



A Skip-Connected 3D DenseNet Networks with Adversarial Training for Volumetric Segmentation

Toan Duc Bui, Sang-il Ahn, Yongwoo Lee, and Jitae Shin (✉)

Department of Electrical and Computer Engineering, Sungkyunkwan University,
Suwon 16419, Republic of Korea
jtshin@skku.edu

Abstract. In this paper, we propose a novel end-to-end adversarial training on volumetric brain segmentation architecture that allows to enforce long-range spatial label contiguity and label consistency. The proposed network consists of two networks: generator and discriminator. The generator network allows to take volumetric image as input and provides a volumetric probability map for each tissue. Then, the discriminator network learns to differentiate ground-truth maps from the probability maps of generator network. We design a discriminator in a fully convolutional manner to differentiate the predicted probability maps from the ground-truth segmentation distribution with the consideration of the spatial information on voxel level, which makes it difficult to learn the discriminator. In order to overcome it, the proposed discriminator provides a 3D confidence map which indicates corresponding regions of the probability maps close to the ground-truth. Based on the 3D confidence map information, the generator network will refine prediction output close to the ground-truth maps in a high-order structure.

Keywords: Adversarial training · Brain segmentation · Deep convolutional networks

1 Introduction

Automatic brain segmentation plays an importance rule for disease diagnosis, progression, and treatment monitoring [11]. With a success of deep learning [9], many algorithms have been proposed to obtain an accurate segmentation result [2, 12, 14]. For example, Moeskops et al. [12] introduced an automatic segmentation method using multiple patch 2D convolutional networks (CNN). Çiçek, Özgün et al. proposed 3D UNet using skip connected path to concatenate low and high level features to produce a full-resolution segmentation. Bui et al. [1] proposed a fully convolutional 3D DenseNet by concatenate multiple contextual information from different level features. Although these methods demonstrated good performance, the contour obtained from thresholding based on the score of

volumetric infant brain sometimes can be imprecise with small isolated regions or holes in the predictions.

Recently, Goodfellow et al. [4] proposed generative adversarial network (GAN) for generative image modeling. It consists of two networks, discriminator and generator. While a discriminator tries to distinguish ground-truth images from the outputs generated by the generator, the generator tries to generate as realistic outputs as the discriminator cannot differentiate from the ground-truth image. Inspired of GAN, Luc et al. [10] introduce an adversarial training approach for training semantic segmentation models. It shows that the adversarial training leads to improve segmentation accuracy.

In this paper, we extend the 3D DenseNet networks [1] by adding an end-to-end adversarial training for volumetric segmentation. While existing methods such as [6, 10, 13] which have both generator and discriminator designed for 2D segmentation problem, we focus on 3D segmentation problem. In order to achieve it, first we introduce a generator network that uses 3D features to explore the information of adjacent slices to enhance volumetric segmentation. The proposed generator network allows to take volumetric image as input and provides a volumetric probability map for each tissue. Then, the discriminator network learns to differentiate ground-truth maps from probability maps of generator network. Inspired of fully discriminator network [6], we extend it from 2D to 3D discriminator network by using $1 \times 1 \times 1$ convolution at final layer. We use skip connection that allows to capture multiple contexture from discriminator network. The generator and discriminator networks are trained in end-to-end manner in order to jointly optimize all the weights in the network using an efficient weight update technique.

2 Method

Figure 1 illustrates the overview of our proposal for volumetric infant brain segmentation. It consists of two networks: generator and discriminator network. The generator network uses 3D features to explore the information of adjacent slices to enhance volumetric segmentation. The proposed generator network allows to take volumetric images as inputs and provides a volumetric probability maps for each tissue. The probability maps through the discriminator network that uses a fully convolutional scheme to obtain spatial 3D confidence. The 3D confidence map indicates which regions of the probability maps are close to the ground truth. Based on the 3D confidence map information, the generator network will refine prediction output close to the ground truth maps in a high-order structure. The generator and discriminator networks are trained in end-to-end manner in order to jointly optimize all the weights in the network using an efficient weight update technique.

For generator network, we modify the network in [1] by adding a squeeze-and-excitation (SE) block [5] after each dense block. In this way, it allows to explore the interdependencies between the channels. Figure 1 illustrates our generator networks for brain MRI segmentation. It consists of contains 47 layers with

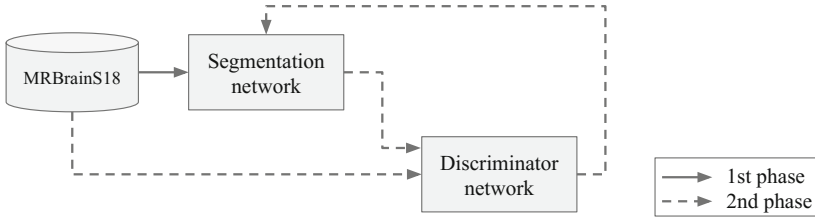


Fig. 1. Our proposed flow chart for brain MRI segmentation

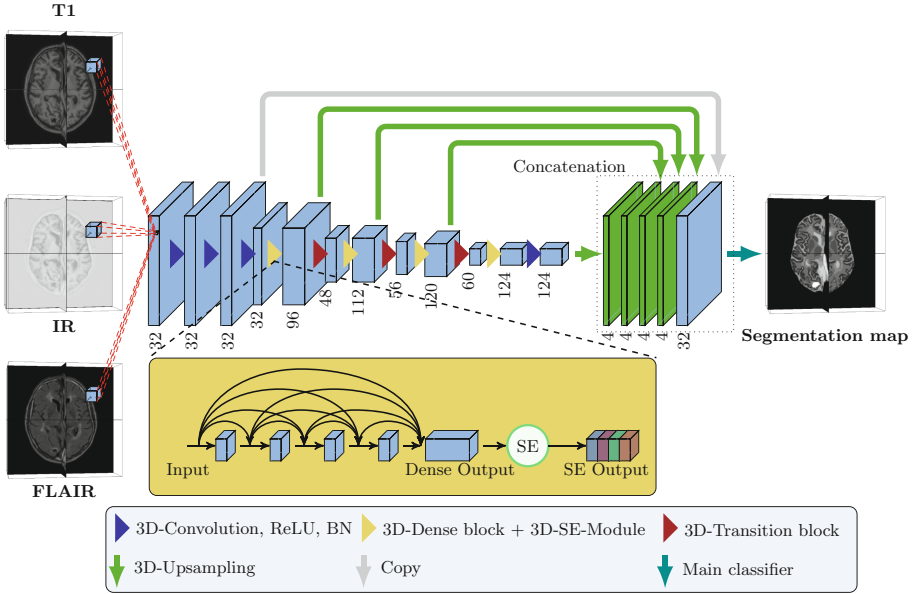


Fig. 2. Modified 3D-SkipDenseSeg network architecture for brain MRI segmentation.

downsampling and upsampling paths. The downsampling path aims to reduce feature resolution and to increase the receptive field. It is performed by four dense blocks with growth rate of $k = 16$. We used four $3 \times 3 \times 3$ convolutions in each dense block. After each dense block, we add an SE block to generate feature interdependencies between the channels. Then the output features from SE block are fed to transition layers that use $1 \times 1 \times 1$ convolution to reduce features size. Meanwhile, the upsampling paths tries to recover original resolution from the low feature resolutions. We upsample low-level features map directed to original resolution and concatenate them together through skip connection before feeding to a classifier. By this way, we can reduce the number of learning parameters. The network can take multi-modalities images, such as T1, IR and FLAIR, as input to generates segmented image as shown in Fig. 2.

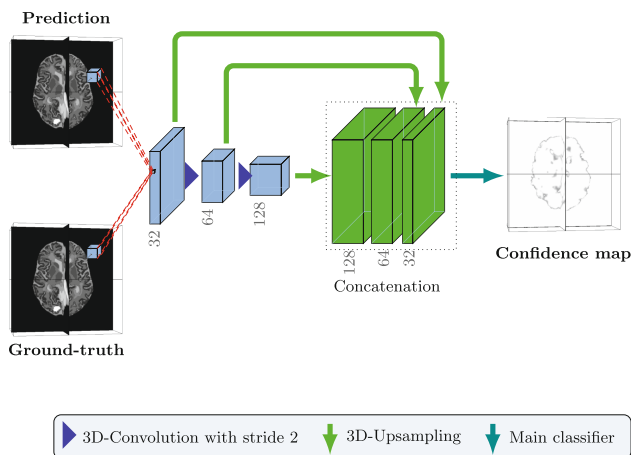


Fig. 3. Discriminator network for brain MRI segmentation. The brighter regions indicate that they are close to the ground truth distribution

Figure 3 shows the structure of discriminator network. The outputs of generator network are fed to the discriminator network that has three $4 \times 4 \times 4$ convolution with stride of 2 to reduce features map. We upsample the result of each convolution into original size and concatenate these feature together to capture multiple contexture information from discrimiatnor network. $1 \times 1 \times 1$ convolution is used to classify the concatenated feature into two classes: real or fake. The discriminator in a fully convolutional manner to differentiate the predicted probability maps from the ground-truth segmentation distribution with the consideration of the spatial information on voxel level, which increases the difficulty to learn the discriminator. The proposed discriminator provides a 3D confidence map which indicates which regions of the probability maps are close to the ground truth. Based on the 3D confidence map information, the generator network will refine prediction output close to the ground truth maps in a high-order structure.

3 Experimental Results

We used the public dataset from the MRBrainS18 Challenge to evaluate the performance of the proposed method. The dataset consists of 7 subjects for training and 23 subjects for testing. For each subject, three modalities are available that includes T1-weighted, T1-weighted inversion recovery and T2-FLAIR with voxel size of $0.958 \text{ mm} \times 0.958 \text{ mm} \times 3.0 \text{ mm}$. Each subject was manually segmented into 11 classes by the challenge organizers. For each participant team, the 7 subjects for training with ground-truth was provided for fine-tuning the network, while 23 testing subjects were retained by the organizers for a fair comparison. The aim of this challenge is to automatically segment the images from each subject into 9 classes and compared with the manual labels using Dice coefficient

(DSC) [3], modified 95th percentile Hausdorff distance (HD) [7] and volumetric similarity (VS).

We implemented and trained the proposed network with an NVIDIA Titan X and Pytorch framework. We first normalized the three modalities input to zero mean and unit variance before feeding it to the network. Because of limited memory resource, we randomly cropped sub-volume samples with a size of $48 \times 64 \times 64$ for input. Both generator and discriminator network were trained with an Adam optimizer [8] with a mini-batch size of four. The learning rate was initially set to 0.0005 for generator network and 0.000002 for discriminator network. The weights then were decreased by a factor of 0.1 every 4000 epochs. We used a weight decay of 0.0001 for generator network. In inference phase, we used the majority voting strategy that smoothed the predictions of the overlapping sub-volumes of size $48 \times 64 \times 64$ with a stride of $1 \times 8 \times 8$, which resulted in an improved result. We use 6 subjects from MRBrainS18 challenge for training, 1 subjects for validation. To make stable training on discriminator network, we first only train the generator network in 500 epochs to provide a reasonable result from the generator network. After that, The generator and discriminator networks are trained in end-to-end manner in order to jointly optimize all the weights in the network. It spends about 12 h for training and 6 min to segment each subject on Titan X Pascal.

Table 1 shows performance results on validation set using with and without adversarial training. The adversarial training not only achieves a better result on DSC but also on other metrics such as HD and VS. From the result, we can conclude that the adversarial training leads improve the segmentation accuracy.

Table 1. Performance comparison of proposed method with or without adversarial training on validation data (DSC: %, HD: mm, VS: mm).

Methods	DSC	HD	VS
Without adversarial training	86.87	3.1232	0.9555
With adversarial training	87.16	2.9845	0.9645

Figure 4 shows the comparison results between with or without adversarial training of a 2D slice on the validation set. From the figure, we can observe that the adversarial training yields a better performance than the case of without adversarial training.

Figure 5 shows our performance of 23 subject on the testing set. We achieve rank of 14 over 22 participant teams. The reason is that the adversarial training is difficult for training small label such as white matter lesions.

3.1 Conclusion

In this paper, we proposed an adversarial training on 3D segmentation task. We extend the discriminator network in a fully convolutional manner to differentiate

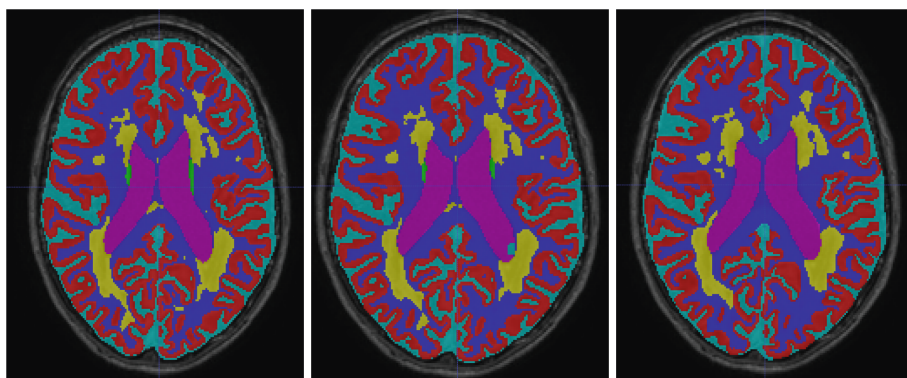


Fig. 4. An comparison of segmentation result on validation set: (left) without adversarial training, (middle) with adversarial training and (right) ground-truth image

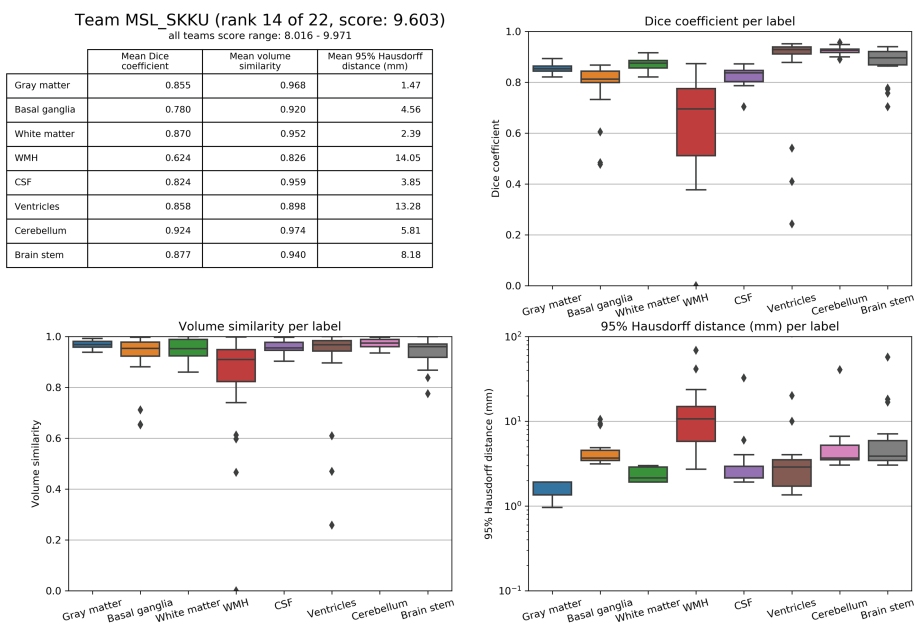


Fig. 5. Segmentation result of 23 subject on testing set

the predicted probability maps from the ground truth segmentation distribution with the consideration of the spatial information on voxel level, which makes it difficult to learn the discriminator. The experiment results show that the adversarial training leads to improve the segmentation accuracy not only on 2D segmentation task, but also on 3D segmentation task such as brain MRI segmentation.

Acknowledgment. This research was supported partly by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2018R1C1B6007472). This research was supported partly by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-2018-0-01798) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. Bui, T.D., Shin, J., Moon, T.: 3D densely convolution networks for volumetric segmentation. arXiv preprint [arXiv:1709.03199](https://arxiv.org/abs/1709.03199) (2017)
2. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
3. Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* **26**(3), 297–302 (1945)
4. Goodfellow, I., et al.: Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, pp. 2672–2680 (2014)
5. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. arXiv preprint [arXiv:1709.01507](https://arxiv.org/abs/1709.01507), vol. 7 (2017)
6. Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.Y., Yang, M.H.: Adversarial learning for semi-supervised semantic segmentation. arXiv preprint [arXiv:1802.07934](https://arxiv.org/abs/1802.07934) (2018)
7. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(9), 850–863 (1993)
8. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
9. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436 (2015)
10. Luc, P., Couprie, C., Chintala, S., Verbeek, J.: Semantic segmentation using adversarial networks. arXiv preprint [arXiv:1611.08408](https://arxiv.org/abs/1611.08408) (2016)
11. Mendrik, A.M., et al.: Mrbrains challenge: online evaluation framework for brain image segmentation in 3T MRI scans. *Comput. Intell. Neurosci.* **2015**, 1 (2015)
12. Moeskops, P., Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J., Išgum, I.: Automatic segmentation of MR brain images with a convolutional neural network. *IEEE Trans. Med. Imaging* **35**(5), 1252–1261 (2016)
13. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint [arXiv:1511.06434](https://arxiv.org/abs/1511.06434) (2015)
14. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28