One-Shot Traumatic Brain Segmentation with Adversarial Training and Uncertainty Rectification

Xiangyu Zhao[1], Zhenrong Shen[1], Dongdong Chen[1], Sheng Wang[1,2], Zixu Zhuang[1,2], Qian Wang[3], and Lichi Zhang[1(✉)]

[1] Shanghai Jiao Tong University, Shanghai, China
lichizhang@sjtu.edu.cn
[2] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China
[3] ShanghaiTech University, Shanghai, China

Abstract. Brain segmentation of patients with severe traumatic brain injuries (sTBI) is essential for clinical treatment, but fully-supervised segmentation is limited by the lack of annotated data. One-shot segmentation based on learned transformations (OSSLT) has emerged as a powerful tool to overcome the limitations of insufficient training samples, which involves learning spatial and appearance transformations to perform data augmentation, and learning segmentation with augmented images. However, current practices face challenges in the limited diversity of augmented samples and the potential label error introduced by learned transformations. In this paper, we propose a novel one-shot traumatic brain segmentation method that surpasses these limitations by adversarial training and uncertainty rectification. The proposed method challenges the segmentation by adversarial disturbance of augmented samples to improve both the diversity of augmented data and the robustness of segmentation. Furthermore, potential label error introduced by learned transformations is rectified according to the uncertainty in segmentation. We validate the proposed method by the one-shot segmentation of consciousness-related brain regions in traumatic brain MR scans. Experimental results demonstrate that our proposed method has surpassed state-of-the-art alternatives. Code is available at https://github.com/hsiangyuzhao/TBIOneShot.

Keywords: One-Shot Segmentation · Adversarial Training · Traumatic Brain Injury · Uncertainty Rectification

1 Introduction

Automatic brain ROI segmentation for magnetic resonance images (MRI) of severe traumatic brain injuries (sTBI) patients is crucial in brain damage assessment and brain network analysis [8,11], since manual labeling is time-consuming

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43901-8_12.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14223, pp. 120–129, 2023.
https://doi.org/10.1007/978-3-031-43901-8_12

and labor-intensive. However, conventional brain segmentation pipelines, such as FSL [14] and FreeSurfer [4], suffer significant performance deteriorations due to skull deformation and lesion erosions in traumatic brains. Although automatic segmentation based on deep learning has shown promises in accurate segmentation [10,12], these methods are still constrained by the scarcity of annotated sTBI scans. Thus, researches on traumatic brain segmentation under insufficient annotations needs further exploration.

Recently, one-shot medical image segmentation based on learned transformations (OSSLT) has shown great potential [3,17] to deal with label scarcity. These methods typically utilize deformable image registration to learn spatial and appearance transformations and perform data augmentation on the single labeled image to train the segmentation, which is shown in Fig. 1(a). Given a labeled image as the atlas, two unlabeled images are provided as spatial and appearance references. Appearance transform and spatial transform learned by deformable registration are applied to the atlas image to generate a pseudo-labeled image to train the segmentation, and the label warped by spatial transform serves as the ground-truth. In this way, the data diversity is ensured by a large amount of unlabeled data, and the segmentation is learned by abundant pseudo images.

However, despite the previous success, the generalization ability of these methods is challenged by two issues in traumatic brain segmentation: 1) Limited diversity of generated data due to the amount of available unlabeled images. Although several studies [6,7] have proposed transformation sampling to introduce extra diversity for alleviating this issue, their strategies rely on a manual-designed distribution, which is not learnable and limits the capacity of data augmentation. 2) The assumption that appearance transforms in atlas augmentation do not affect semantic labels in the images [6,17]. However, this assumption neglects the presence of abnormalities in traumatic brains, such as brain edema, herniation, and erosions, which affect the appearance of brain tissues and introduce label errors.

To address the aforementioned issues, we propose a novel one-shot traumatic brain segmentation method that leverages adversarial training and uncertainty rectification. We introduce an adversarial training strategy that improves both the diversity of generated data and the robustness of segmentation, and incorporate an uncertainty rectification strategy that mitigates potential label errors in generated samples. We also quantify the segmentation difference of the same image with and without the appearance transform, which is used to estimate the uncertainty of segmentation and rectify the segmentation results accordingly. The main contributions of our method are summarized as follows: First, we develop an adversarial training strategy to enhance the capacity of data augmentation, which brings better data diversity and segmentation robustness. Second, we notice the potential label error introduced by appearance transform in current one-shot segmentation attempts, and introduce uncertainty rectification for compensation. Finally, we evaluate the proposed method on brain segmentation of sTBI patients, where our method outperforms current state-of-the-art methods.

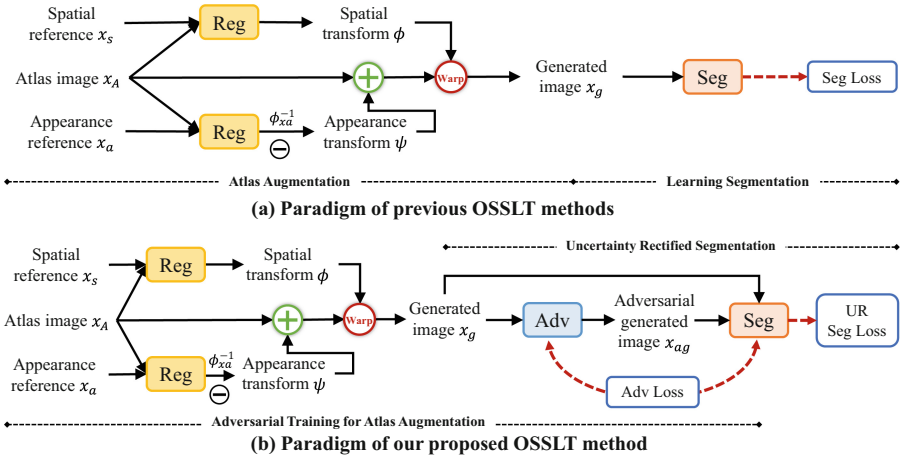

Fig. 1. An illustration of one-shot segmentation based on learned transformations (OSSLT). Compared with previous methods, we aim to use adversarial training and uncertainty rectification to address the current challenges in OSSLT.

2 Method

2.1 Overview of One-Shot Medical Image Segmentation

After training the unsupervised deformable registration, one-shot medical image segmentation based on learned transformations typically consists of two steps: 1) Data augmentation on the atlas image by learned transformations; 2) Training the segmentation using the augmented images. The basic workflow is shown in Fig. 1(a). Specifically, given a labeled image x_A as the atlas and its semantic labels y_A, two reference images, including the spatial reference x_s and appearance reference x_a, are provided to augment the atlas by spatial transform ϕ and appearance transform ψ that are calculated by the same pretrained registration network.

For spatial transform, given an atlas image x_A and a spatial reference x_s, the registration network performs the deformable registration between them and predicts a deformation field ϕ, which is used as the spatial transform to augment the atlas image spatially. For appearance transform, given an atlas image x_A and an appearance reference x_a, we warp x_a to x_A via an inverse registration ϕ_{xa}^{-1} and generates a inverse-warped $\tilde{x_a} = x_a \circ \phi_{xa}^{-1}$, and appearance transform $\psi = \tilde{x_a} - x_A$ is calculated by the residual of inverse-warped appearance reference $\tilde{x_a}$ and the atlas image x_A. It should be noted that the registration here is diffeomorphic to allow for inverse registration.

After acquiring both the spatial and appearance transform, the augmented atlas $x_g = (x_A + \psi) \circ \phi$ is generated by applying both transformations. The corresponding ground-truth $y_g = y_A \circ \phi$ is the atlas label warped by ϕ, as it is hypothesized that appearance transform does not alter the semantic labels in

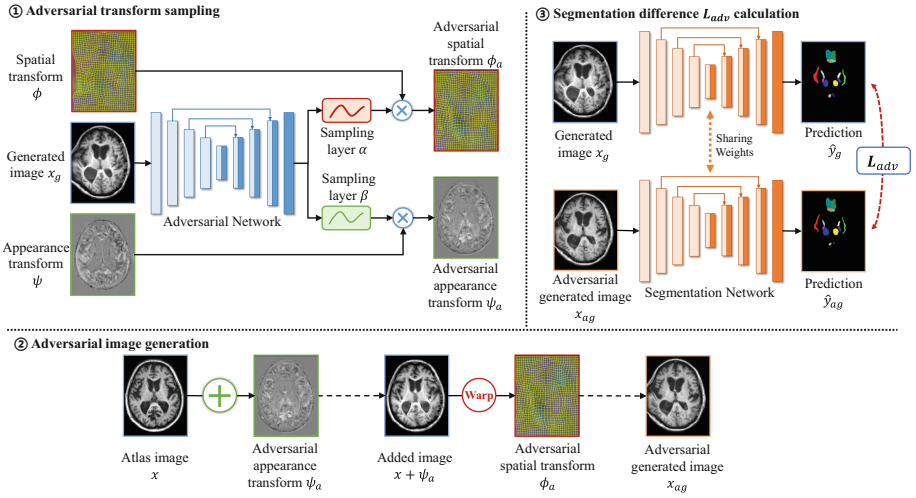

Fig. 2. Framework of adversarial training, which includes adversarial transform sampling, adversarial image generation and segmentation difference calculation.

the atlas image. During segmentation training, a large amount of spatial and appearance reference images are sampled to ensure the diversity of x_g, and the segmentation is trained with generated pairs of x_g and y_g.

In this work, we focus on both of the two steps in OSSLT by adversarial training and uncertainty rectification, which is shown in Fig. 1(b). Specifically, we generate an adversarial image x_{ag} along with x_g by adversarial training to learn better data augmentation for the atlas image, and uncertainty rectification is utilized during segmentation learning to bypass the potential label errors introduced by appearance transforms. We discuss our proposed adversarial training and uncertainty rectification in Sect. 2.2 and Sect. 2.3, respectively.

2.2 Adversarial Training

Although the diversity of generated pairs of x_g and y_g is ensured by the increased number of unlabeled images as references, such a setting requires a large amount of unlabeled data. Inspired by [2,9], we adopt the adversarial training strategy to increase both the data diversity and the segmentation robustness, which is shown in Fig. 2. Given a learned spatial transform ϕ, appearance transform ψ, and a generated image x_g augmented by ϕ and ψ, our adversarial training is decomposed into the following 3 steps:

First, we feed x_g into the adversarial network and generate two sampling layers α and β activated by Sigmoid function. The sampling layers α and β have the same spatial shape with ϕ and ψ respectively, and each location in the sampling layers represents the sampling amplitude of the original transform, ranging from 0 to 1. In this way, the diversity of spatial and appearance transforms is

significantly improved by the infinite possibilities of sampling layers:

$$\phi_a = \phi \times \alpha, \psi_a = \psi \times \beta \tag{1}$$

Second, by applying the sampled transformations ϕ_a and ψ_a to the atlas x_A, we acquire an adversarial generated image x_{ag}. We expect x_{ag} to add extra diversity of data augmentation and maintain realistic as well:

$$x_{ag} = (x_A + \psi_a) \circ \phi_a \tag{2}$$

Finally, both the original generated image x_g and the adversarial generated image x_{ag} are fed to the segmentation network, and the training objective is the min-max game of the adversarial network and the segmentation network. Thus, we ensure the diversity of generation and robustness of segmentation simultaneously by adversarial training:

$$\min_{g(\cdot;\theta_h)} \max_{f(\cdot;\theta_g)} \mathcal{L}_{adv}(\hat{y}_g, \hat{y}_{ag}) \tag{3}$$

$$\mathcal{L}_{adv}(\hat{y}_g, \hat{y}_{ag}) = \frac{\hat{y}_g \cdot \hat{y}_{ag}}{\parallel \hat{y}_g \parallel_2 \cdot \parallel \hat{y}_{ag} \parallel_2} \tag{4}$$

where $f(\cdot;\theta_g)$ and $g(\cdot;\theta_h)$ denote the adversarial network and the segmentation network, respectively. $\hat{y}_g$ and $\hat{y}_{ag}$ are the segmentation predictions of x_g and x_{ag}. It should be noted that since the spatial transformation applied to x_g and x_{ag} is different, the loss calculation is performed in atlas space by inverse registration.

2.3 Uncertainty Rectification

Most of the current methods hypothesize that appearance transformation does not alter the label of the atlas. However, in brain scans with abnormalities such as sTBI, the appearance transformation may include edema, lesions, and *etc*, which may affect the actual semantic labels of the atlas and weaken the accuracy of segmentation. Inspired by [18], we introduce uncertainty rectification to bypass the potential label errors.

Specifically, given a segmentation network, fully augmented image $x_g = (x_A + \psi) \circ \phi$ and spatial-augmented image $x_{As} = x_A \circ \phi$ are fed to the network. The only difference between x_g and x_{As} is that the latter lacks the transformation on appearance. Thus, the two inputs x_g and x_{As} serve different purposes. Fully augmented image x_g is equipped with more diversity compared with x_{As}, as appearance transform has been applied to it, while the spatial augmented image x_{As} has more label authenticity and could guide a more accurate segmentation.

The overall supervised loss consists of two items. First, the segmentation loss $\mathcal{L}_{seg} = \mathcal{L}_{ce}(\hat{y}_{As}, y_g)$ of spatial-augmented image x_{As} guides the network to learn spatial variance only, where $\hat{y}_{As}$ is the prediction of x_{As}, and $\mathcal{L}_{ce}$ denotes cross-entropy loss. Second, the rectified segmentation loss $\mathcal{L}_{rseg}$ of x_g guides the network to learn segmentation under both spatial and appearance transformations. We adopt the KL-divergence D_{KL} of the segmentation results $\hat{y}_{As}$ and

$\hat{y}_g$ as the uncertainty in prediction [18]. Compared with Monte-Carlo dropout [5], KL-divergence for uncertainty estimation does not require multiple forward runs. A voxel with a greater uncertainty indicates a higher possibility of label error in the corresponding location of x_g, thus, the supervision signal of this location should be weakened to reduce the effect of label errors:

$$\mathcal{L}_{rseg} = \exp[-D_{KL}(\hat{y}_g, \hat{y}_{As})]\mathcal{L}_{ce}(\hat{y}_g, y_g) + D_{KL}(\hat{y}_g, \hat{y}_{As}) \tag{5}$$

Thus, the overall supervised loss $\mathcal{L}_{sup} = \mathcal{L}_{seg} + \mathcal{L}_{rseg}$ is the segmentation loss $\mathcal{L}_{seg}$ of x_{As} and the rectified segmentation loss $\mathcal{L}_{rseg}$ of x_g. We apply the overall supervised loss $\mathcal{L}_{sup}$ on both x_g and x_{ag} in practice. During segmentation training, the linear summation of supervised segmentation loss $\mathcal{L}_{sup}$ and adversarial loss $\mathcal{L}_{adv}$ is minimized.

3 Experiments

3.1 Data

We have collected 165 MR T1-weighted scans with sTBI from 2014–2017, acquired on a 3T Siemens MR scanner from Huashan hospital. Among the 165 MR scans, 42 scans are labeled with the 17 consciousness-related brain regions (see appendix for details) while the remaining are left unlabeled, since the manual labeling requires senior-level expertise. Informed consent was obtained from all patients for the use of their information, medical records, and MRI data. All MR scans are linearly aligned to the MNI152 template using FSL [14]. For the atlas image, we randomly collect a normal brain scan at the same institute and label its 17 consciousness-related brain regions as well. During training, the labeled normal brain scan serves as the atlas image, and the 123 unlabeled sTBI scans are used as spatial or appearance references. For one-shot setting, the labeled sTBI scans are used for evaluation only and completely hidden during training. In order to validate the effectiveness of the proposed one-shot segmentation method, a U-Net [13] trained on the labeled scans by 5-fold cross validation is used as the reference of fully supervised segmentation.

3.2 Implementation Details

The framework is implemented with PyTorch 1.12.1 on a Debian Linux server with an NVIDIA RTX 3090 GPU. In practice, the registration network is based on VoxelMorph [1] and pretrained on the unlabeled sTBI scans. During adversarial training, the registration is fixed, while the adversarial network and segmentation network are trained alternately. Both the adversarial network and the segmentation network is based on U-Net [13] architectures and optimized by SGD optimizer with a momentum of 0.9 and weight decay of 1×10^{-4}. The initial learning rate is set to 1×10^{-2} and is slowly reduced with polynomial strategy. We have pretrained the registration network for 100 epochs, and trained the adversarial network and segmentation network for 100 epochs as well. The batch size is set to 1 during training.

Table 1. Ablation studies on different components. **Rand.** denotes uniform transform sampling following [7], **Adv.** denotes adversarial training, and **UR** denotes uncertainty rectification.

Experiments	Dice Coefficient (%)
(1) Baseline	51.4 ± 20.2
(2) Baseline + Rand	53.1 ± 20.6
(3) Baseline + Adv	54.3 ± 19.7
(4) Baseline + UR (wo/ $\mathcal{L}_{seg}$)	48.8 ± 23.5
(5) Baseline + UR (w/ $\mathcal{L}_{seg}$)	53.3 ± 19.0
(6) Baseline + Adv. + UR (w/ $\mathcal{L}_{seg}$)	$\mathbf{56.3 \pm 18.8}$
(Upper Bound) Fully-Supervised U-Net	61.5 ± 16.4

Table 2. Comparison with alternative segmentation methods (%).

	IR	IL	TR	TL	ICRA	ICRP
BrainStorm	46.2±21.9	43.2±21.3	**66.7±18.1**	57.1±7.3	46.8±11.7	34.5±21.0
LT-Net	45.4±24.6	**52.4±21.0**	59.5±17.9	**58.3±10.4**	47.0±19.7	**42.6±17.3**
DeepAtlas	**50.3±25.4**	44.2±27.0	56.4±16.1	57.5±8.7	**50.4±13.8**	38.8±18.7
Proposed	**52.6±24.8**	**54.4±22.9**	**62.0±16.2**	**58.4±10.2**	**51.1±16.6**	**44.2±17.5**
	ICLA	ICLP	CRA	CRP	CLA	CLP
BrainStorm	46.1±14.3	38.8±14.6	50.1±16.6	48.0±13.2	**50.9±11.1**	54.2±10.8
LT-Net	43.5±19.2	42.6±17.2	**52.2±18.4**	**48.4±12.5**	48.5±13.0	**56.5±11.9**
DeepAtlas	**53.8±15.6**	**46.2±16.9**	46.4±19.3	42.5±13.9	43.4±15.8	52.1±13.0
Proposed	**53.0±16.7**	**44.7±18.1**	**56.1±15.0**	**53.2±12.5**	**51.7±12.9**	**60.8±11.1**
	MCR	MCL	IPL	IPR	B	**Average**
BrainStorm	52.5±18.7	57.9±11.9	**50.1±17.1**	52.2±16.5	86.4±4.3	51.9±19.0
LT-Net	**56.1±20.1**	**59.2±12.3**	43.0±15.4	**53.5±18.2**	87.0±6.2	**52.7±19.6**
DeepAtlas	55.9±17.9	54.6±13.2	44.0±17.5	50.8±15.3	**88.9±4.0**	51.5±19.8
Proposed	**61.0±16.9**	**61.1±11.1**	**48.1±18.9**	**53.7±17.0**	**90.1±4.2**	**56.3±18.8**

3.3 Evaluation

We have evaluated our one-shot segmentation framework on the labeled sTBI MR scans in the one-shot setting, which means that only one labeled image is available to learn the segmentation. We explore the effectiveness of the proposed adversarial training and uncertainty rectification, and make the comparison with state-of-the-art alternatives, by reporting the Dice coefficients.

First, we conduct an ablation study to evaluate the impact of adversarial training and uncertainty rectification, which is shown in Table 1. In No. (1), the baseline OSSLT method yields an average Dice of 51.42%. In No. (3), adversarial training adds a performance gain of approximately 3% compared with No. (1), and is also superior to predefined uniform transform sampling in No. (2). The results indicate that adversarial training brings both extra diversity of

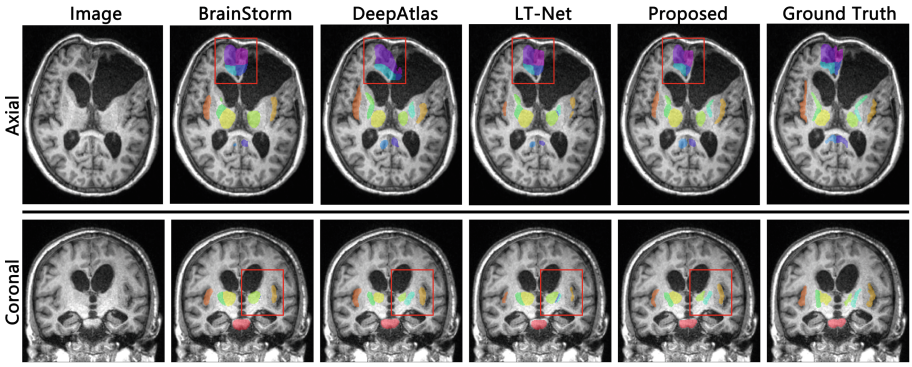

Fig. 3. Visual comparison of different segmentation methods. (Color figure online)

the generated samples and robustness of the segmentation network. The introduction of uncertainty rectification along with segmentation loss $\mathcal{L}_{seg}$ brings a performance gain of approximately 2%, which is shown in No. (5). But in No. (4), if we use rectified segmentation loss $\mathcal{L}_{rseg}$ only (without the segmentation loss $\mathcal{L}_{seg}$ of spatial-augmented image x_{As}), the segmentation performance drops significantly compared with No. (1). This is because the rectified segmentation loss $\mathcal{L}_{rseg}$ does not provide enough supervision signal in regions where the segmentation uncertainty is too high, and thus we need segmentation loss $\mathcal{L}_{seg}$ to compensate it. Finally, by applying both adversarial training and uncertainty rectification, the proposed method yields the best results with an improved Dice coefficient of approximately 5% and a lower standard deviation of segmentation performance, which is shown in No. (5). Experimental results demonstrate that the proposed adversarial training and uncertainty rectification can both contribute to the segmentation performance, compared with the baseline setting.

Then, we compare the proposed method with three cutting-edge alternatives in one-shot medical image segmentation, including BrainStorm [17], LT-Net [15], and DeepAtlas [16], which is shown in Table 2. The proposed method outperforms other segmentation methods with an average Dice score of 56.3%, higher than all of the previous state-of-the-art methods, and achieves the highest and second highest segmentation performance in all of the 17 brain regions. Also, it should be noted that the proposed method has a lower standard deviation in terms of segmentation performance, which also demonstrates the robustness of our method. However, despite the promising results of the proposed method, we have observed that the performance gain of proposed method in certain brain regions that are usually very small is not significant. The plausible reason is that the uncertainty of these small brain regions is too high and affects the segmentation.

For qualitative evaluation, we have visualized the segmentation results of the proposed method and the above-mentioned alternatives, which are shown in Fig. 3. The red bounding boxes indicate the regions where our method achieves

better segmentation results compared with the alternatives. Overall, our method achieves more accurate segmentation, especially in the brain regions affected by ventriculomegaly, compared with BrainStorm, LT-Net, and DeepAtlas.

4 Conclusion

In this work, we present a novel one-shot segmentation method for severe traumatic brain segmentation, a difficult clinical scenario where limited annotated data is available. Our method addresses the critical issues in sTBI brain segmentation, namely, the need for diverse training data and mitigation of potential label errors introduced by appearance transforms. The introduction of adversarial training enhances both the data diversity and segmentation robustness, while uncertainty rectification is designed to compensate for the potential label errors. The experimental results on sTBI brains demonstrate the efficacy of our proposed method and its advantages over state-of-the-art alternatives, highlighting the potential of our method in enabling more accurate segmentation in severe traumatic brains, which may aid clinical pipelines.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (No. 62001292).

References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging **38**(8), 1788–1800 (2019)
2. Chen, C., et al.: Enhancing MR image segmentation with realistic adversarial data augmentation. Med. Image Anal. **82**, 102597 (2022)
3. Ding, Y., Yu, X., Yang, Y.: Modeling the probabilistic distribution of unlabeled data for one-shot medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 1246–1254 (2021)
4. Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012)
5. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: International Conference on Machine Learning, pp. 1050–1059. PMLR (2016)
6. He, Y., et al.: Learning better registration to learn better few-shot medical image segmentation: authenticity, diversity, and robustness. IEEE Trans. Neural Netw. Learn. Syst. (2022)
7. He, Y., et al.: Deep complementary joint model for complex scene registration and few-shot segmentation on medical images. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12363, pp. 770–786. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58523-5_45
8. Huang, Z., et al.: The self and its resting state in consciousness: an investigation of the vegetative state. Hum. Brain Mapp. **35**(5), 1997–2008 (2014)
9. Olut, S., Shen, Z., Xu, Z., Gerber, S., Niethammer, M.: Adversarial data augmentation via deformation statistics. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12374, pp. 643–659. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58526-6_38

10. Qiao, Y., Tao, H., Huo, J., Shen, W., Wang, Q., Zhang, L.: Robust hydrocephalus brain segmentation via globally and locally spatial guidance. In: Abdulkadir, A., et al. (eds.) MLCN 2021. LNCS, vol. 13001, pp. 92–100. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87586-2_10
11. Qin, P., et al.: How are different neural networks related to consciousness? Ann. Neurol. **78**(4), 594–605 (2015)
12. Ren, X., Huo, J., Xuan, K., Wei, D., Zhang, L., Wang, Q.: Robust brain magnetic resonance image segmentation for hydrocephalus patients: hard and soft attention. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 385–389. IEEE (2020)
13. 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
14. Smith, S.M., et al.: Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage **23**, S208–S219 (2004)
15. Wang, S., et al.: LT-Net: label transfer by learning reversible voxel-wise correspondence for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9162–9171 (2020)
16. Xu, Z., Niethammer, M.: DeepAtlas: joint semi-supervised learning of image registration and segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 420–429. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_47
17. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V.: Data augmentation using learned transformations for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8543–8553 (2019)
18. Zheng, Z., Yang, Y.: Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. Int. J. Comput. Vision **129**(4), 1106–1120 (2021)