Learning Data Augmentation for Brain Tumor Segmentation with Coarse-to-Fine Generative Adversarial Networks

Tony C. W. Mok[✉] and Albert C. S. Chung

Lo Kwee-Seong Medical Image Analysis Laboratory,
Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology, Kowloon, Hong Kong
cwmokab@connect.ust.hk, achung@cse.ust.hk

Abstract. There is a common belief that the successful training of deep neural networks requires many annotated training samples, which are often expensive and difficult to obtain especially in the biomedical imaging field. While it is often easy for researchers to use data augmentation to expand the size of training sets, constructing and generating generic augmented data that is able to teach the network the desired invariance and robustness properties using traditional data augmentation techniques is challenging in practice. In this paper, we propose a novel automatic data augmentation method that uses generative adversarial networks to learn augmentations that enable machine learning based method to learn the available annotated samples more efficiently. The architecture consists of a coarse-to-fine generator to capture the manifold of the training sets and generate generic augmented data. In our experiments, we show the efficacy of our approach on a Magnetic Resonance Imaging (MRI) image, achieving improvements of 3.5% Dice coefficient on the BRATS15 Challenge dataset as compared to traditional augmentation approaches. Also, our proposed method successfully boosts a common segmentation network to reach the state-of-the-art performance on the BRATS15 Challenge.

1 Introduction

Accurate segmentation of a brain tumor from medical images is a crucial step for clinical diagnosis, evaluation, and follow-up treatment. Currently, the automatic segmentation methods which achieve state-of-the-art results are often using a deep learning approach. Modern deep learning models often consist of millions of parameters and learning these parameters requires massive annotated datasets to avoid overfitting to the training set. However, the problem is made challenging by the number of annotated training datasets often being limited in the medical imaging domain due to a couple of reasons. First, it is time-consuming and expensive for experts to accurately delineate the pixel-wise brain tumor region. Second, manual labeling also suffers from considerable intra-rater and inter-rater inconsistencies [5]. Third, there are various modalities and imaging protocols,

therefore a training set generated for one study is difficult to transfer to another study in practice.

To address these problems, we propose an automatic data augmentation approach for network-based brain tumor segmentation. Specifically, we present and evaluate a method for augmenting multimodal brain MRI images of high-grade (HG) and low-grade (LG) glioma patients, in which the generic augmented data enable the network-based method to learn the available annotated datasets more efficiently. Experimental results demonstrate that the proposed method effectively improves the segmentation accuracy of the network-based method, compared to the traditional data augmentation approach. It achieves improvements of 3.5% dice coefficient on the BRATS15 Challenge dataset as compared to traditional augmentation approaches.

2   Related Work

Data Augmentation. Data augmentation is essential to teach the network the desired invariance and robustness properties when only a few training samples are available. For medical image segmentation, different combinations of affine transformations are commonly used as data augmentation to teach the network the desired invariance and robustness properties. Ronneberger et al. [7] applied shift, rotation and elastic deformations to the microscopical images during training, while Milletari et al. [6] applied the random deformation to prostate MRI volumes using dense deformation field with B-spline interpolation. For brain tumor segmentation, scaling, rotation and flipping have also been applied to multimodal brain MR images for data augmentation [9]. Typical data augmentation approaches fail to increase the diversity of the training data, i.e., different parameters for MR imaging protocol, tumor size, shape, location, and appearance. The contribution of this work is that we have developed an automatic way to learn a more generic augmentation so that not only the rotational and scaling invariance, high-level information such as the shape of tumor and contextual information can also be augmented.

Generative Adversarial Networks. In the domain of computer vision, Generative adversarial networks (GANs) [2] have elicited considerable attention. GANs aim to model the data distribution by forcing the generated sample to be indistinguishable from the data. They have also proven successful in a wide variety of applications such as image generation [1,8], image manipulation [13] and image inpainting [11]. Recently, various coarse-to-fine frameworks of GANs have been proposed [4,10] to generate high-quality and high-resolution images, e.g., 1024×1024 pixels. Inspired by their successes, we propose a new coarse-to-fine boundary-aware GANs suitable to generate generic augmented MR images for brain tumor segmentation.

3   Methods

3.1   Preliminaries of Generative Adversarial Networks

Typical Generative Adversarial Networks (GANs), comprise a generator G and a discriminator D that are trained to compete with each other alternatively. The generator G is optimized to generate the data distribution p_{data} by generating the images that are indistinguishable for the discriminator D to differentiate from real images. While D is optimized to distinguish real images and synthetic images generated by G. The training objective is similar to a two-player min-max game as follows:

$$\min_{G} \max_{D} \mathcal{L}_{GAN}(D, G) = \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log\left(1 - D(G(z))\right)]. \quad (1)$$

where x is a real sample from the target data distribution p_{data}, and z is a noise vector sampled from distribution p_z.
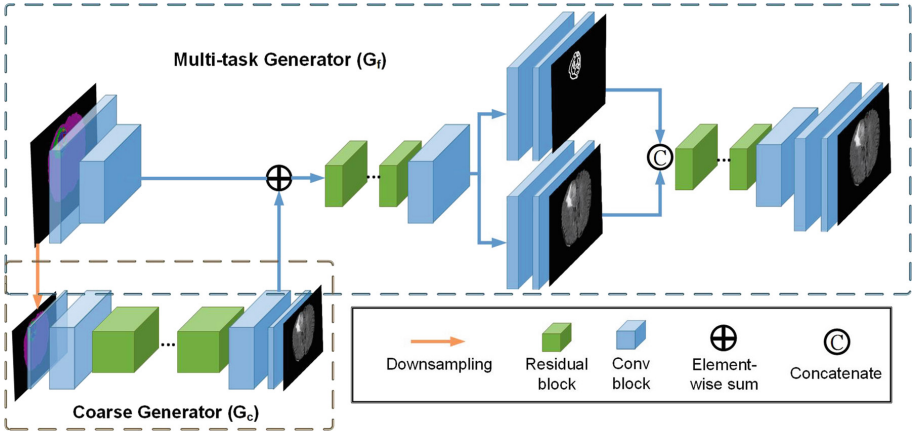

Fig. 1. Network architecture of proposed generator.

3.2   Coarse-to-Fine Boundary-Aware Generator

To generate high-resolution MR images of brains with realistic detail, we propose a Coarse-to-fine Boundary-aware Generative Adversarial Networks (CB-GANs). In our proposed method, the noise vector in traditional GANs is replaced by a label map of 2D axial slice from a 3D MR volume as a conditioning variable. We explain how we diversify the generated data without using a noise vector as input in Sect. 3.4.

Coarse-to-Fine Generator. Our generator is decomposed into two different sub-generators: G_c and G_f. G_c is the coarse generator while G_f is the fine generator. The generator G is then given by the tuple $G = \{G_c, G_f\}$. The coarse generator G_c aims to sketch the primitive shape and texture of multimodal brain MR images from a label map at a lower resolution and the fine generator G_f aims to correct the defects and completes the details of the low-resolution MR images from the coarse generator G_c.

The G_c and G_f consist of three components, namely, a convolutional down-sampling block, a set of residual blocks, and a transposed convolutional block. The resolution of the input label map to G_f is the same as the training data, while the resolution of the input label map to G_c is 4× smaller than the training data (2x smaller along each axis). Different from the residual block in coarse generator G_c, the residual block in fine generator G_f takes the element-wise sum of the output of G_c and the input feature maps of previous layers from G_f as the input. The element-wise sum operation helps integrate the global and local information from G_c and G_f.

Boundary-Aware Generator. Although the above coarse-to-fine framework can already produce high resolution natural images, it remains a challenge to produce a high quality synthetic MR image of a brain tumor that serves the purpose of data augmentation, given the corresponding label map. Because the size of the tumor core in MR images is often small compared to the other encephalic regions, in which the networks may fail to notice that the details of the tumor core are important. Preserving accurate tumor boundaries is important for augmented data to teach the network the desired invariance and robustness properties. To address this problem, we propose a multi-task generator G_f to replace the original fine generator.

The structure of the proposed generator is illustrated in Fig. 1. Instead of treating the image generation task as a single problem, we formulate it as a multi-task problem by exploring auxiliary information, which can simultaneously infer the location and boundary of the complete tumor. Specifically, two different branches are added to the final layer of G_f in order to output the MR image of a brain with a tumor and the boundaries of the complete tumor. After that, the outputs from the two new branches are concatenated and fed into a residual block followed by a non-linear activation layer. Therefore, the boundary and texture information from the new branches are fused together to output the final image. The mean-square-error loss is used for the boundary extraction task, as shown in the following:

$$\mathcal{L}_b(x, y) = \frac{1}{n_i} \sum_n \sum_i \left(P(x_{n,i}; \theta) - y_{n,i} \right)^2, \tag{2}$$

where θ is the weight parameters in the generator. $\mathcal{L}_b$ refers to the mean-square-error loss for the boundary extraction task. $x_{n,i}$ and $y_{n,i}$ are the i-th pixel and ground truth in the n-th image used for training, respectively. P refers to the predicted probability for the pixel $x_{n,i}$.
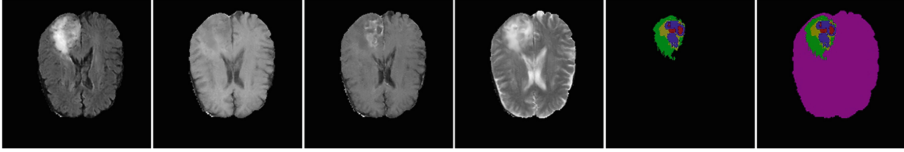
Fig. 2. Example of a synthetic high-grade glioma tumor. Left to right: FLAIR, T1, T1c, T2, expert delineation (Ground truth) and semantic label map (Red: necrosis, Green: edema, Yellow: non-enhancing tumor, Blue: enhancing tumor). (Color figure online)

3.3   Adversarial Training

Multi-discriminators. As the resolution of the synthetic image increases, the difficulty for the discriminator to differentiate real and synthetic images also increase. When there is only a single discriminator, the discriminator needs to have a large receptive field that is able to capture both global, i.e., tumor location, and local, i.e., tumor texture and shape, information from the input image. However, this may not be a good idea as it implies we will need a discriminator which has either a deep network or large convolutional kernels. Both options require a large memory and may easily suffer from overfitting to the training data.

To address this challenge, we adopt multi-discriminators with different scales of input as our discriminator $D = \{D_1, D_2, D_3, D_4\}$. The four discriminators have identical architectures but operate at different image scales, which is similar to [10]. Specifically, real MR images and synthesized MR images are downsampled by factors 2, 4 and 8 using the bilinear interpolation to create input for D of 4 scales. Throughout the experiments, we find that using four discriminators can achieve the optimal performance and further increasing the number of discriminators cannot improve the quality of synthetic image.

Perceptual Loss. We further improve the GAN's loss by incorporating a modified perceptual loss. The main idea of the perceptual loss function is that if the synthetic image is similar to the real image, the ith-layer feature maps of the discriminator should be also similar when the synthetic and real images pass through it. The modified perceptual loss $\mathcal{L}_P(G, D_k)$ is calculated as:

$$\mathcal{L}_P(G, D_k) = \mathbb{E}_{(x,c)} \sum_{i=1}^{L} \frac{1}{N_i} [||D_k^{(i)}(x, c) - D_k^{(i)}(G(z), z)||_2^2], \tag{3}$$

where $D_k^{(i)}$ represents the ith-layer feature maps of discriminator D_k, L is the total number of layers, N_i denotes the number of elements in each layer, x denotes the real MRI image, c denotes the label map and z denotes the deformed label map.
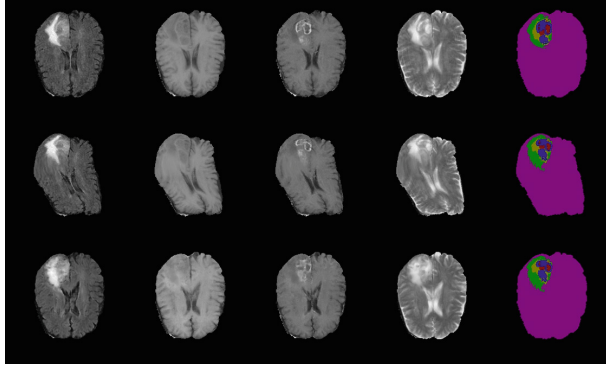
Fig. 3. Comparison of traditional augmentation and our proposed method. First row: Original image. Second row: Augmented image using elastic deformation. Third row: synthetic image generated by our approach. Left to right: FLAIR, T1, T1c, T2 and semantic label map

Therefore, our full training objective combines both GANs loss and modified perceptual loss as:

$$\min_G \Big(\Big(\max_{D_{k \in \{1,2,3,4\}}} \sum_{k=1,2,3,4} \mathcal{L}_{GAN}(D_k, G) \Big) + \lambda_1 \mathcal{L}_b(x, y) + \lambda_2 \sum_{k=1,2,3,4} \mathcal{L}_P(G, D_k) \Big). \tag{4}$$

3.4 Diversity of Augmented Data

Using Deformed Semantic Label Maps. The traditional augmentation approach for object segmentation often uses different combinations of affine transformations, such as shifting, rotation, and zoom, to leverage the knowledge of invariances in a task. However, such knowledge implied by these affine transformations is limited. For example, the shape, location and appearance of a complete tumor in a multi-modal MR image can vary significantly in the testing data, but the augmented image produced by the typical data augmentation fails to "simulate" such changes. Although some interpolation-based techniques such as elastic deformation can cause a slight variation in the shape of the augmented image, it may bring about damage and noise to the training data, as shown in Fig. 3, if the deformation field varies a lot.

Instead, we propose applying the elastic deformation to the label map. After that, we create a set of semantic labels from the deformed label maps. Specifically, we label 1 to 5 for necrosis, edema, non-enhancing tumor, enhancing tumor, non-tumor brain regions and 0 for everything else in the semantic labels. We use the semantic labels instead of the label maps as input for our proposed CB-GANs. By providing the information of the contour of the brain to the generator, it further diversifies the synthetic brain MR image with different shapes and prevents model collapse, i.e., prevents the model from generating a set of

realistic MR images with the identical shape and context of the brain. Figures 2 and 3 show the example of the synthetic image generated by CB-GANs with corresponding semantic label map.

4   Experiments and Results

4.1   Data and Pre-processing

Experiments have been performed using brain MRI sequences from BRATS15 datasets. The dimensions of each MR volume are $240 \times 240 \times 155$ pixels. BRATS15 provides both the training and test sets. The training set consists of 220 high-grade glioma (HGG) and 54 low-grade glioma (LGG) cases. For each case, it includes 4 modalities (Flair, T1, T1-contrast (T1c) and T2) that were skull-stripped and co-registered. Pixel-wise ground truths that annotate the complete tumor, which are verified by radiologists, are provided in the training set. For the testing set, it consists of 110 cases each with 4 modalities. Unlike the training set, the ground truth labels of the test cases are hidden from the public and evaluation is carried out via an online system. Extensive evaluation has been carried out on three tasks: (1) the complete tumor (necrosis, edema, non-enhancing and enhancing tumor) (2) the tumor core (necrosis, non-enhancing and enhancing tumor) (3) enhancing tumor region. For each MR image, we normalize the intensities of each modality to have zero-mean and unit variance.

4.2   Network Architectures

Generator Architectures. For generator networks, we adopt our backbone architectures from Wang et al. [10] with some modifications. Below, we follow the naming convention used in the Wang et al. [10]. Let $c7s1$-k denote a 7×7 Convolution-BatchNorm-ReLU layer with k filters and stride 1. dk denotes a 3×3 Convolution-BatchNorm-ReLU layer with k filters and stride 2. Rk denotes a residual block that contains two 3×3 convolutional layers with the same number of filters on both layers. uk denotes a 3×3 fractional-strided-Convolution-BatchNorm-ReLU layer with k filters and stride $\frac{1}{2}$. Note that we will replace the activation layer from ReLU to Tanh for the final layer of each generator.

Our coarse generator G_c:
$c7s1$-64, d128, d256, d512, d1024, R1024, R1024, R1024, R1024, u512, u256, u128, u64, $c7s1$-4

Our fine generator G_f:
$c7s1$-32, d64, R64, R64, R64, u32, $c7s1$-4, $concat\{c7s1$-2, $c7s1$-4$\}$, R64, R64, $c7s1$-4

Discriminator Architectures. For discriminator networks, we use 4 Convolution-BatchNorm-LeakyReLU blocks for each discriminator. Let Ck denote 4×4 Convolution-BatchNorm-LeakyReLU blocks with k filters and stride

2. At the last layer, we add a sigmoid activation layer at the end to produce a 1-dimensional output. We use leaky ReLUs with default slope 0.2. All our four discriminators share the identical architecture as follows:
$C64, C128, C256, C512$

4.3   Network Configuration and Training

During the experiments, we employ two sets of convolution neural networks (CNNs). The first set of CNNs is the proposed CB-GANs as shown in Fig. 1, which is used for generic data augmentation. While the second set of CNNs is the U-Net [7], which is used for the segmentation task. We first trained the CB-GANs with back-propagation using the Adam optimizer with initial learning rate 0.0002 and momentum 0.5 for both generators and discriminators. We use CB-GANs to augment the training data during the training phase of U-Net. U-Net is trained with the same learning rate as CB-GANs. All the network are trained from scratch. The method is implemented using Pytorch.

In terms of computation time, it takes about 4 days to train the CB-GANs and 20 h to train the segmentation network for 100 epochs on a Nvidia GTX1080 Ti GPU. Moreover, we define the typical augmentation to be a combination of rotation (-10 to $10°$), zoom (0.98x to 1.02x) and random horizontal flip (50%) that apply to the training data.

4.4   Evaluation

We validate our approach by using it to augment the annotated training sets for the segmentation tasks and show that we have achieved strong gains, in terms of the Dice overlap metric between the automated segmentation and the radiologist annotation label map, over traditional augmentation baselines. We randomly split the training set in BRATS15 into two subsets, resulting in 234 training and 40 validation multimodal volumes. The full test set in BRATS15 is used as our test set, which includes 110 patients. First, we conduct the component testing on the test set to evaluate the impact of coarse-to-fine framework and proposed boundary loss function. Table 1 compares the segmentation performance between a baseline GANs, a coarse-to-fine GANs and the proposed CB-GANs. It shows that if the coarse-to-fine framework and boundary loss function were added, there is an improvement in Dice values for the tumor core task, giving an average 3.6% improvement in Dice. This is probably because the coarse-to-fine framework GANs and boundary loss function can correct defects and generate a clear boundary for small tumor regions in synthetic images.

We also compare the performance of the proposed method to the traditional augmentation method as listed in Table 2. Both data augmentation methods are able to improve the segmentation performance by a significant Dice value. Our proposed method further improves the performance over traditional data augmentation methods on average by 3.5% of Dice values and achieves a significant improvement in Dice for the complete tumor task.

Table 1. Segmentation performance on the BRATS15 testing set. GANs: proposed architecture without coarse-to-fine framework and boundary loss function. C-GANs: coarse-to-fine GANs. CB-GANs: our proposed method.

Method	Dice			Precision			Sensitivity		
	Complete	Core	Enh.	Complete	Core	Enh.	Complete	Core	Enh.
GANs	0.80	0.58	0.55	0.84	0.80	0.62	0.80	0.55	0.51
C-GANs	0.82	0.60	0.55	**0.87**	0.80	**0.66**	0.81	0.55	0.52
CB-GANs (ours)	**0.84**	**0.63**	**0.57**	**0.87**	**0.82**	0.65	**0.84**	**0.57**	**0.54**

Table 2. Performance on the BRATS15 testing set. w/o DA: without any data augmentation. w/ DA: with typical data augmentation. w/ Proposed: with proposed generic data augmentation method.

Method	Dice			Precision			Sensitivity		
	Complete	Core	Enh.	Complete	Core	Enh.	Complete	Core	Enh.
w/o DA	0.79	0.54	0.43	0.85	0.79	**0.66**	0.78	0.47	0.37
w/ DA	0.81	0.61	0.55	0.85	**0.82**	0.64	0.80	0.54	**0.54**
w/ Proposed	**0.84**	**0.63**	**0.57**	**0.87**	**0.82**	0.65	**0.84**	**0.57**	**0.54**

Table 3. Comparison to the state-of-the-art results on the BRATS15 testing set.

Method	Dice			Precision			Sensitivity		
	Complete	Core	Enh.	Complete	Core	Enh.	Complete	Core	Enh.
Kamnitsasa17 [3]	**0.85**	0.67	**0.63**	0.85	**0.84**	0.63	**0.87**	0.60	0.66
Zhao17 [12]	0.84	**0.73**	0.62	**0.89**	0.76	0.63	0.82	**0.76**	**0.67**
2D U-net w/ proposed	0.84	0.63	0.57	0.87	0.82	**0.65**	0.84	0.57	0.54

Finally, we compare our proposed method with two state-of-the-art methods as listed in Table 3. Kamnitsasa et al. [3] method, achieving a top ranking in both BRATS15 and ISLES15 Challenge, using a dual pathway deep 3D CNNs to segment the tumor region and 3D fully connected Conditional Random Field to reduce the false positive, while Zhao [12] joins three segmentation models which uses 2D image patches from different views. Our results are competitive with both methods and give better result for the enhancing tumor task in terms of Dice precision.

Also, one advantage of our proposed model is its relatively low computational cost in both the training and testing phases as we only leverage simple 2D CNNs with no post-processing method. Kamnitsasa et al. [3] reported a running time of 3 min using a 3 GB GPU to segment one case, while 6–12 min was reported by

Zhao [12]. With our proposed method, we achieve 2.1 s for one case in inference time as the architecture of U-Net has much fewer learning parameters.

5   Conclusion

In this paper, we propose a novel, automatic and network-based data augmentation method for brain tumor MR image segmentation. The main contribution is that we propose a generic way to augment training data that is able to teach network-based methods the desired invariance and robustness properties for segmentation tasks. We have shown that the proposed coarse-to-fine framework and boundary loss function in GANs lead to improved augmented data and segmentation quality. We have also shown that our method can boost a common segmentation network to reach the state-of-the-art multi-scale deep networks' performance with the relatively low computational cost at inference time and outperforms the traditional augmentation methods.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International Conference on Machine Learning, pp. 214–223 (2017)
2. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
3. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. MedIA **36**, 61–78 (2017)
4. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196 (2017)
5. Menze, B., Jakab, A., Bauer, S., et al.: The multimodal brain tumorimage segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015)
6. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
7. 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
8. Rosca, M., Lakshminarayanan, B., Warde-Farley, D., Mohamed, S.: Variational approaches for auto-encoding generative adversarial networks. arXiv preprint arXiv:1706.04987 (2017)
9. Shen, H., Wang, R., Zhang, J., McKenna, S.J.: Boundary-aware fully convolutional network for brain tumor segmentation. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 433–441. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_49
10. Wang, T.C., Liu, M.Y., et al.: High-resolution image synthesis and semantic manipulation with conditional GANs. arXiv preprint arXiv:1711.11585 (2017)

11. Yeh, R.A., Chen, C., Lim, T.Y., et al.: Semantic image inpainting with deep generative models. In: CVPR, pp. 5485–5493 (2017)
12. Zhao, X., Wu, Y., Song, G., et al.: A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. MedIA **43**, 98–111 (2018). https://doi.org/10.1016/j.media.2017.10.002
13. Zhu, J.-Y., Krähenbühl, P., Shechtman, E., Efros, A.A.: Generative visual manipulation on the natural image manifold. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9909, pp. 597–613. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46454-1_36