Brain Tumor Segmentation Using an Adversarial Network

Zeju Li[1], Yuanyuan Wang[1,2(✉)], and Jinhua Yu[1,2(✉)]

[1] Department of Electronic Engineering, Fudan University, Shanghai, China
{yywang, jhyu}@fudan.edu.cn
[2] Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention of Shanghai, Shanghai, China

Abstract. Recently, the convolutional neural network (CNN) has been successfully applied to the task of brain tumor segmentation. However, the effectiveness of a CNN-based method is limited by the small receptive field, and the segmentation results don't perform well in the spatial contiguity. Therefore, many attempts have been made to strengthen the spatial contiguity of the network output. In this paper, we proposed an adversarial training approach to train the CNN network. A discriminator network is trained along with a generator network which produces the synthetic segmentation results. The discriminator network is encouraged to discriminate the synthetic labels from the ground truth labels. Adversarial adjustments provided by the discriminator network are fed back to the generator network to help reduce the differences between the synthetic labels and the ground truth labels and reinforce the spatial contiguity with high-order loss terms. The presented method is evaluated on the Brats2017 training dataset. The experiment results demonstrate that the presented method could enhance the spatial contiguity of the segmentation results and improve the segmentation accuracy.

Keywords: Brain tumor segmentation · Adversarial network · Deep learning

1 Introduction

Automatic segmentation of brain tumors in magnetic resonance (MR) images is of great clinical value. Nevertheless, the task is technically challenging because tumor regions vary a lot in the shape and location [1]. Among the existing segmentation methods, the convolutional neural network (CNN) provides very outstanding results and attracts increasing attentions [2]. A CNN obtains stacked features and produces segmentation results by classifying image voxels based on these features. One defect of a CNN-based segmentation method is that the receptive field is always limited by the size of convolutional kernels [3]. In other words, each pixel in the image is predicted almost independently with each other. This problem could become more apparent in brain tumor segmentation because the appearance of brain tumors is unpredictable and MR images are inhomogeneous in the intensity. Therefore, the segmentation results of CNN-based methods always have rough boundaries and perform poorly on details of tumor sub-regions.

A. Crimi et al. (Eds.): BrainLes 2017, LNCS 10670, pp. 123–132, 2018.
https://doi.org/10.1007/978-3-319-75238-9_11

Actually, many efforts have been made to reinforce the spatial contiguity of the segmentation results. One solution is to add another CNN pathway with a larger receptive field [4]. The multiple pathway CNN structure could take advantage of both the visual details and larger context of the region around a certain pixel. However, the long-range connection would make the network more complex and bring a huge increase in the network computation. The conditional Markov random field (CRF) is another common approach to enhance the spatial contiguity of the output labels [5, 6]. A fully connected CRF is usually taken as the post-processing procedure of the CNN and takes the feature maps of the CNN as unary potentials to build the global probabilistic models. By the inference of graphical models, the CRF makes it possible to recover fine details in the output maps. Nevertheless, in practical applications, the global optimal parameters of the CRF are hard to find for the multiple classification problem. Refinement of different tumor sub-regions require different parameters of the CRF model, thus a CRF model with fixed parameters is only applicable to a specific kind of tumor sub-regions [6]. Instead of adding the context information or CRF, we want to strength the spatial contiguity by using an auxiliary high-order loss term, which can make the network more perceptive towards spatial-connected tumor regions. This idea is inspired by the generative adversarial network (GAN) [7]. The GAN, through the use of adversarial loss, has successfully been applied to the generation of real-life images in an unsupervised way [8].

In this paper, we present a novel CNN-based brain tumor segmentation method by using an adversarial network. A fully connected CNN is firstly applied to provide the segmentation results. A discriminator network, also taking advantage of the original images, is designed to discriminate the synthetic labels from the ground truth labels. The adversarial loss provided by the discriminator network encourages the generator network to correct the segmentation mistakes and produce more accurate and space continuous results. The presented method was evaluated on the Brats2017 training dataset. Experiment results demonstrated the effectiveness of adversarial network and showed that the presented method could provide competitive segmentation results.

2 Method

The proposed method consists of two CNN networks, named the generator network and the discriminator network. These two networks are tightly connected. The detailed structure is described in Fig. 1.

The generator network is similar to a regular CNN which is a feedforward neural network architecture. Patches of MR images are set as the input of the generator network. Ground truth labels of smaller patches with the same center are set as the targets. The backpropagation procedure helps the generator network learn to produce segmentation results.

The discriminator network is another separate CNN network. MR images and segmentation results are input to the discriminator network and then concatenated together in the network. The discriminator network is trained to distinguish between the synthetic labels and ground truth labels. The derivatives of the synthetic labels are computed by the backpropagation from the output at the top where the network takes the synthetic labels as real.

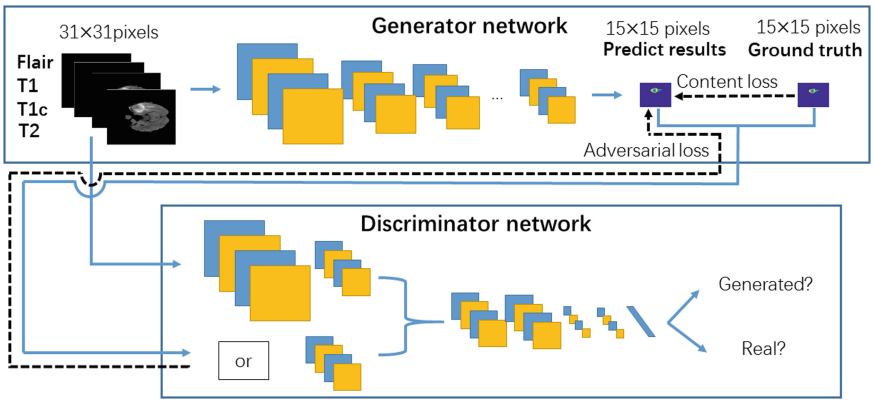

Fig. 1. Schematic diagram of the proposed framework. Forward propagations are depicted using blue lines and backward propagations are depicted using black dotted lines. (Color figure online)

The derivatives of synthetic labels are fed back to the generator network to provide external adjustments. The adversarial loss is added up with the pixel-wise content loss in the output layer of the generator network. These two networks are trained in an iterative way. Specifically, a single batch of images is input to the generator network to create the synthetic labels with fixed network parameters. Then, the synthetic labels together with the ground truth labels are input to the discriminator network. Along with the training process of the discriminator network, the gradient is propagated through all the way to the bottom and forms the adversarial loss. Lastly, the adversarial loss is delivered to the generator network and participates in the training of the generator network. The detailed network structures were demonstrated in Fig. 2. Implement details of the presented method are described in the following paragraphs.

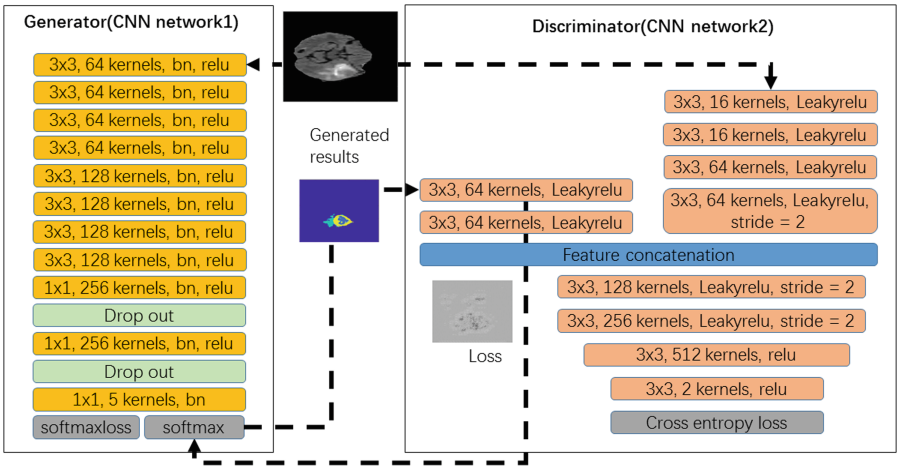

Fig. 2. The network structures of the two CNN networks used in this work.

2.1 Generator Network

Taking into account that the portion of tumor regions is quite small in the brain MR images, the network training process is carried out with image patches, instead of taking the full-size image as the input. The selection of the patches is uneven for the better recognition of tumor regions. Patches with tumor regions are more likely to be selected. During the training phase, image patches I with the size of 31 × 31 pixels are input to the network. The generator network consists of 11 layers. All first eight layers include three stages of convolutional layers with 3 × 3 kernels, relu layers and batch normalization layers. The convolutional layers are built without padding, therefore the output features maps have the size of 15 × 15 pixels ($31 - 2 \times 8$). The last three layers are taken as dense layers, where convolutional kernels with the size of 1 × 1 are utilized. Two dropout layers are added before the last two layers.

The lost function of the generator network consists of two parts, called the content loss and the adversarial loss. Specifically, the loss function is calculated as:

$$L^G = L_{softmax}(G(I), label_{GT}) + L^G_{Ad} \tag{1}$$

The first term $L_{softmax}$, namely the content loss, is the pixel-wise derivatives. The content loss is calculated based on the softmax-loss between the prediction and ground truth of four given labels. I represent the patches of four modalities MR images. $label_{GT}$ represent the given ground truth labels which are corresponding to I. In order to take advantage of more context information, the size of the predicted label is small than the input images. Specifically, $label_{GT}$ are the ground truth labels with the size of 15 × 15 pixels around the same image center of I. The second term, L^G_{Ad} namely the adversarial loss, is produced simultaneously by the adversarial network and is discussed in the next section. It is worth being mentioned that only the generator network is utilized during the test phase. Therefore, the presented method would not increase the complexity of the CNN network.

2.2 Discriminator Network

Adversarial Training. The discriminator network is a core of the presented method. The synthetic labels and the ground truth labels, with the form of the one-hot coding, are input to the discriminator network. The original MR images are also input to the network as references. The discriminator network consists of 10 convolutional layers with the kernel size of 3 × 3 and full padding. Leaky relu layers are utilized but no pooling layers are included in the discriminator network, as suggested by a previous GAN study [9]. Images are down sampled to the same size of the labels using convolutional layers with stride 2. The discriminator network is trained to distinguish two kinds of labels obtained from different places. The loss function is calculated as:

$$L^D = \frac{1}{2}[L_{bce}(I, D(label_{GT}), 0) + L_{bce}(I, D(G(I)), 1)] \tag{2}$$

As usual, the binary cross entropy (bce) loss is calculated to differ two types of inputs. The batch size of the discriminator network is twice the size of the generator network. Therefore, the discriminator network could process both the synthetic labels and ground truth labels in a single batch.

Adversarial Loss. The discriminator network provides the adversarial loss by minimizing the probability that the adversarial predicts $D(label_{GT})$ are similar to the one of synthetic labels. The adversarial loss is calculated as:

$$L_{Ad}^{G} = L_{bce}(I, D(G(I)), 0) \tag{3}$$

The adversarial loss contains high order derivatives and are fed back to the generator network in (1) Gradients from the adversarial network encourage the generator network to correct the mistakes of the segmentation results with high-order terms. The discriminator network could assess the mutual information of many label varies and reinforce the spatial contiguity of the segmentation results with the global information. The adversarial loss is produced by the discriminator network after a long period of training and could not be assessed by the pixel-wise content loss.

The adversarial loss is produced along with the training process of the generator network and the discriminator network. However, the discriminator network with the poor discrimination ability may not provide the effective adjustment for the generator network. The effectiveness of pretrain of the model is evaluated in this study. During the pretrain procedure, the generator network is firstly trained for a while separately, and the discriminator is trained with settled generator network until the network is stable.

2.3 Pre-processing and Post-processing

Before entering the patches to the networks, some pre-processing procedures are taken to make the input data have similar distributions. The intensity of the patches are normalized to from 0 to 1 for each channel, separately. Then the mean values of the patches are subtracted and the variances of the patches are united, separately for each channel. In the post-processing stage, we want to reduce the false positive segmentation results. In particular, the tumor region is always gathered in a certain area. However, the network may mistake other regions as tumor because of the intensity inhomogeneity of MR images. Regions with similar gray intensity but in other brain areas could be likely to be mistaken. Therefore, the largest three-dimensional (3D) connection region of the segmentation results is firstly chosen as the tumor candidate. 3D bounding box slightly larger than the tumor candidate is built, and then recognized region inside the bounding box are finally identified as the tumor region.

3 Experimental Results

All the experiments were developed on the top of Matconvnet. Both the generator network and the discriminator network were trained using Adam optimizer with beta1 = 0.9, beta2 = 0.999 and epsilon = 1e-8. The primary learning rate was 1e-3 and

1e-4 for the generator network and discriminator network, separately. The learning rates were decreased exponentially. The two CNN networks were both initialized using Xavier method. The batch size of the generator network and the discriminator network is 128 and 256 respectively. About 280,000 patches are extracted from the training data. Patch samples inside and around tumor regions are more likely to be taken. Consequently, about 40% of the selected patches contain tumor regions.

3.1 Material

The presented tumor segmentation methods were evaluated on the Brats2017 training dataset [10–12]. The dataset consists of four MR images modalities (T1, T1c, T2 and Flair) of 285 patients. The tumor regions were manually segmented into three sub-regions including the enhancing tumor (label4), the peritumoral edema (label2) and the necrotic and non-enhancing tumor (label1). Whole regions (label1 + label2 + label4), core regions (label1 + label3) and enhancing regions (label4) of the segmentation results were evaluated, separately. The Dice similarity coefficient (DSC), positive predictive value (PPV) and sensitivity (Sen.) were calculated for segmentation results with the corresponding manual segmentation as the ground truth. Calculation details of those indices could be found in the previous study [1].

3.2 Evaluation on the BRATS2017 Training Dataset

The effectiveness of adversarial training. It is interesting for us to observe the effect of the addition of the discriminator network. Thus, two experiments were carried out with exactly the same condition except for the existence of the adversarial network. Qualitative results could be found in Fig. 3. With the help of adversarial loss, the amount of over segmentation is reduced. False positive segmentation results are also decreased. Moreover, the spatial contiguity of segmentation of all tumor sub-regions is enhanced. The segmentation results become smoother in the sub-regions and preserve good boundaries. It is hard to achieve by the post-processing of CRF with certain parameters.

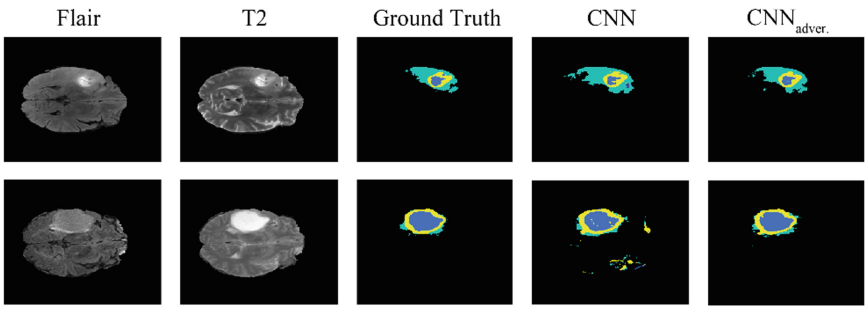

Fig. 3. Visual segmentation results of two examples using the direct outputs of CNN network.

To quantitative exhibit the effectiveness of the adversarial loss, the segmentation results of all 285 training data are summarized in Table 1. To better illustrate the effectiveness of the adversarial loss, the segmentation results were extracted directly from the generator network without any post-processing. We can see that the improvement by the adversarial loss is small but consistent in all sub-regions.

Table 1. Quantitative segmentation results of 285 training data using the direct outputs of a CNN network. $CNN_{adver.}$ corresponds to the network trained with the adversarial loss.

Method	Whole			Core			Enhancing		
	DSC	PPV	Sen.	DSC	PPV	Sen.	DSC	PPV	Sen.
CNN	86.1	82.2	90.0	86.3	85.5	87.2	77.1	74.2	80.1
$CNN_{adver.}$	**87.9**	**86.8**	89.0	**86.8**	**86.6**	87.0	**77.5**	**77.1**	77.9

To better illustrate the effectiveness of the loss from adversarial network, the losses are visualized in Fig. 4. The pixel-wise content loss, which is the most common loss for the neural network, is calculated simply by the derivatives between the feedforward network output and the given ground truth. The pixel-wise content losses directly represent the differences in the given cases but might not be the best for the overall dataset.

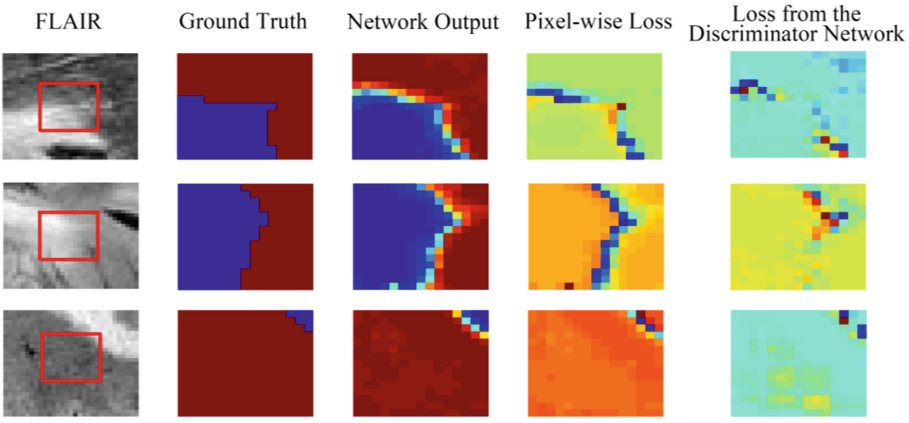

Fig. 4. Comparison between the pixel-wise content loss and the loss from the discriminator network. In the last two columns of image, the more obvious the color change, the greater the value of loss in the corresponding area.

On the other hand, loss from the discriminator network could be more perceptual. The discriminator network is trained by discriminating the synthetic labels from the ground truth labels with the use of plenty of MR image and corresponding label patches. The adversarial network can discover the mutual characteristics of the

mistakes. As we can see from Fig. 4, the adversarial losses do not simply be obtained by the first-order derivative but generated in a high-order form. Actually, the adversarial loss could figure out the more important parts of the mistaken regions, which have the greater impact on the identification of the discriminator network. The adversarial loss would help the output of the generator network visually closer to the ground truth and more likely to fool the discriminator network. This would drive the generator network produce more realistic and space continuous results. The adversarial loss may be not the best option for the single given case but could lead the generator network to pursue the best solutions for the overall dataset. What is more, this kind of high-order loss has a stronger adaptability, contains more global information, and cannot be provided by per-pixel loss. Thus, this kind of loss is more meaningful.

Results of adding adversarial loss with pretrain. As mentioned in the method section, the pretrain of the model seems to be useful. Understandably, the better the discriminator is trained, the more effective the adversarial loss is. The experiment results illustrate this point. As demonstrated in Fig. 5, segmentation results of the pretrain model show more accurate recognition ability, especially in details.

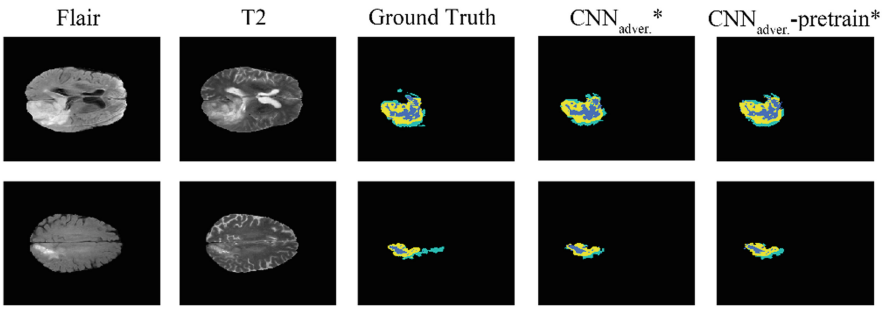

Fig. 5. Visual segmentation results of two examples using a CNN with the adversarial loss.

Similarly, the quantitative segmentation results of all 285 training data are summarized in Table 2. It could be seen that the post-processing procedure could improve the segmentation accuracy. Better segmentation results could be carried out with the pretrain of the model. Compared with the results of previous methods on the previous Brats dataset [1], segmentation results provided by the presented method is competitive.

Table 2. Quantitative segmentation results of 285 training data using a CNN with the adversarial loss. Results with post-processing is marked followed by *.

Method	Whole			Core			Enhancing		
	DSC	PPV	Sen.	DSC	PPV	Sen.	DSC	PPV	Sen.
$CNN_{adver.}$	87.9	86.8	89.0	86.8	86.6	87.0	77.5	77.1	77.9
$CNN_{adver,}$*	89.5	90.6	88.4	88.0	89.0	87.0	78.4	79.1	77.8
$CNN_{adver.}$-pretrain*	**89.7**	89.8	88.5	**88.4**	**89.7**	**87.2**	**79.1**	78.1	**80.0**

3.3 Cross Validation on the Brats2017 Training Dataset

The presented method was also evaluated on a separate dataset. The proposed model was trained once again using 250 cases randomly selected from the Brats2017 training dataset. The rest 35 cases were set as separate validation dataset. Quantitative results are summarized in Table 3. The experiment results also demonstrate the effectiveness of the adversarial network.

Table 3. Quantitative results of separate validation data from Brats2017 training dataset.

Method	Whole			Core			Enhancing		
	DSC	PPV	Sen.	DSC	PPV	Sen.	DSC	PPV	Sen.
CNN*	85.8	86.9	84.6	72.3	76.8	68.3	67.7	71.3	67.5
$CNN_{adver.}$*	**87.0**	86.7	**87.2**	72.0	71.4	**72.5**	**68.2**	68.3	**68.1**

4 Conclusion

In this study, a novel CNN-based tumor segmentation method with adversarial network is presented. A high-order adversarial loss provided simultaneously by the discriminator network is added to encourage the generator network to produce more precise results. Experiments on the Brats2017 training dataset demonstrate that the presented methods could enhance the spatial contiguity of the segmentation results in all tumor sub-regions and improve segmentation accuracy. Moreover, the effectiveness of pre-train of the model is demonstrated in this study. Compared with segmentation results of previous researches on the previous Brats dataset, our method could provide competitive segmentation results. It should be mentioned that the presented method is also efficient and energy-saving since the complexity is not added to the network at test time. In the future, we will extent the presented method into 3D to better make use of the 3D information of MR images. Further, we will exploit the best way to take advantage of the adversarial loss.

Acknowledgments. This work was supported by the National Basic Research Program of China (2015CB755500), the National Natural Science Foundation of China (11474071).

References

1. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015)
2. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in MR images. IEEE Trans. on Med. Imaging **35**(5), 1240–1251 (2016)
3. Zhao, X., Wu, Y., Song, G., Li, Z., Fan, Y., Zhang, Y.: Brain tumor segmentation using a fully convolutional neural network with conditional random fields. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Winzeck, S., Handels, H. (eds.) BrainLes 2016. LNCS, vol. 10154, pp. 77–80. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-55524-9_8

4. Havaei, M., Davy, A., Wardefarley, D., Biard, A., Courville, A., Bengio, Y., et al.: Brain tumor segmentation with deep neural networks. Med. Image Anal. **35**, 18–31 (2017)
5. Li, Z., Wang, Y., Yu, J., Shi, Z., Guo, Y., Chen, L., et al.: Low grade glioma segmentation based on CNN with fully connected CRF. J. Healthc. Eng. **2017** (2017). https://doi.org/10.1155/2017/9283480. Article No. 9283480
6. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., et al.: Efficient multi-scale 3D CNN with fully connected crf for accurate brain lesion segmentation. Med. Image Anal. **36**, 61–78 (2017)
7. Luc, P., Couprie, C., Chintala, S., Verbeek, J.: Semantic segmentation using adversarial networks. arXiv:1611.08408 (2016)
8. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al.: Generative adversarial networks. In: NIPS, pp. 2672–2680 (2014)
9. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434 (2015)
10. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., et al.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data **4** (2017). https://doi.org/10.1038/sdata.2017.117. Article No.170117
11. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., at al.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017)
12. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., at al.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017)