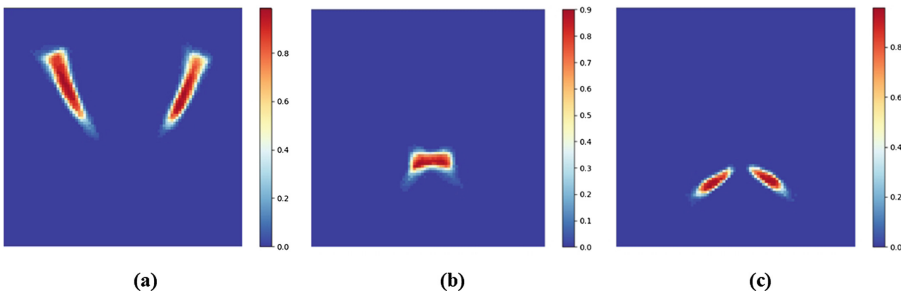
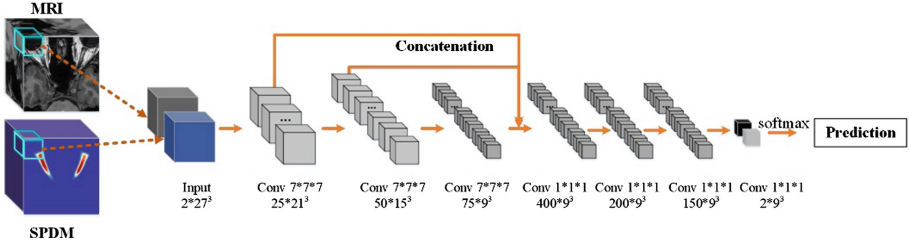


Fig. 2. The 2D representation of SPDM, which reflected the probability that a voxel belonged to a visual pathway. (a) optic nerves; (b) optic chiasm; (c) optic tracts.

## 2.2 3D FCN Architecture

The 3D FCN architecture used in this study was derived from Kamnitsas et al. [14] and Dolz et al. [15]. The baseline network architecture is shown in Fig. 3.



**Fig. 3.** The baseline 3D FCN architecture

The baseline network was composed of three convolutional layers with kernels of size  $7 \times 7 \times 7$ . Many studies have shown the benefits of using deeper network architectures. Following Simonyan et al. [16], the final network used three consecutive  $3 \times 3 \times 3$  convolution kernels instead of the  $7 \times 7 \times 7$  convolution kernel. Both cases have the same receptive field sizes. By using these smaller kernels, we obtained a deeper architecture that could learn a more discriminative hierarchy of features, with a reduced risk of overfitting. In addition, there were fewer parameters when smaller convolution kernels were used. Compared with CNN, FCN replaced the fully connected layer with  $1 \times 1 \times 1$  convolution kernels, which allowed networks to be applied to images of arbitrary size. The dense training strategy enabled FCN to obtain predictions for multiple or all voxels in an image per step, thus avoiding redundant convolution and making the network more efficient. Each convolutional layer was followed by a PRelu nonlinear activation layer, and the final convolutional layer was followed by the Softmax function.

Furthermore, to use different levels of information, the network concatenated the feature maps of convolution layers at different stages (Fig. 3).

### 2.3 Experimental Details

After the SPDM was obtained, it was respectively registered to each image so that the SPDM and the image had one-to-one correspondence. At the same time, we obtained the ROI, including visual pathway, which removed most of the irrelevant information. Following the network of Dolz et al. [15], we obtained a network with the input size of  $27 \times 27 \times 27$  and the network's output was the prediction of the input's central  $9 \times 9 \times 9$  size patch. Thus, the volume data were cropped into  $27 \times 27 \times 27$  size patches that corresponded to the central  $9 \times 9 \times 9$  size labels. Image intensity information and SPDM were simultaneously fed into the neural network as two channels.

During the training stage, to alleviate the problem of class imbalance, we chose the target patches (patches that contain target voxels) and added the same number of background patches (all voxels belong to the background). However, owing to the intensity similarity between a few background patches and target patches, some voxels in the background patches may have been misclassified as targets during the testing stage because not all background patches were sent to the network for training. The introduction of SPDM effectively solved this problem. For background patches, the value of SPDM was 0, meaning that the target did not appear in this spatial location. During

the test stage, the output predictions were stitched to reconstruct the final segmentation results for each test image.

The proposed method has been implemented in Python, using elastix [13] for registration and Keras for network building. The initial learning rate was set as 0.001, which was divided by a factor of 10 every three epochs when the validation loss stopped improving. The weights of the network were initialized by Xavier initialization [17] and optimized by Adam algorithm [18]. The loss function used by the network was cross-entropy. To prevent over-fitting, early stopping strategy was likewise utilized in the work if no improvement arose in the validation loss after 10 epochs.

### 3 Results

MR 3T T1 brain images with voxel sizes of  $1.0 \times 1.0 \times 1.0 \text{ mm}^3$  were used in this experiment. A total of 93 images were provided by Beijing Tongren Hospital, 23 of which were used for testing. For each item of data, the ground truth for the visual pathway was manually delineated by radiologists. Intensity normalization was performed on the original image.

DSC was used as the quantitative evaluation metric, which reflects the overlap rate between the segmentation result and the ground truth. DSC is defined as follows:

$$\text{DSC}(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

where  $A$  and  $B$  represent the automatic segmentation result and the ground truth, respectively.

For the spatial distance-based metric, average symmetric surface distance (ASD) was used as another metric. ASD is defined as follows:

$$\text{ASD}(A, B) = \frac{\sum_{a \in S_A} d(a, S_B) + \sum_{b \in S_B} d(b, S_A)}{|S_A| + |S_B|} \quad (2)$$

$$d(a, S_B) = \min_{b \in S_B} (a - b) \quad (3)$$

where  $S_A$  and  $S_B$  represent the surface of the automatic segmentation result and the ground truth, respectively;  $d(a, S_B)$  represents the shortest distance of an arbitrary point  $a$  on  $S_A$  to  $S_B$ ; and  $\|\cdot\|$  denotes the Euclidean distance.

Table 1 shows the quantitative analysis of the segmentation results under different methods, expressed as mean  $\pm$  standard deviation. The text in bold indicates the best metrics. ‘‘FCN(MRI)’’ indicates that the network only learns image intensity information, and ‘‘FCN(SPDM)’’ indicates that only spatial probabilistic distribution information is learned. ‘‘Cascaded FCN’’ is a cascaded training strategy. On the basis of the previous FCN that relied only on MRI, the training set is sent to the network for prediction, in which the misclassified patches continue to be trained to fine-tune the network. Thus, the network can learn patches that are not easily discernible. ‘‘Our method’’ represents a combination of image intensity information and spatial probabilistic distribution information, which is the method proposed in this work.

**Table 1.** Quantitative analysis results of different methods

Methods	FCN(MRI)	Cascaded FCN	FCN(SPDM)	Our method
DSC(%)	$82.92 \pm 2.87$	$83.47 \pm 3.62$	$68.58 \pm 5.20$	<b><math>84.89 \pm 1.40</math></b>
ASD(mm)	$0.94 \pm 0.35$	$0.42 \pm 0.08$	$0.68 \pm 0.10$	<b><math>0.37 \pm 0.04</math></b>

Figure 4 shows the box plot of the DSC comparison results. The results of quantitative analysis indicate that FCN combined with SPDM achieves better performance and is more robust. Compared with the FCN that relied only on image intensity information, the introduction of spatial probability information in this paper has increased the DSC from  $(82.92 \pm 2.87) \%$  to  $(84.89 \pm 1.40) \%$ .

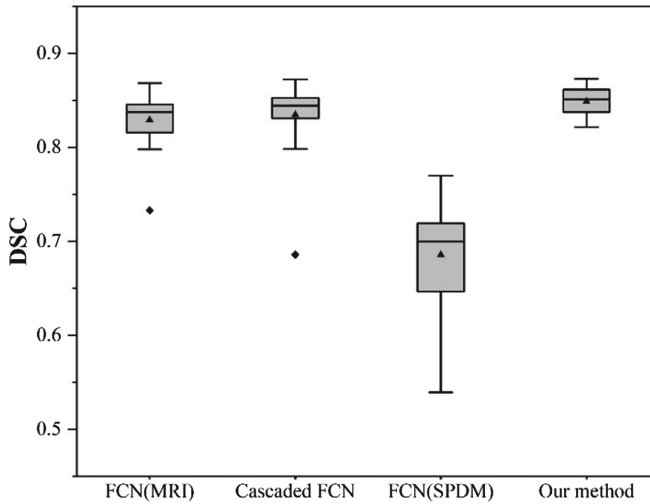
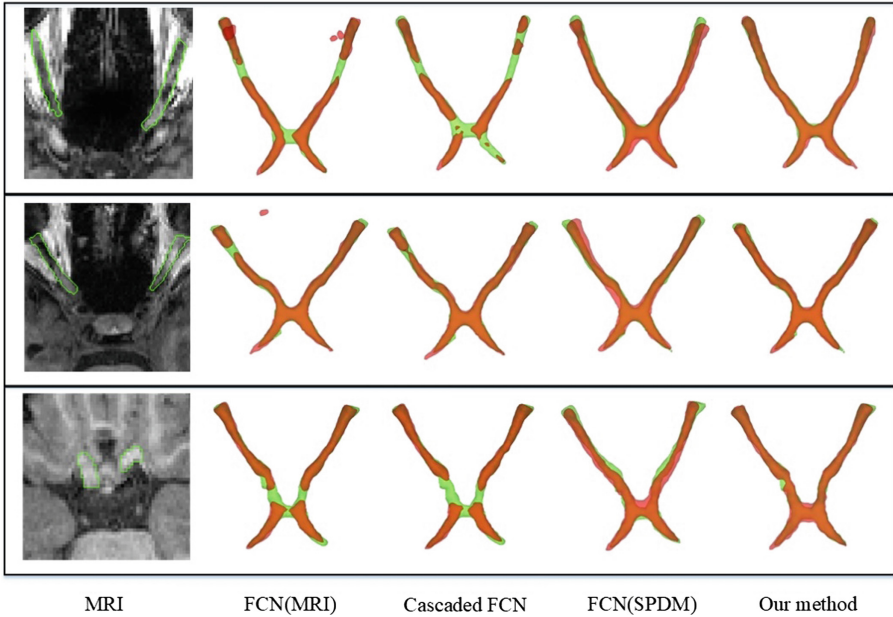
**Fig. 4.** The box plot of DSC comparison results

Figure 5 demonstrates the visual comparison of segmentation results for several cases of data. Each row represents the segmentation results of an example of MRI under different methods, in which the ground truth is represented by green and the automatic segmentation result is represented by red. The green curves in MRI outline the blurring boundary of the target. The segmentation results obtained by FCN that relied only on image intensity information have noise and breakage. Although the cascaded training strategy can suppress noise, it cannot identify the blurring boundary of the target; worse outcomes are possible as well. SPDM provides a complete shape of the visual pathway but deviates from the ground truth because of a few variations between individuals. To a certain extent, learning only with SPDM can be a type of multi-atlas-based segmentation, in which a network is used to determine the label fusion strategy. The obtained DSC also confirmed that it is similar to multi-atlas-based methods. Owing to

the integration of shape and locality information, FCN combined with SPDM effectively solved the problem of background misclassification and discontinuity of segmentation results caused by low contrast.



**Fig. 5.** Visual comparison of segmentation results. Each row represents the segmentation results of an example of MRI under different methods, in which the ground truth is represented by green and the automatic segmentation result is represented by red. The green curves in MRI outline the blurring boundary of the target. (Color figure online)

## 4 Conclusion

This study proposed a novel method based on 3D FCN combined with SPDM for visual pathway segmentation in MRI. SPDM reflects the probability that a voxel belongs to a given tissue and preserves the shape and locality information of the target. Experimental results show that compared with the FCN relying only on image intensity information, the introduction of SPDM effectively overcame the problem of background misclassification and discontinuity of segmentation results caused by low contrast. FCN combined with SPDM achieves improved segmentation performance, in which DSC is improved from  $(82.92 \pm 2.87)\%$  to  $(84.89 \pm 1.40)\%$ .

Considering the similarity in the anatomy and location of tissues in medical images, useful shape and locality priori information may be provided for tissue segmentation. CNNs can automatically extract features from an image but may fail if the image is blurred. The experimental results in this work demonstrated that the combination of traditional shape and location information and FCN is promising. In the future, we



will try to apply the approach to other tissue segmentation tasks that encounter similar problems.

**Acknowledgments.** This work was supported by the National Key Research and Development Program of China (2017YFC0112000), National Science and Technology Major Project of China (2018ZX10723-204-008), and the National Science Foundation Program of China (61672099, 61527827, 61771056).

## References

1. Ren, X., et al.: Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images. *Med. Phys.* **45**(5), 2063–2075 (2018)
2. Dolz, J., Leroy, H.A., Reyens, N., Massotier, L., Vermandel, M.: A fast and fully automated approach to segment optic nerves on MRI and its application to radiosurgery. In: *IEEE International Symposium on Biomedical Imaging*, pp. 1102–1105. IEEE, New York (2015)
3. Iglesias, J.E., Sabuncu, M.R.: Multi-atlas segmentation of biomedical images: a survey. *Med. Image Anal.* **24**(1), 205–219 (2015)
4. Gensheimer, M., Cmelak, A., Niermann, K., Dawant, B.M.: Automatic delineation of the optic nerves and chiasm on CT images. *Medical Imaging*, p. 10. SPIE, San Diego (2007)
5. Asman, A.J., DeLisi, M.P., Mawn, L.A., Galloway, R.L., Landman, B.A.: Robust non-local multi-atlas segmentation of the optic nerve. *Medical Imaging 2013: Image Processing*, pp. 86691L. International Society for Optics and Photonics (2013)
6. Harrigan, R.L., et al.: Robust optic nerve segmentation on clinically acquired computed tomography. *J. Med. Imaging* **1**(3), 034006 (2014)
7. Bekes, G., Mate, E., Nyul, L.G., Kuba, A., Fidrich, M.: Geometrical model-based segmentation of the organs of sight on CT images. *Med. Phys.* **35**(2), 735–743 (2008)
8. Noble, J.H., Dawant, B.M.: An atlas-navigated optimal medial axis and deformable model algorithm (NOMAD) for the segmentation of the optic nerves and chiasm in MR and CT images. *Med. Image Anal.* **15**(6), 877–884 (2011)
9. Yang, X., et al.: Weighted partitioned active shape model for optic pathway segmentation in MRI. In: Linguraru, M.G., et al. (eds.) *CLIP 2014*. LNCS, vol. 8680, pp. 109–117. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-13909-8\\_14](https://doi.org/10.1007/978-3-319-13909-8_14)
10. Mansoor, A., et al.: Deep learning guided partitioned shape model for anterior visual pathway segmentation. *IEEE Trans. Med. Imaging* **35**(8), 1856–1865 (2016)
11. Dolz, J., et al.: A deep learning classification scheme based on augmented-enhanced features to segment organs at risk on the optic region in brain cancer patients (2017)
12. 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)
13. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* **29**(1), 196–205 (2010)
14. Kamnitsas, K., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017)
15. Dolz, J., Desrosiers, C., Ben Ayed, I.: 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study. *Neuroimage* **170**, 456–470 (2018)
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *Comput. Sci.* (2014)

17. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256 (2010)
18. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)