Robust Hydrocephalus Brain Segmentation via Globally and Locally Spatial Guidance

Yuanfang Qiao, Haoyi Tao, Jiayu Huo, Wenjun Shen, Qian Wang, and Lichi Zhang(✉)

School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
lichizhang@sjtu.edu.cn

Abstract. Segmentation of brain regions for hydrocephalus MR images is pivotally important for quantitatively evaluating patients' abnormalities. However, the brain image data obtained from hydrocephalus patients always have large deformations and lesion occupancies compared to the normal subjects. This leads to the disruption of the brain's anatomical structure and the dramatic changes in the shape and location of the brain regions, which poses a significant challenge to the segmentation task. In this paper, we propose a novel segmentation framework, with two modules to better locate and segment these highly distorted brain regions. First, to provide the global anatomical structure information and the absolute position of target regions for segmentation, we use a dual-path registration network which is incorporated into the framework and trained simultaneously together. Second, we develop a novel Positional Correlation Attention Block (PCAB) to introduce the local prior information about the relative positional correlations between different regions, so that the segmentation network can be guided in locating the target regions. In this way, the segmentation framework can be trained with spatial guidance from both global and local positional priors to ensure the robustness of the segmentation. We evaluated our method on the brain MR data of hydrocephalus patients by segmenting 17 consciousness-related ROIs and demonstrated that the proposed method can achieve high performance on the image data with high variations of deformations. Source code is available at: https://github.com/JoeeYF/TBI-Brain-Region-Segmentation.

Keywords: Image segmentation · Image registration · Hydrocephalus

1 Introduction

Hydrocephalus is an abnormal accumulation of cerebrospinal fluid (CSF) in the patient's brain with persistent ventricular dilatation, which is a secondary injury from Traumatic Brain Injury (TBI). It is usually caused by the obstruction of cerebrospinal fluid pathways, impaired cerebrospinal fluid circulation, etc., which

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-87586-2_10) contains supplementary material, which is available to authorized users.

A. Abdulkadir et al. (Eds.): MLCN 2021, LNCS 13001, pp. 92–100, 2021.
https://doi.org/10.1007/978-3-030-87586-2_10

leaves the patient in a state of impaired consciousness [3]. The evaluations of the brain's abnormalities and the correspondence to the patient's consciousness state play an important role to assist the clinical assessments of the disease progression. Specifically, it is demonstrated in [8] that there are 17 brain regions whose functional and anatomical shape states have certain correlations with the improvements of the consciousness level. Therefore, the identification and parcellation of these consciousness-related brain regions are demanding. However, it is generally impractical to conduct segmentation manually, which is tedious, time-consuming, and introduces inter-observer variability. The issues become more deteriorated for hydrocephalus brain images, which contain even higher variability and extent brain changes than the normal brain images. Therefore, the development of an accurate and automatic brain parcellation method for hydrocephalus images would be highly beneficial.

With the development of deep learning, the convolutional neural network (CNN) and its extensions have dominated the field of medical image segmentation. Specifically, UNet [5] combined the high-level and low-level features with different context information to estimate precise segmentation results, and has become the most applied method in this field. Many attempts were also made specifically for brain image segmentation. For example, Moeskops et al. [11] used a CNN to achieve the automatic segmentation of MR brain images into some tissue classes. Ghafoorian et al. [6] integrated the anatomical location information into the network to get explicit location features and improve the segmentation results substantially. The alignment-based brain mapping methods, which adopt the registration techniques such as VoxelMorph [2] to align the brain template to the target for brain parcellation, have also been widely applied.

However, most brain region segmentation methods are designed for normal brain images, while the hydrocephalus images have much higher variations of anatomical structures compared to the normal, due to the large deformations and lesion occupancies caused by the diseases as shown in Fig. 1. The brain anatomical structure information is much more complicated to be encoded for constructing the segmentation model, especially when the training data are also limited. Some attempts have been made to resolve these issues in hydrocephalus brain segmentation. Ledig et al. [10] intended to develop a multi-atlas-based method to segment the brain regions on TBI data, but the experiments reported the failed cases. Ren et al. [12] proposed a two-stage framework with hard and soft attention modules to segment the brain regions of hydrocephalus brain images and demonstrated that it outperforms the state-of-art methods such as UNet. However, it is not an end-to-end framework, where the two modules are not fully integrated into a single network to fully share their conducted features.

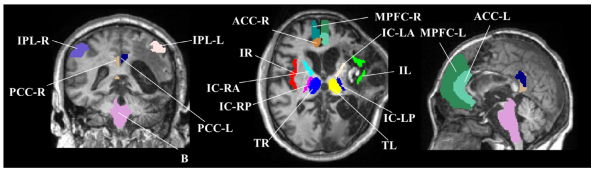

Fig. 1. Exemplar hydrocephalus MR brain images with consciousness-related regions.

In this paper, we intend to parcellate the 17 consciousness-related regions according to [8,12] from the MR hydrocephalus images. To resolve the issues in hydrocephalus brain segmentation, we focus on developing an end-to-end novel framework, with two modules to locate and segment the target regions from two different perspectives: (1) We use a registration guidance module to produce the segmentation network with more anatomical structure information about the absolute position of target regions in the whole brain. (2) We propose a novel Positional Correlation Attention Block (PCAB) integrated into the UNet to improve performance by conducting more explicitly structural features. In the PCAB, we design a Positional Correlation Layer (PCL) to extract the relative positional relationship between different brain regions and use it to refine the segmentation estimation via the attention layer. The absolute and relative position information can provide complementary and comprehensive guidance to the segmentation network, which is designed as an end-to-end network for better information integration. Experiments show that it can outperform the alternatives with a statistical significance, which is evaluated by 5-fold cross-validation on 17 brain regions with great deformations from the collected hydrocephalus brain images.

2 Method

To reduce the impact of the deformation caused by the occupancy and erosion of lesion areas on normal brain areas, we propose two novel modules to extract more comprehensive anatomical structure information from the absolute location in the whole brain and the correlation with each other's location. Figure 2 shows our proposed segmentation framework with dual-path registration module and Positional Correlation Attention Block (PCAB) module. The two networks are trained simultaneously but only the segmentation network is used in the inference stage.

2.1 Guidance with Registration Module

We use an UNet-like network with a dual-path encoder to non-rigidly align the hydrocephalus subject (moving set) with the brain template (fixed set), which is shown in Fig. 2(b). Specifically, we adopt a dual-path encoder that takes the original brain image (I_M, I_F) and the brain segmentation mask (M_M, M_F) as two channels input to make the registration network focus on the ROI regions. The features extracted from the encoder are concatenated with the corresponding decoder features by skip connections to make a low-level and high-level feature fusion. The decoder predicts a deformation field ϕ to align the mask of different brain regions. Finally, the hydrocephalus brain image and brain regions mask are warped using Spatial Transformation Network (STN) according to ϕ. To introduce more supervised information into the segmentation framework and improve the robustness [13], we warp the output prediction map of the segmentation network according to ϕ above and calculate the similarity with the template dataset. Then the segmentation and registration network can be trained simultaneously.

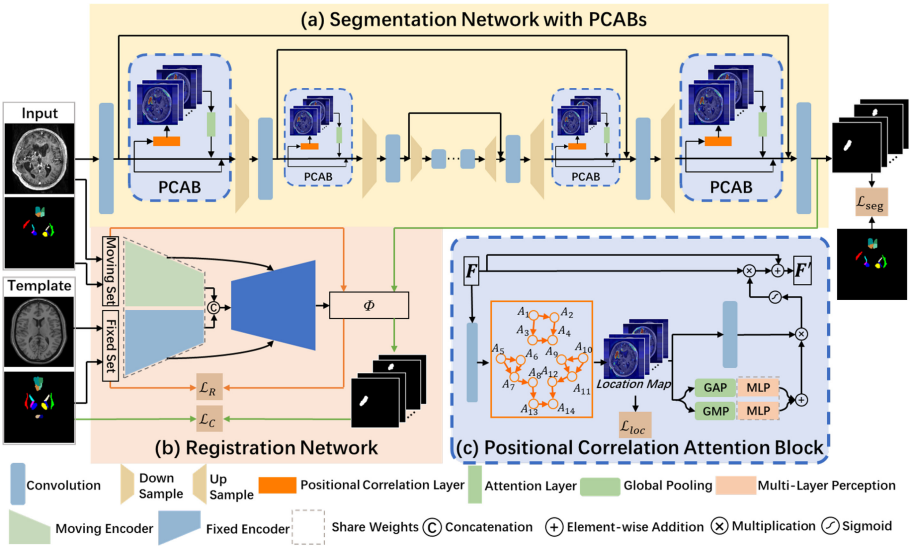

Fig. 2. The overview of the proposed method. The framework consists of a segmentation network and a registration network. The Positional Correlation Attention Blocks (PCAB) are integrated into the segmentation network. Note that only the segmentation network is used in the inference stage.

2.2 Segmentation with Positional Correlation Attention Block

We use a 3D-UNet as our main segmentation network, and PCAB is integrated into different levels. The purpose of PCAB is to use the location correlation of different brain regions to generate the attention map with anatomical structure information. The PCAB includes a Positional Correlation Layer (PCL) to estimate the location map and an attention layer to generate the attention map to refine the input feature. As shown in Fig. 2(a), the PCAB is a plug-and-play block and is integrated into the first two levels of the encoder and the last two levels of the decoder in our study.

Positional Correlation Layer. Localization probability maps for each brain region $\mathrm{A} \in \mathbb{R}^{\mathrm{l}\times\mathrm{w}\times\mathrm{h}\times 17}$ are calculated from the input feature $\mathrm{F} \in \mathbb{R}^{\mathrm{l}\times\mathrm{w}\times\mathrm{h}\times\mathrm{c}}$ by a convolutional layer. Our PCL forms the target brain regions into two directed graph structures according to their correlation with each other's location and the graph is shown in Fig. 2(c). The probability maps for every brain region are passed to the adjacent region using a convolutional layer according to the linkage between each brain region in the graph above. The method has been used in pose estimation [4]. To further expand receptive fields, we use the 3D dilated convolution with dilation rate of 2 and the filter size of 7×7. Note that due to the large distance between some regions, we exclude 3 brain regions and constructed 2 directed graph structures using the remaining 14 brain regions.

Specifically, let A_k be the original feature maps and k is the index of brain regions. The positional correlation for A_k can be defined as

$$\tilde{\mathrm{A}}_k = f\left(\mathrm{A}_k + \sum_{j \in N} C(\mathrm{A}_j)\right), \tag{1}$$

where $j \in N$ means that A_j is the brain region that has linkage with A_k, f is the ReLU function, and C is the dilated convolution. Take A_{11} as an example, A_{11} is refined by receiving information from $\tilde{\mathrm{A}}_9$ and A_{10}, so the updated $\tilde{\mathrm{A}}_{11}$ after PCL layer is

$$\tilde{\mathrm{A}}_{11} = f(\mathrm{A}_{11} + C(\tilde{\mathrm{A}}_9) + C(\mathrm{A}_{10})). \tag{2}$$

Since the graph is starting from A_1, A_5 and A_{10}, and they do not receive information from other regions, they remain the same as the original ones. The other region's feature map is refined in a similar way to A_{11}.

Attention Layer. The 17 location maps generated by PCL may have different priorities for the segmentation task. Here we use an attention layer consisting of spatial attention and channel attention to exploiting the most significant features to refine the input feature. First, we use a convolution layer with an output channel number of c to generate a feature map $\alpha \in \mathbb{R}^{\mathrm{l}\times\mathrm{w}\times\mathrm{h}\times\mathrm{c}}$, which has the same size as the input feature. Next, we use a global max pooling (GMP) and a global average pooling (GAP) to get the global information of each channel, which are represented as $\mathrm{F}_{max} \in \mathbb{R}^{1\times1\times1\times17}$ and $\mathrm{F}_{avg} \in \mathbb{R}^{1\times1\times1\times17}$, respectively. Then, the channel attention coefficient $\beta \in \mathbb{R}^{1\times1\times1\times\mathrm{c}}$ is calculated by a multiple layer perception consisting of two fully connected layers and a ReLU activation. The output feature of PCAB $\tilde{\mathrm{F}}$ can be obtained as $\tilde{\mathrm{F}} = \mathrm{F} + \sigma(\alpha \cdot \beta) \cdot \mathrm{F}$ and is fed into the next convolution block, where σ denotes a sigmoid function.

2.3 Training Strategy

The registration network loss consists of three following components:

$$\mathcal{L}_{\mathrm{R}} = \beta_1 \mathcal{L}_{\mathrm{NCC}}(\mathrm{I}_M(\phi), \mathrm{I}_F) + \beta_2 \mathcal{L}_{\mathrm{mask}}(\mathrm{M}_M(\phi), \mathrm{M}_F) + \beta_3 \mathcal{L}_{\mathrm{smooth}}(\phi), \tag{3}$$

where the first term is local cross-correlation loss between the warped image and fixed image which is the same as VoxelMorph, the second term calculates the mean square error (MSE) between the brain region mask, and the last term penalizes local spatial variations ϕ to keep it smooth.

The segmentation network is optimized by two aspects of supervision information. The loss function is written as follows:

$$\mathcal{L}_{\mathrm{S}} = \alpha_1 \mathcal{L}_{\mathrm{seg}}(\hat{\mathrm{Y}}, \mathrm{Y}_G) + \alpha_2 \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_{\mathrm{loc}}^{i}(\tilde{\mathrm{A}}^i, \mathrm{A}_G^i). \tag{4}$$

The first term in Eq. (4) calculates the cross-entropy between the prediction map of the segmentation network $\hat{\mathrm{Y}}$ and the ground-truth Y_G. The second term

calculates the weighted binary-cross-entropy between the location maps $\tilde{\mathrm{A}}^i$ from PCL at each level i and the ground-truth A_G^i. A_G^i is the center of each brain region which is dilated into ball areas with a radius of 3. To balance the loss of foreground and background, we use the inverse of the average pixel value of ground-truth as the loss weight of the foreground. The equation for the second term is shown below, where V^i denotes the sum of voxels for level i:

$$\mathcal{L}_{\mathrm{loc}}^i(\tilde{\mathrm{A}}, \mathrm{A}_G) = -\frac{\mathrm{V}^i}{\sum \mathrm{A}_G^i} \cdot \mathrm{A}_G^i \cdot \log \tilde{\mathrm{A}}^i - \left(1 - \mathrm{A}_G^i\right) \cdot \log\left(1 - \tilde{\mathrm{A}}^i\right). \tag{5}$$

During the training stage, we warp the segmentation map according to ϕ and calculate the MSE with brain template ROIs, which is defined as

$$\mathcal{L}_{\mathrm{C}} = \gamma \mathcal{L}_{\mathrm{mask}}(\hat{\mathrm{Y}}(\phi), \mathrm{M}_F), \tag{6}$$

where we use $\mathcal{L}_{\mathrm{C}}$ to train the segmentation network and registration network simultaneously, but only when the training epoch is greater than 15 to keep the training process stable. In this way, the overall loss function can be defined as $\mathcal{L} = \mathcal{L}_{\mathrm{S}} + \mathcal{L}_{\mathrm{R}} + \mathcal{L}_{\mathrm{C}}$. Note that β_1, β_2, β_3 in Eq. (3), α_1, α_2 in Eq. (4) and γ in Eq. (6) are hyper-parameters and γ is 0 when the train epochs are less than 15.

3 Experiments and Results

3.1 Datasets and Experiments

The proposed method was evaluated on in-house MR brain data in T1 from 44 hydrocephalus patients and all subjects have hematoma volume and hydrocephalus disease. The brain template used in the registration network is Colin 27 Average Brain [7]. The 17 consciousness-related ROIs shown in Fig. 1 was manually delineated and used for training. Before the data was fed into the network, we preprocessed the data by the following steps: The images' voxel spacing was resized to $1\,\mathrm{mm} \times 1\,\mathrm{mm} \times 1\,\mathrm{mm}$. Then the histogram matching was conducted for intensity normalization. In the registration module, the moving set was firstly affine-registered to the fixed set using the ANTs package [1]. The data were grouped into 5-fold for the cross-validation and ablation study.

The network was implemented using PyTorch1.6 and trained on NVIDIA RTX Titan GPU. We used different learning rate settings for the two networks: 1e−3 for the registration network and 1e−4 for the segmentation network. All learning rate settings have a decay of 0.1 every 5 epochs and the epochs of training are 100. Adam optimizer was adopted with a weight decay of 1e−4. Due to the limitation of the GPU's memory size, the batch size was set to 1. Therefore, we replaced the BatchNorm with GroupNorm to reduce the effect of small batch size, and the number of groups were set to 4.

3.2 Results

To verify the effect of the proposed method, we make the ablation studies by adjusting the number of PCABs denoted as N in Eq. (4) and the incorporation of

Table 1. Dice Coefficient of ablation studies and comparison with other methods. PCABs(N) means that there are N PCABs integrated into the framework.

Methods	Dice coefficient(%)
(a) Ablation studies	
UNet(Baseline)	61.64 ± 21.60
UNet+Registration	63.85 ± 18.71
UNet+PCABs(2)	64.83 ± 18.67
UNet+PCABs(4)	66.00 ± 17.76
UNet+PCABs(4)+Registration	$\mathbf{69.03 \pm 14.85}$
(b) Comparison with other methods	
VoxelMorph[2]	39.23 ± 22.23
UNet [5]	61.64 ± 21.60
nnUNet [9]	63.37 ± 23.62
Ren et al. [12]	67.19 ± 17.18
Proposed	$\mathbf{69.03 \pm 14.85}$

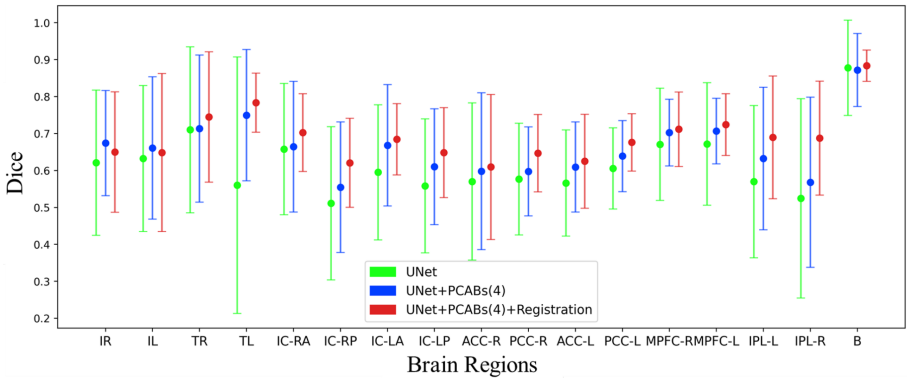

Fig. 3. Dice Coefficient of each brain region in different studies. (Color figure online)

the registration network, which are shown in Table 1(a). PCABs(2) means that the PCAB is only integrated into the first level and the last level of decoder, while PCABs(4) integrated two more PCABs in the deeper levels like Fig. 2. We use Dice Coefficient Score (DCS) as the evaluation metrics.

As presented in Table 1(a), the DSC score is increased by 4.36% by applying PCAB into baseline UNet, and with the growth of the number of PCABs the performance also increases. The results indicate that the PCABs can help the network to encode the variations of the anatomical structure by extracting the positional correlations information. By training the segmentation network and registration network together, the DSC further improved 3.03% and the standard deviation decreased 2.91% which proved that the registration network further confirmed the absolute position of the regions in the whole brain. The results also show that the both two networks helps to achieve the best segmentation

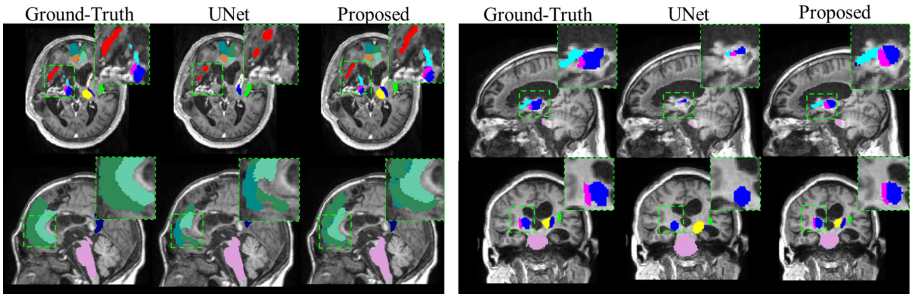

Fig. 4. Visualization of segmentation results comparison of UNet and the proposed method.

performance. Figure 3 shows the distribution of dice scores of each brain region in the different ablation studies. It can be concluded that the method proposed has better performance than the baseline method. The final proposed methods (red) have higher average dice and lower standard deviation in most of the brain regions, indicating that our method can achieve more accurate and more stable results for the hydrocephalus dataset.

We also compared our method with other state-of-the-art segmentation methods. The results are shown in Table 1(b). Our method achieved the average dice of 69.03% with a standard deviation of 14.85%, which is 1.84% higher than the recent state-of-the-art method [12]. The lower standard deviation also indicates that the result has improved more on the hard subject samples which have larger deformation than the others, where the proposed method achieves more robust performance than the alternatives.

Figure 4 visualizes the results from our segmentation framework and the UNet method. The first column is ground-truth and the next column are the results of UNet and the method proposed, respectively. As shown in the zoomed block, our method can locate and segment the brain regions, while the UNet fails to handle. For the cases with large deformation, there is a huge improvement in accuracy and robustness with our proposed method.

4 Conclusion

We have proposed a novel segmentation framework based on globally registration guidance and locally relative positional correlations guidance for the hydrocephalus dataset. First, we used a dual-path registration network to provide global anatomical structure information which was important to the final segmentation results. Besides, to make the network focus more on the target regions, we integrated a Positional Correlation Attention Block (PCAB) into UNet to generate a location attention map to refine the feature of the encoder and decoder. The comparative results show that the proposed method can achieve higher results and have a greater improvement for cases with larger deformation, indicating that our method has high accuracy and robustness.

Acknowledgements. This work was supported by the National Key Research and Development Program of China (2018YFC0116400), National Natural Science Foundation of China (NSFC) grants (62001292), Shanghai Pujiang Program (19PJ1406800), and Interdisciplinary Program of Shanghai Jiao Tong University.

References

1. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage **54**(3), 2033–2044 (2011)
2. 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)
3. Chari, A., Czosnyka, M., Richards, H.K., Pickard, J.D., Czosnyka, Z.H.: Hydrocephalus shunt technology: 20 years of experience from the Cambridge shunt evaluation laboratory. J. Neurosurg. **120**(3), 697–707 (2014)
4. Chu, X., Ouyang, W., Li, H., Wang, X.: Structured feature learning for pose estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4715–4723 (2016)
5. 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
6. Ghafoorian, M., et al.: Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. Sci. Rep. **7**(1), 1–12 (2017)
7. Holmes, C.J., Hoge, R., Collins, L., Woods, R., Evans, A.C.: Enhancement of MR images using registration for signal averaging. J. Comput. Assist. Tomogr. **3**(2), 324–333 (1998)
8. Huo, J., et al.: Neuroimage-based consciousness evaluation of patients with secondary doubtful hydrocephalus before and after lumbar drainage. Neurosci. Bull. (9) (2020)
9. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021)
10. Ledig, C., et al.: Robust whole-brain segmentation: application to traumatic brain injury. Med. Image Anal. **21**(1), 40–58 (2015)
11. 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)
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. Xu, Z., Niethammer, M.: DeepAtlas: joint semi-supervised learning of image registration and segmentation. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 420–429. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_47