Semi-supervised Segmentation with Self-training Based on Quality Estimation and Refinement

Zhou Zheng[1], Xiaoxia Wang[2], Xiaoyun Zhang[1(✉)], Yumin Zhong[2(✉)], Xiaofen Yao[2], Ya Zhang[1], and Yanfeng Wang[1]

[1] Cooperative Medianet Innovation Center, Shanghai Jiao Yong University, Shanghai, China
xiaoyun.zhang@sjtu.edu.cn
[2] Shanghai Children's Medical Center, Shanghai, China
zhongyumin@scmc.com.cn

Abstract. Building a large dataset with high-quality annotations for medical image segmentation is time-consuming and highly depends on expert knowledge. Therefore semi-supervised segmentation has been investigated by utilizing a small set of labeled data and a large set of unlabeled data with generated pseudo labels, but the quality of pseudo labels is crucial since bad labels may lead to even worse segmentation. In this paper, we propose a novel semi-supervised segmentation framework which can automatically estimate and refine the quality of pseudo labels, and select only those good samples to expand the training set for self-training. Specifically the quality is automatically estimated in the view of shape and semantic confidence using variational auto-encoder (VAE) and CNN based network. And, the selected labels are refined in an adversarial way by distinguishing whether a label is the ground truth mask or not at pixel level. Our method is evaluated on the established neuroblastoma(NB) and BraTS18 dataset and outperforms other state-of-the-art semi-supervised medical image segmentation methods. We can achieve a fully supervised performance while requiring ∼4x less annotation effort.

Keywords: Medical image segmentation · Semi-supervised learning · Quality estimation and refinement

1 Introduction

Recently, deep Fully Convolutional Neural networks (FCN) [1] have gained much popularity in the medical image segmentation because of its ability to learn the most discriminative pixel-wise features. Based on the encoder-decoder structure of FCN, U-Net [2] proposes a skip connection between the encoder and decoder layers which can utilize the low level features to improve the segmentation. However, supervised learning requires laborious pixel-level annotation which is very

M. Liu et al. (Eds.): MLMI 2020, LNCS 12436, pp. 30–39, 2020.
https://doi.org/10.1007/978-3-030-59861-7_4

time-consuming and needs expert knowledge. In practice, usually only a small set of data with pixel-level annotations are affordable while most of the data left unlabeled. Semi-supervised learning(SSL) aims to solve this challenge by exploring the potential of unlabeled data to improve the performance of a segmentation model.

Because of the characteristics of fuzzy texture structure, low-contrast intensity and limited amount of data, medical image segmentation is extremely challenging in learning based semi-supervised setting, and some methods have been proposed to address the problem. The work in [3,4] have explored the consistency of transformation between labeled and unlabeled data, where a consistency loss is incorporated into the loss function and provides regularization for training the network. Another direction in semi-supervised segmentation is co-training [5–7], which tries to improve segmentation with the assistance of other auxiliary tasks. Also, transfer learning is adopted in semi-supervised segmentation such as [8,9] by employing external labeled data sets via domain adaption.

One practical direction for semi-supervised learning is self-training [10] which is the earliest SSL approach and became popular in deep learning schemes [15, 16]. In this setting, after finishing the supervised training stage of a segmentation model, it is possible to continue the learning process on new unlabeled data by creating pseudo labels for the unlabeled data. However, the quality of the generated pseudo labels is not guaranteed for retraining the segmentation model, which limits their potential for improvements from the data with pseudo label and sometimes even makes the updated model worse. The work in [11] just selects the confident region in the segmentation map to train the network which focuses more on the background area and ignores the tumor region. Such method can limit the negative impact of bad pseudo labels but it can not make full use of the unlabeled data.

In this paper, we propose a novel self-training method which can automatically estimate and refine the quality of pseudo labels, and select only those good samples to expand the data set for retraining the segmentation network. The quality of the pseudo label is estimated from the view of shape confidence and semantic confidence. The former estimates the shape matching between the prediction map and the label mask, while the latter evaluates the semantic matching between the prediction map and the raw image. By ranking the quality of the predicted segmentation, we choose the top K samples as the pseudo labels. In considering that the selected pseudo labels may still have some obvious mis-segmentation for possible improvement, we further refine the label in an adversarial way by distinguishing whether it is the ground truth mask or not at pixel level. After quality estimation and refinement, the samples with good quality pseudo labels are added to expand the training set, which is then utilized to retrain and update the segmentation network. This process can be iterated until to a satisfied result. In addition, the refinement network can also be employed during inference to obtain a refined and better segmentation.

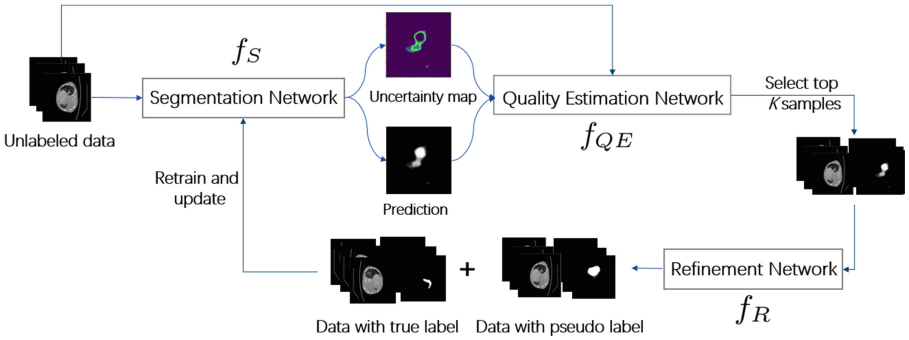

Fig. 1. Overview of the architecture of our proposed semi-supervised approach.

In this work, our contributions can be summarized as follows:

- We propose a novel semi-supervised segmentation method which can automatically estimate the quality of a segmentation and refine it in an adversarial way. Then a segmentation model can be retrained by the expanded dataset with the selected good pseudo labels in a self-training way.
- We design a robust quality estimation network by estimating the shape confidence and the semantic confidence with Variational Auto-encoder(VAE) and VGG [12] based network. Also, a refinement network is proposed to refine the generated pseudo label for higher quality expanded dataset.
- We establish a neuroblastoma segmentation dataset which contains 430 cases of young children's CT with manually-annotated label by doctors. Experiments on NB and BraTS18 dataset demonstrate the robustness and effectiveness of our method compared to other semi-supervised methods.

2 Method

2.1 Overview

The overview structure of our proposed framework is shown in Fig. 1 which consists of three modules: segmentation f_S, quality estimation f_{QE} and refinement f_R.

Given a set of labeled data $X_L = \{x_L^1, x_L^2, ..., x_L^m\}$ with corresponding label $Y_L = \{y_1, y_2, \ldots, y_m\}$ and a large set of unlabeled data $X_U = \{x_U^1, x_U^2, \ldots, x_U^n\}$, we can generate pseudo label $Y_U^{'}$ and uncertainty map U for X_U with f_S. However, the quality of generated pseudo label $Y_U^{'}$ is not guaranteed for retraining the segmentation model. The addition of good quality training data can improve the performance of the segmentation model greatly so we need to filter the pseudo label $Y_U^{'}$ to ensure that the model can obtain a high-quality subset $Y_{sub}^{'}$ and $X_{sub}^{'}$ to expand the training data. To this end, we design a quality estimation module f_{QE} which can provide a reliable estimation about the quality of the

pseudo label $Y_U^{'}$. In considering that the selected pseudo labels may still have some obvious mis-segmentation for possible improvement, we introduce a refinement module f_R to refine the pseudo labels $Y_{sub}^{'}$ in an adversarial way. Our goal is to retrain the segmentation model f_S from training data $X_T = X_L \cup X_{sub}^{'}$, training label $Y_T = Y_L \cup Y_{sub}^{'}$ in a self-training way. We repeat the above steps till to obtain a final model f_S with expected results.

2.2 Network Architecture and Loss Functions

Quality Estimation Module: A segmentation network is very vulnerable to the quality and quantity of annotations as it implements segmentation at pixel level. This quality estimation module can provide a reliable estimation about the quality of the predicted segmentation or the pseudo label $Y_U^{'}$ so we can use it to select a subset $Y_{sub}^{'}$ of $Y_U^{'}$ to expand the training dataset.

We consider the quality of the pseudo label from the view of shape confidence and semantic confidence. As shown in Fig. 2(a), we utilize a VAE for shape representation learning where the encoder learns a low dimensional space for the underlying distribution of the shape prior such as the continuity and boundary smoothness etc. We utilize the latent vector z to distinguish whether it is the real mask or the prediction map in order to provide a shape confidence of the prediction map.

We also introduce the semantic confidence network which looks into the relationship between the prediction map and its surrounding tissue to evaluate the semantic matching between the prediction map and the input image. Besides the original image and the segmentation map, we also feed the semantic confidence network with the uncertainty map U produced by the segmentation network to provide it with information about the uncertainty region in the segmentation map. VGG16 [12] is the backbone architecture of the semantic confidence network.

We then fuse the shape and semantic confidence to form the final quality by several FC layers. We adopt mean absolute error loss L_q, binary cross entropy loss L_d and mean square error loss L_{vae} to train the semantic and fusion branch, discriminator and VAE respectively.

$$L_q = |q - y| \tag{1}$$

$$L_d = -(y_z \log(D(z)) + (1 - y_z) \log(1 - D(z))) \tag{2}$$

$$L_{vae} = \sum |x_i' - x_i| + KL(p(z|x)||q(z)) \tag{3}$$

Where q is the quality output, y is the true DSC, z is the latent vector of the VAE and y_z is $\{0, 1\}$ where 0 for prediction and 1 for label, x_i and x_i' is the input and reconstruction data, $q(z) \sim N(0, I)$.

Refinement Module: The selected pseudo labels may still have some obvious mis-segmentation, so we design a refinement module to refine the pseudo labels in

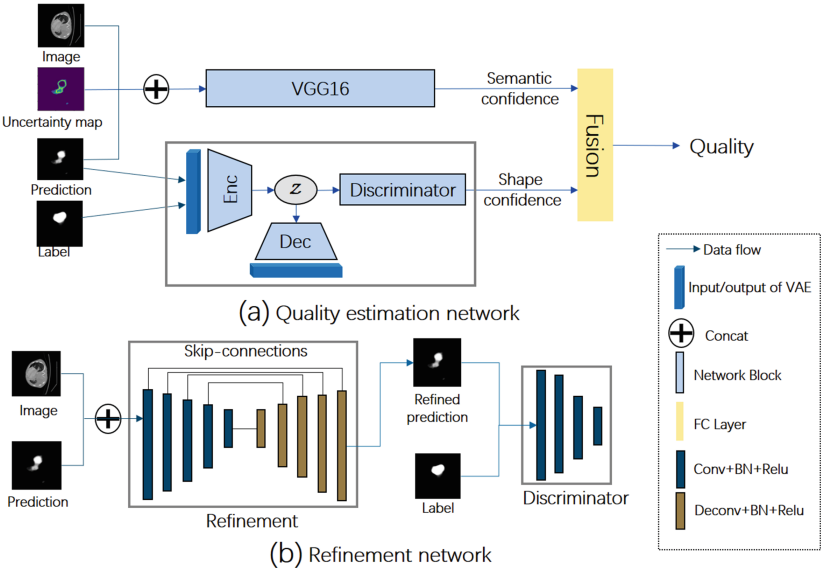

Fig. 2. (a) Quality estimation network consists of two branches. One is to estimate the shape matching by distinguish the latent vector z of prediction and label while reconstructing them, the other is to estimate the semantic matching between the prediction and the image from VGG16. Two confidences are fused by FC layers to form the estimated quality. (b) Refinement network is to improve the prediction by adversarial learning.

an adversarial way as shown in Fig. 2(b). Our generator utilizes the U-Net [2] as the backbone and takes the original image and the corresponding segmentation map as input and outputs a refined map. Different from the typical discriminator which discriminates the map at the image level, we propose a fully convolutional discriminator like [13] that learns to differentiate the predicted probability maps from the ground truth segmentation distribution at pixel level.

We train the refinement network f_R by minimizing a weighted joint of binary cross entropy and adversarial loss L_r:

$$L_r = -((y_i \log(f_R(x_i)) + (1 - y_i) \log(1 - f_R(x_i))) + \lambda L_{adv}) \tag{4}$$

$$L_{adv} = \mathbb{E}_{y_i \sim Y}[\log(D(y_i))] + \mathbb{E}_{y_i' \sim Y'}[\log(1 - D(y_i'))] \tag{5}$$

where L_{adv} is the adversarial loss, λ is the weighted coefficient.

Segmentation Module: The architecture of the segmentation module is same to U-Net [2] which has skip-connections, allowing the transfer of low-level features from the encoder to the decoder. It accepts the image and produces a segmentation probability map with an uncertainty map which is sent to the quality estimation network f_{QE}. The objective function L_s of this module is a

sum of binary cross entropy L_{bce} and the Kullback–Leibler divergence loss L_{kl} which can provide a distribution constraint between mask Y and prediction P.

$$L_{bce} = -(y_i \log(f_S(x_i)) + (1 - y_i) \log(1 - (f_S(x_i)))) \tag{6}$$

$$L_{kl} = y_i * (log(y_i/p_i)) \tag{7}$$

$$L_s = L_{bce} + L_{kl} \tag{8}$$

The output uncertainty map $U = \{u_i\}$ is obtained from the prediction map P by calculating the margin between the positive and negative probability $u_i = 1 - |p_i - (1 - p_i)|$, where p_i is the predicted probability for pixel x_i.

2.3 Training Strategy

Our entire algorithm is summarized in Algorithm 1. To improve the performance of our core quality estimation module f_{QE}, we can pre-train it on any public datasets and tranfer it to our target dataset. The training of f_{QE} follows that the VAE and discriminator are firstly trained with L_{vae} and L_d respectively, and then train the CNN network and fusion layers with L_q by fixing the VAE and discriminator.

Algorithm 1. Training process of our method.

Step 1: Pre-train f_{QE} with any public datasets.
Step 2: Train f_S, f_R and fine-tune f_{QE} with labeled data X_L, Y_L by Eq.1 – Eq.8.
Step 3: Generate pseudo label $Y_U^{'}$ for X_U with f_S. Select a subset $Y_{sub}^{'}$ and refine it by f_{QE} and f_R respectively.
Step 4: Expand train dataset $X_T = X_L + X_{sub}^{'}$, $Y_T = Y_L + Y_{sub}^{'}$. Retrain and update f_S with X_T, Y_T.
Step 5: Repeat step 3 and step 4 for several iterations.

3 Experiment Results

3.1 Dataset and Implementation Details

NB Dataset: We establish a neuroblastoma segmentation dataset which consists of 430 CT scans of children, with a manually-annotated label by expert doctors from **Shanghai Children's Medical Center**. The dataset is divided into two parts: training set (344 cases) and testing set (86 cases). The intra-slices resolution is 512 × 512, and the number of slices varies from 67 to 395 and the voxel size is $0.49 \times 0.49 \times 1.24mm^3$ in average.

BraTS18 Dataset [17]**:** 210(train:160, test:50) MRI scans from patients with high grade glioma and 75(train:60, test:15) MRI scans from patients with low grade glioma are split into training set and testing set. To simplify comparison between different segmentation methods, we perform binary classification and segment only the whole tumor with the FLAIR sequence.

Table 1. Experiment results on NB dataset by different methods.

Labeled (unlabeled)	33(311)			106(238)			169(175)			344(0)		
Methods	DSC	HD	ASD	DSC	HD	ASD	DSC	HD	ASD	DSC	HD	ASD
U-Net	54.76	24.49	3.54	68.21	17.69	0.82	71.85	16.65	0.88	77.91	14.09	0.80
MASSL [5]	63.94	18.53	1.67	72.72	17.41	1.07	74.92	14.88	0.84	-	-	-
ASDNet [11]	65.39	21.01	1.43	72.93	17.47	1.41	75.88	14.45	0.66	-	-	-
TCSM [4]	68.15	17.55	1.12	73.09	17.31	1.08	76.70	13.98	0.81	-	-	-
Ours	**71.79**	**17.46**	**1.09**	**77.64**	**15.13**	**0.78**	**80.01**	**13.44**	**0.61**	-	-	-

Implementation Details: The proposed method is implemented on a NVIDIA GeForce GTX1080Ti GPU in Keras [14]. The adaptive moment estimation optimizer(ADAM) and weight decay are used. The initial learning rate is set to be 0.001, 0.0001 and 0.001 for segmentation module and quality estimation module and refinement module respectively. The coefficient λ is set to be 1 and the number of iterations is 3 in our experiment as the results remain stable when the number of iteration is greater than 3 so we just set it to be 3 for simplicity. In our experiment, K has an important impact on the results. If K is too small, the selected data with pseudo label won't be enough. If K is too large, it will lead to an increase of low-quality data with pseudo labels, so we set K to be 50% for balance. When segmenting NB(BraTS18) dataset, we pre-train f_{QE} on BraTS18(NB) dataset.

3.2 Quantitative and Qualitative Analysis

Metrics: Dice Similarity Coefficient (DSC), Hausdorff Distance(HD) and Average Surface Distance (ASD) are used as the evaluation metrics. For fair comparison, 5-fold cross validation is employed.

We compare with the backbone U-Net [2] and other state-of-the-art semi-supervised segmentation methods TCSM [4], MASSL [5], ASDNet [11]. Some methods are not originally used for binary segmentation and we re-implement all above methods and apply them to our experiment dataset. In Table 1, our methods can outperform other methods at least 3.64%, 4.55% and 3.31% in DSC with 10%, 30%, 50% of labeled data. Besides using quality estimation to guarantee the quality of the pseudo label, the refinement module is reused in the inference to get a better result and Kullback–Leibler divergence is adopted in the loss function which is proved effective in segmentation, so we can achieve a competing performance with only 106 labelled data to a fully supervised model with 344 labelled data.

We further investigate the robustness of our proposed semi-supervised segmentation algorithm on BraTS18 dataset. In Table 2, we achieve a DSC of 78.03%, 80.31% and 80.61% with 20, 50 and 110 labeled data, with obvious margin compared with other methods. Futhermore, we can reach a fully supervised performance with only 50 labeled data.

Table 2. Experiment results on BraTS18 dataset by different methods.

Labeled (unlabeled)	20(200)			50(170)			110(110)			220(0)		
Methods	DSC	HD	ASD	DSC	HD	ASD	DSC	HD	ASD	DSC	HD	ASD
U-Net	70.17	17.99	2.46	72.82	17.77	2.39	74.42	17.59	2.17	80.28	15.38	1.97
MASSL [5]	76.35	19.97	2.82	77.34	17.36	2.44	78.26	16.32	1.88	-	-	-
ASDNet [11]	75.47	17.59	2.75	77.18	18.20	2.32	78.59	15.10	2.21	-	-	-
TCSM [4]	74.25	17.58	**2.01**	78.24	15.34	1.88	79.41	15.86	**1.38**	-	-	-
Ours	**78.03**	**15.47**	2.24	**80.31**	**14.74**	**1.75**	**80.61**	**14.68**	1.59	-	-	-

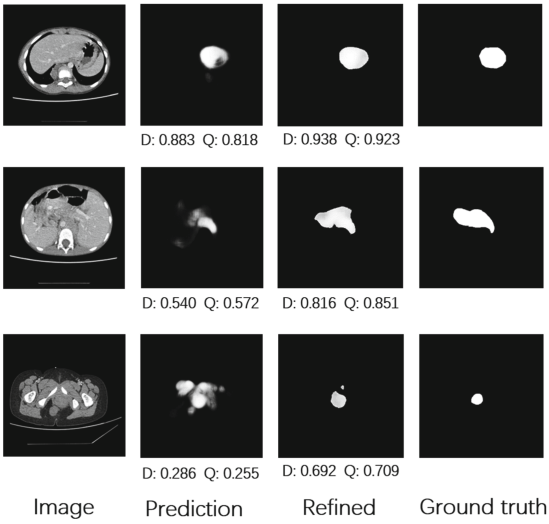

Fig. 3. Visual results of quality estimation and refinement on NB dataset(D: DSC Q: Estimated Quality).

Table 3. Ablation study of our method on NB and BraTS18 dataset with 169 and 110 labeled data. We evaluate the efficiency of each component in our method (Selecting strategy, shape confidence branch, semantic confidence branch and refinement module).

Data	NB	BraTS18
Methods	DSC	DSC
U-Net	71.85	74.42
U-Net+All unlabeled without selection	69.87	72.93
U-Net+random 50% selection	70.56	73.75
U-Net+Shape	75.94	77.69
U-Net+Shape+ Semantic	78.13	79.05
U-Net+Shape+ Semantic+Refinement	**80.01**	**80.61**

Visual results on NB dataset are also presented in Fig. 3, which illustrates the effectiveness of the quality estimation and refinement modules. The second and third column indicate that the refinement module can produce a better prediction with higher DSC and more accurate segmentation. The DSC and Quality under the prediction and refined prediction show the quality estimation module can provide a reliable quality with less than 10% estimation error.

Ablation Study: We analyze the efficiency of each component in our proposed method by performing five ablation studies on NB and BraTS18 dataset as Table 3 shows. First, we examine the effect of using all unlabeled data and randomly selecting 50% of unlabeled data with pseudo labels to expand the training set. From the fourth and fifth row of the table, the DSC even decreases

to 69.87%, 70.56% from 71.85% in NB and 72.93%, 73.75% from 74.42% in BraTS18 for many low-quality pseudo labels have been added to the train set. Then we analyse the performance of the shape confidence branch, semantic confidence branch and refinement module. We can see the DSC increases to 75.94%, 78.13%, 80.01% in NB and 77.69%, 79.05%, 80.61% in BraTS18 respectively. The results show that the quality estimation module has the greatest improvement on the proposed framework.

4 Conclusion

In this paper, we propose a novel semi-supervised segmentation with self-training based on quality estimation and refinement. We select the good segmentation samples to expand the training set by estimating their quality and refine them in an adversarial way by distinguishing the generated pseudo label with the ground truth. Moreover, the refinement network can be reused during inference to obtain more accurate segmentation result. Our method is evaluated on the established neuroblastoma(NB) and BraTS18 dataset and outperforms other state-of-the-art semi-supervised medical image segmentation methods. We can achieve a fully supervised performance while requiring ∼4x less annotation effort.

References

1. Long, J., et al.: Fully convolutional networks for semantic segmentation. In: CVPR, pp. 3431–3440 (2015)
2. 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
3. Baur, C., Albarqouni, S., Navab, N.: Semi-supervised deep learning for fully convolutional networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 311–319. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_36
4. Li, X., et al.: Transformation consistent self-ensembling model for semi-supervised medical image segmentation. arXiv preprint (2019). arXiv:1903.00348
5. Chen, S., Bortsova, G., García-Uceda Juárez, A., van Tulder, G., de Bruijne, M.: Multi-task attention-based semi-supervised learning for medical image segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11766, pp. 457–465. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32248-9_51
6. Peng, J., et al.: Deep co-training for semi-supervised image segmentation. arXiv preprint (2019). arXiv:1903.11233
7. Zhou, Y., Wang, Y., et al.: Semi-supervised multi-organ segmentation via multi-planar co-training. arXiv preprint (2018). arXiv:1804.02586
8. Fu, Y., et al.: More unlabelled data or label more data? a study on semi-supervised laparoscopic image segmentation. In: Wang, Q., et al. (eds.) DART/MIL3ID -2019. LNCS, vol. 11795, pp. 173–180. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33391-1_20

9. Cui, W., et al.: Semi-supervised brain lesion segmentation with an adapted mean teacher model. In: Chung, A.C.S., Gee, J.C., Yushkevich, P.A., Bao, S. (eds.) IPMI 2019. LNCS, vol. 11492, pp. 554–565. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20351-1_43
10. You, X., Peng, Q., Yuan, Y., Cheung, Y.M., Lei, J.: Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. Pattern Recognit. **44**(10–11), 2314–2324 (2011)
11. Nie, D., Gao, Y., Wang, L., Shen, D.: ASDNet: attention based semi-supervised deep networks for medical image segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 370–378. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_43
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint (2014). arXiv:1409.1556
13. Hung, W.-C., et al.: Adversarial learning for semi-supervised semantic segmentation. arXiv preprint (2018). arXiv:1802.07934
14. Keras: Deep learning library for theano and tensorflow (2015). http://keras.io
15. Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 408–416. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_47
16. Radosavovic, I., Dollár, P., Girshick, R., Gkioxari, G., He, K.: Data distillation: towards omni-supervised learning. In: CVPR, pp. 4119–4128 (2018)
17. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE TMI **34**(10), 1993–2024 (2015)