Semi-Supervised Attention-Guided CycleGAN for Data Augmentation on Medical Images

Zhenghua Xu[1,2,⋆], Chang Qi[1,2,⋆], Guizhi Xu[1,2,†]
[1]State Key Laboratory of Reliability and Intelligence of Electrical Equipment,
Hebei University of Technology, China
[2]Key Laboratory of Electromagnetic Field and Electrical Apparatus Reliability of Hebei Province,
Hebei University of Technology, China
⋆Co-first authors, contributed to this work equally
†Corresponding author, email: gzxu@hebut.edu.cn

***Abstract*—Recently, deep learning methods, in particular, convolutional neural networks (CNNs), have made a massive breakthrough in computer vision. And a big amount of annotated data is the essential cornerstone to reach this success. However, in the medical domain, it is usually difficult (and sometimes even impossible) to get sufficient data for some specific learning tasks. Consequently, in this work, a novel data augmentation solution, called semi-supervised attention-guided CycleGAN (SSA-CycleGAN) is proposed to resolve this problem. Specifically, a cycle-consistency GANs-based model is first proposed to generate synthetic tumor (resp., normal) images from normal (resp., tumor) images. Then, a semi-supervised attention module is further proposed to enhance the model's capability in learning the important details of the training images, which in turns help the generated synthetic images become more realistic. To verify its effectiveness, experimental studies are conducted on three medical image datasets with limited amounts of MRI images, and the proposed SSA-CycleGAN is applied to generate synthetic tumor and normal MRI images for data augmentation. Experimental results show that i) SSA-CycleGAN can add (resp., remove) tumor lesions on (resp., from) the original normal (resp., tumor) images and generate very realistic synthetic tumor (resp. normal) images; and ii) in the ResNet18-based MRI image classification tasks based on these datasets, data augmentation using SSA-CycleGAN achieves much better classification performances than the classic data augmentation methods.**

***Index Terms*—Generative adversarial networks (GANs), Data Augmentation, Attention module, Medical image processing**

I. Introduction

In recent years, there are great progresses in the field of Generative Adversarial Networks (GANs) [1] and its extensions. The wide and successful applications of GANs in image generation tasks [2]–[4] have attracted growing interests across many communities, including medical imaging. GANs is a minimax game between two Neural Networks, where a generator is used to generate samples, and a discriminator identifies the source of samples. In this game, the adversarial loss brought by the discriminator provides a clever way to capture high dimensional and complex distributions, which imposes higher-order consistency that is proven to be useful in many cases, such as domain adaptation, data augmentation, and image-to-image translation.

It is widely known that a big amount of annotated data is important to reach the successes of deep learning based computer vision tasks. Data that are very imbalanced on classes or with poor diversity usually lead to bad model performances. This often proves to be problematic in the field of medical imaging, where abnormal findings are, as literally shown, uncommon. Although traditional data augmentation schemes (e.g., crop, rotation, flip, and translation) can mitigate some of these issues, the augmented images usually have similar distributions to the original images, leading to limited performance improvements. Also, the diversity that the augmented images can bring is relatively small. Motivated by GANs, researchers try to add synthetic samples to the training process for data augmentation. GANs-based data augmentation can improve performance by introducing additional distribution information that uncovered by original images. Consequently, it can generate new and realistic medical images, and help achieve outstanding performance in medical image analysis [5]–[7].

Despite these efforts, some specific GANs-based data augmentation tasks in medical imaging, e.g., converting a normal image without tumor to a tumor image, are still challenging. Essentially, this is an image generation problem, but unlike the style translation task, e.g., synthesizing PET image from CT scans [8] or MRI [9], due to the requirement of only modifying some image features while keeping others unchanged, the attribute manipulation in these tasks is a challenging problem. One solution is to translate medical images to another domain firstly, and capture tumor feature leading by tumor annotations in the another domain [10], and then merge the tumor feature to normal images with randomly selected place and size in the another domain, finally, the merged images in the another domain is translated to medical image domain. This method convert the attribute manipulate task to style transform task. The other solution is to generate tumor masks manually, and then generate tumor in the area under the tumor mask. However, these solutions define the position and size of tumors either manually or randomly, which may break image prior, and cause unacceptable additional false positives in the following image processing tasks.

To overcome the above problem, in this paper, we propose a novel GAN-based data augmentation model guided by semi-supervised attention mechanisms. Inspired by CycleGAN [11], the proposed data augmentation model com-

prises of two generators and two discriminators. The generator $G_{N\to T} : Normal \to Tumor$ and the inverse generator $G_{T\to N} : Tumor \to Normal$ introduce two cycle-consistence losses $\mathcal{L}_{cycle}(N, G_{T\to N}(G_{N\to T}(N)))$, $\mathcal{L}_{cycle}(T, G_{N\to T}(G_{T\to N}(T)))$ to add constraint to the training procedures of mapping functions $G_{T\to N}$ and $G_{N\to T}$, because they are highly under-constrained. Then, attention modules are introduced to locate the place where the generators should pay more attention to translate. In order to locate the place to translate more accurately, the attention module inside the generator $G_{T\to N}$ is trained by both adversarial loss (unsupervised) and pixel-wise loss (supervised), so we call it semi-supervised attention mechanisms. Moreover, we add spectral normalization [12] to stabilize the training of discriminators.

We evaluate the generated images' realism by observing the quality of synthetic images. The results showed that our model could generate more realistic images compared to baseline models. And we evaluate the performance of the proposed SSA-CycleGAN data augmentation model by conducting tumor classification experiments with/without synthetic images. The results showed that our model can boost 9.78% sensitivity and 5.97% specificity compare to classic data augmentation methods.

The contributions of this paper are summarized as follows:

- **SSA-CycleGAN model for medical image generating:** We propose a novel semi-supervised attention-guided CycleGAN to add (resp., remove) tumor lesions on (resp., from) the original normal (resp., tumor) images and generate very realistic synthetic tumor (resp. normal) images. We are the first that generate tumor lesions in normal images without human interupt, so that the image prior will not be demage.
- **Synthetic medical image for data augmentation:** In the ResNet18-based medical image classification task, models trained on datasets combines synthetic medical images with origin images achieve much better classification performances than models trained on origin datasets with classic data augmentation methods.

II. Related Work

Generative Adversarial Networks(GANs) [1] have two models to train: a model G to learn the target data distribution $p_{\text{data}}(\boldsymbol{x})$, a model D to assess the source of D's input, is it from $p_{data}(x)$ or from model $G(\boldsymbol{z})$. The aim of the training model G is to maximize the chance of model D making mistakes, while the aim of training model D is to maximize the probability of assigning the correct label to both training examples and samples from G. The adversarial loss is the key to GAN's success. It forces the model to generate images that indistinguishable from real images.

Attention-guided GANs can generate more realistic images since the learned attention enforce the generator pay more attention to the place that lead to a realistic image generation. Mejjati *et al.* [13] proposed an attention mechanism that is jointly trained with the generators and discriminators. Chen *et al.* proposed AttentionGAN [14], an extra attention network is used to generate attention maps. Kastaniotis *et al.* [15] proposed ATAGAN, a teacher network is used to produce attention maps. Zhang *et al.* [16] proposed a Self-Attention Generative Adversarial Networks (SAGAN), the Non-Local Module [17] was used to produce the attention map. Liang *et al.* [18] proposed a Contrasting GAN that takes the segmentation mask as the attention map. Sun *et al.* [19] proposed an attention GANs using FCN to generate a facial mask for face attribute manipulation.

To the best of our knowledge, there are two research groups tried to generate brain tumor images. Shin et al. [10] tried to duplicate tumors from BRATS dataset to normal MRI images from ADNI dataset. They proposed a two-stage model for the translation task. The first is an GANs-based image-to-brain model to generate brain masks with white matter, grey matter, and CSF. The second stage is a GANs-based brain-to-image model. The brain masks generated in the stage one was merged with tumor masks. The merged brain masks as input to the brain-to-image model generate abnormal MRI images with brain tumors. However, there are some limits in this two-stages model. Firstly, there are no brain mask annotations in the BRATS dataset. Thus the tumor mask that merged to brain mask is inferred by the model trained on ADNI dataset. But the ADNI dataset does not contain tumor information, so the quality of the generated tumor mask is doubtful. Secondly, the position of the tumor merged to brain mask was decided randomly. However, the tumor location is related to other feature of the tumor, such as size, shape, degree of malignancy, the attempt that randomly locate a wrong tumor location that unpaired to the specific tumor features may damage the image prior, causing higher false positives and less robustness.

Han et al. [20] proposed a CPG-GAN model for generating tumors from noise. The 'condition' in their model is a $[0, 1]$ mask, where 0 stands for the normal area, 1 stands for the area that needs to generate a tumor. However, there are some problems in their work: Firstly, like shin et al.'s work [10], the position of the tumor is decided randomly, which may damage the image prior. The experiment results in their work prove it: the FPs per slice increases 3.52 with only 0.1 increase of sensitivity in the detection task. Secondly, the adversarial loss is not enough to generate a realistic tumor image. The synthetic tumor images are in poor image quality.

III. Our approach

The goal of our work is to generate tumor (resp., normal) images from normal (resp., tumor) images by introducing additional distribution information uncovered by original images. As illustrated in Fig. 1, our data augmentation method includes two mapping function $G1 : Normal \to Tumor$ and $G2 : Tumor \to Normal$. Meanwhile, there are two adversarial discriminators D1: distinguish between real tumor images $\{T_r\}$ and generated tumor images from real normal images $\{T_g^r\}$, and D2: distinguish between real normal images $\{N_r\}$ and generated normal images from real tumor images $\{N_g^r\}$.

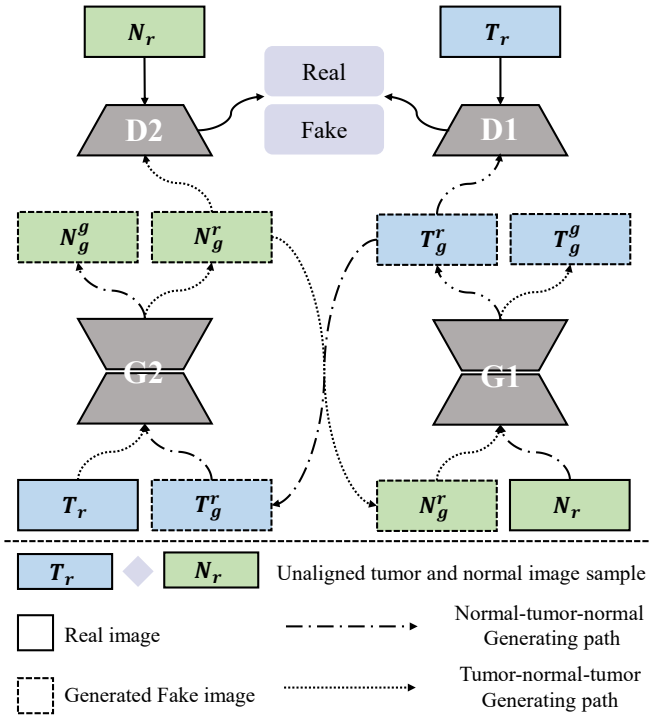

Fig. 1. The framework of the proposed data augmentation model. 'T' represents for Tumor samples, 'N' represents for Normal samples. G1/G2: The Attention-guided generator, D1/D2: The Attention-guided discriminator.

A. Attention-Guided Generator

Unlike style translation tasks in domain translation tasks, translation between normal images to tumor images require to solve two tasks: 1) location the area to translate, and 2) taking the proper translation in the located area. So we proposed two attention networks A_N and A_T to achieve this. Where A_N aims to select the area to generate tumor that maximizes the probability that the discriminator makes a mistake and minimizes the probability that the generator makes a mistake; A_T aims to locate the place that has tumor and generates the possibility map, which will guide the generator recover the normal images from tumor images.

In the forward processing, the generated image contains two parts, the foreground from the generator and the background from the input image. Take the translation from normal samples to tumor samples as an expmple. Firstly, the normal brain MRI $\{N_r\} \in N$ is fed into the generator $G_{N\to T}$, which maps $\{N_r\}$ to the target domain T, resulting the generated tumor image $G_{T'} = G_{N\to T}(\{N_r\})$. Then, the same input $\{N_r\}$ is fed into the attention module A_N, resulting in the attention map $M_{N'} = A_N(\{N_r\})$. To create the 'foreground' object $\{T'_f\} \in T$, we apply $M_{N'}$ to $G_{T'}$ via an element-wise product: $\{N'_f\} = M_{N'} \odot G_{T'}$. Secondly, the inverse of attention map $M'_{N'} = 1 - M_{N'}$ will be applied to the input image via an element-wise product as the background. Thus, the mapped image$\{T_g\}$ is obtained by:

$$T_g = \underbrace{M_{N'} \odot G_{T'}}_{\text{Foreground}} + \underbrace{M'_{N'} \odot N_r}_{\text{Background}}. \quad (1)$$

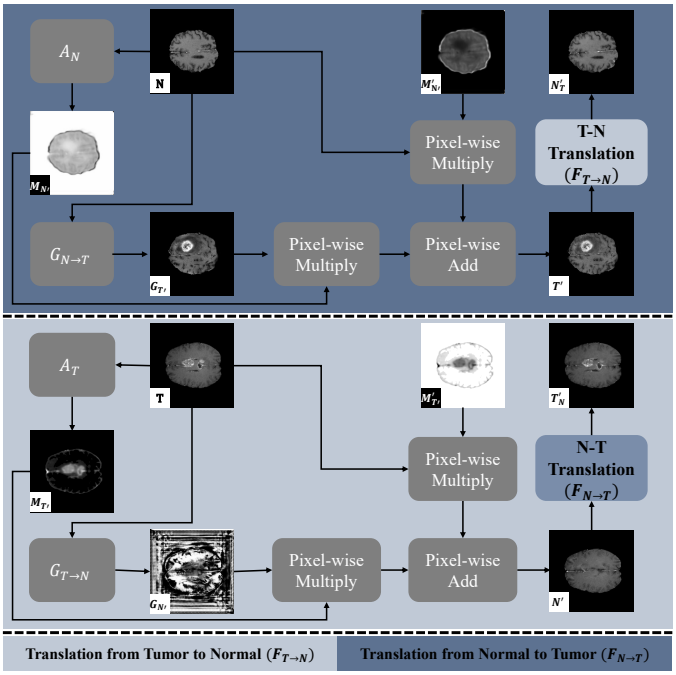

Fig. 2. The framework of the proposed attention-guided generator. The attention-guided generators have built-in attention mechanism, which can detect the most discriminative part of images. Then, we mix the input image and the attention mask to synthesize the targeted image.

We only described the mapping $F_{N\to T}$; the inverse mapping $F_{T\to N}$ is defined similarly. Fig. 2 visualizes those processes.

B. Attention-Guided Discriminator

Eq. 1 constrains the generators to modify only on the attention regions, however, the discriminatorsconsider the whole image currently. Vanilla discriminator D_T takes the whole generated image $\{T_g\}$ and the whole real image $\{T_r\} \in T$ as input and tries to distinguish them. We add attention mechanism to discriminators so that the discriminators only consider the regions inside the attention map. The attention-guided discriminator takes the attention mask, the generated images, and the real images as inputs. For example, attention-guided discriminator D_T^A tries to distinguish between synthetic tumor images with attention maps $M_{N'} \odot T_g$ and real tumor images with attention map $M_{N'} \odot T_r$. Similar to D_T^A, D_N^A tries to distinguish between the synthetic normal images with attention maps $M_{T'} \odot N_g$ and real normal images with attention maps $M_{T'} \odot N_r$. Discriminators can focus on the most discriminative content by this attention-guided method.

C. Spectral Normalization

It is wildly known that GANs is challenging to train because of the objective function of the vanilla GANs is equivalent to the J-S divergence between the distribution p_g of the generated data and the distribution p_r of the real data. However, J-S metric fails to provide a meaningful value when two distributions are disjoint. It makes no guarantee convergence to a unique solution such that $p_r = p_g$. Then, WGAN [3] was proposed to replace the J-S divergence in the vanilla

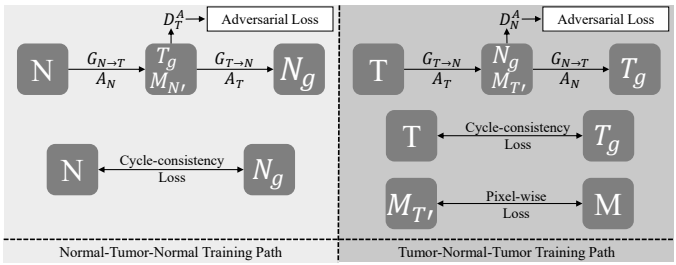

Fig. 3. The training loss of the proposed GAN-based network.

GANs with Wasserstein distance. The KR duality principle is used to transform the Wasserstein distance problem into a solution to the optimal Lipschitz continuous function. The spectral normalization proposed by Miyato et al. [12] use a more elegant way to make the discriminator meet Lipschitz continuity.

D. Semi-supervised Attention Mechanism

Tumor segmentation annotations are available in our case. And the attention map in tumor $\rightarrow$ normal translation is exactly the whole tumor region. Therefore, we supervise the training process of attention network A_T by segmentation label. Given a training set $\{(T_1, M_1), \cdots, (T_N, M_N)\}$ of N examples, where M_i stands for the tumor annotation of segmentation. To reduce changes and constrain generators, we adopt pixel-wise loss between the tumor annotation M_i and the generated attention map $M_{T_i'}$. We express this loss as:

$$\mathcal{L}_M(M_i, M_{T_i'}) = \|M_i - M_{T_i'}\|_1. \tag{2}$$

This added loss makes our model more robust by encouraging the attention maps to be sharp (converging towards a binary map), while the attention mask of normal areas will always be zero.

E. Optimization Objective

Fig. 3 shows the training loss of the proposed data augmentation model.

Attention-guided Adversarial Loss. The attention-guided adversarial loss is proposed to training the attention-guided discriminators. It can be formulated as follows:

$$\begin{aligned}\mathcal{L}^N_{AGAN}(G_{N\to T}, D^A_T) = \mathbb{E}_{t\sim p_{\text{data}}(t)}\left[\log D^A_T(M_{N'}\odot t)\right] + \\ \mathbb{E}_{n\sim p_{\text{data}}(n)}[\log(1 - D^A_T(M_{N'}\odot G_{N\to T}(n)))].\end{aligned} \tag{3}$$

$G_{N\to T}$ aims to translation the normal image to tumor image and maximize the probability that the discriminator makes a mistake. D^A_T is trained to distinguish between the generated image with its attention mask $M_{N'}\odot G_{N\to T}(n)$ and the real image $(M_{N'}\odot t)$. Which means $G_{N\to T}$ tries to minimize the attention-guided adversarial loss $\mathcal{L}_{GAN}(G_{N\to T}, D^A_T)$, while D^A_T tries to maximize it. There is an another loss for the discriminator D^A_N and the generator $G_{T\to N}$:

$$\begin{aligned}\mathcal{L}^T_{AGAN}(G_{T\to N}, D^A_N) = \mathbb{E}_{n\sim p_{\text{data}}(n)}\left[\log D^A_N(M_{T'}\odot n)\right] + \\ \mathbb{E}_{t\sim p_{\text{data}}(t)}[\log(1 - D^A_N(M_{T'}\odot G_{T\to N}(t)))].\end{aligned} \tag{4}$$

Cycle-Consistency Loss. The cycle-consistency loss can be used to enforce forward and backward consistency. For example, if a tumor image translates to a normal image, the translation from the synthetic normal image to the tumor image should be brought back to a cycle. Thus, the loss function of cycle-consistency is defined as:

$$\begin{aligned}&\mathcal{L}_{cycle}(G_{N\to T}, G_{T\to N}) = \\ &\mathbb{E}_{n\sim p_{\text{data}}(n)}[\|G_{T\to N}(G_{N\to T}(n)) - n\|_1] + \\ &\mathbb{E}_{t\sim p_{\text{data}}(t)}[\|G_{N\to T}(G_{T\to N}(t)) - t\|_1].\end{aligned} \tag{5}$$

Loss Function. We obtain the final loss function by combining the adversarial loss, cycle-consistency loss, and semi-supervised pixel losses for both source and target domains:

$$\begin{aligned}&\mathcal{L}(G_{N\to T}, G_{T\to N}, A_N, A_T, D^A_N, D^A_T) = \\ &\lambda_{gan}(\mathcal{L}^N_{AGAN} + \mathcal{L}^T_{AGAN}) + \\ &\lambda_{cyc}\times\mathcal{L}_{cycle}(G_{N\to T}, G_{T\to N}) + \mathcal{L}_M(M_i, M_{T_i'}).\end{aligned} \tag{6}$$

IV. Experiments and Results

We present sets of experiments and results in this section. To evaluate the performance of proposed CycleGAN-based data augmentation method, we employed a convolutional neural network with the deep residual block (ResNet18) [21] to compare the classification results using generated tumor images to the classification results of real images. We implemented five models to generate tumor images, as described in Section IV.A.b.

For the implementation of the tumor classification model ResNet18 and GAN-based data augmentation architecture, we used the Pytorch framework. All training processes were performed in an NVIDIA GeForce GTX 1080 Ti GPU.

A. Dataset Evaluation and Implementation Datails

1) Classification: For brain tumor classification, we chose ResNet18 [21] because it is an small model with limited number of parameters, making it potentially more portable and less prone to overfitting. In this study, we split the datasets into 70% training, 20% validation, and 10% test images.

We calculated sensitivity and specificity to meature the performance of our data augmentation method in tumor classification task. In the following equations, we present these measures:

$$Sensitivity/Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

where T stands for correct classification, in the opposite, F stands for the wrong classification. P stands for the

TABLE I
GAN-BASED DATA AUGMENTATION PREFERENCE IN CLASSIFICATION

Dataset	Measures	Real Images (With classic DA methods)	Ours	CycleGAN	Spec-CycleGAN	Attention-CycleGAN	SpecAtte-CycleGAN
Brats19	Sensitivity	67.66	**77.44**	69.15	72.69	68.07	72.82
	Specificity	72.69	**78.66**	77.17	77.98	73.91	**78.66**
Brats18	Sensitivity	69.83	**78.38**	71.60	75.27	71.32	75.27
	Specificity	73.77	**80.16**	75.81	79.07	76.36	77.17
Brats15	Sensitivity	69.45	**80.93**	70.60	72.69	71.73	74.04
	Specificity	71.32	**80.43**	78.38	77.52	76.63	76.90

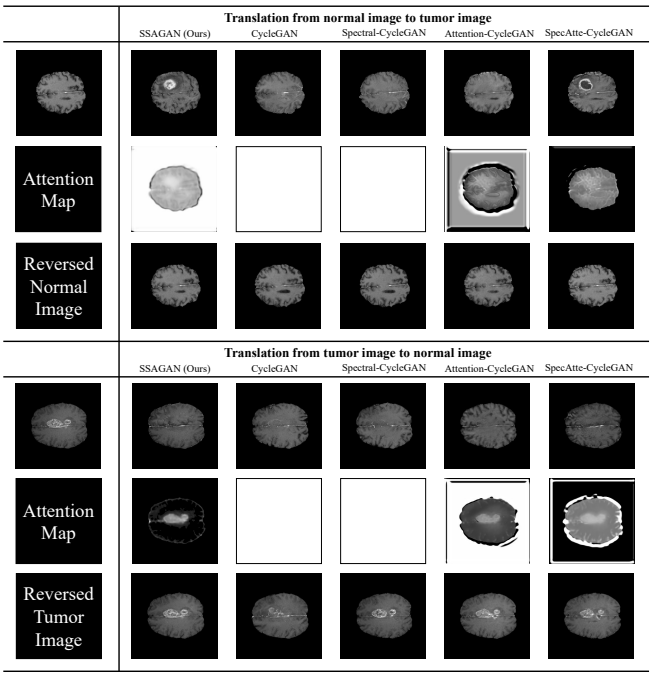

Fig. 4. Results on normal-tumor translation and the inverse tumor-normal translation. Top row shows the input normal image (upper left) and the generated tumor images by difference models. The second row shows the attention mask generated by upper model. The third row shows the reverse normal images generated by feeding the generated tumor images to the tumor-normal translation.

TABLE II
CLASSIFICATION RESULTS OF DIFFERENT RATIO BETWEEN REAL IMAGES AND SYNTHETIC IMAGES

Dataset	Composition	Sensitivity	Specificity
Brats19	130 real cases	67.66	72.69
	+65 synthetic cases	71.32	72.14
	+130 synthetic cases	77.44	**78.66**
	+260 synthetic cases	**77.52**	77.44
	+390 synthetic cases	74.86	75.27
Brats18	105 real cases	69.83	73.77
	+53 synthetic cases	72.47	76.94
	+105 synthetic cases	**78.38**	**80.16**
	+210 synthetic cases	78.09	77.80
	+315 synthetic cases	77.52	76.36
Brats15	110 real cases	69.45	71.32
	+55 synthetic cases	73.16	80.16
	+110 synthetic cases	**80.93**	**80.43**
	+220 synthetic cases	76.08	77.44
	+330 synthetic cases	76.36	75.81

classification result is tumor category, and N stands for the classification result is normal category. So, TP stands for the tumor image is classification to the tumor category, TN stands for the normal image is classification to the normal category, FP stands for the normal image is classification to the tumor category, and FN stands for the tumor image classification to the normal category.

2) Baselines: We compare our model with leading image to image translation model: CycleGAN [11], and the extension of CycleGAN: Attention-guided CycleGAN [13]. For a fair comparison, we then add the spectral normalization method to those models.

3) Datasets: We use the Brats15, Brats18, and Brats19 datasets provided by Menze et al. [22] to evaluate our data augmentation method. These datasets contain a segmentation mask for each case, and for each case, there are four modal MRI images: T1, T1c, T2, and Flair. We choose T1c modal MRI images as the input of both the data augmentation task and classification task because T1c modal MRI images can grade the tumor stage and determine the region of the tumor. Which means that the T1c modal MRI images contain more information about tumor compare to other modal images.

Brats19 dataset has 259 cases of 3D brain MRI images with glioma. To test the performance of the proposed data augmentation model under the situation that the training data is limited, we randomly select half cases as the training data, and then randomly select 10% from another half cases as the testing data. We did the same to Brats18 and Brats15 datasets.

B. Evaluation of the Data Augmentation

1) Qualitative results: Fig. 4 illustrates examples of synthetic images by our data augmentation method. Observing the generated images and learned attention maps by our model, we can see that our model successfully captures the T1c-specific texture and tumor appearance in the right position.

2) Quantitative results: Table. I and Table. II show the performance of our data augmentation method in the im-

age classification task. The results in Table. I proved our hypothesis that adding generated samples can improve the classification performance. The results in Table. II showed the best ratio of real samples and generated samples in the classification task. The results showed that the best ratio is 1:1.

V. Conclusions

This work focus on generating tumor images from normal images and recovering normal images from tumor images with SSA-CycleGAN. This CycleGAN-based data augmentation method can enlarge small medical datasets, fulfill data distribution. While recent GAN-based data augmentation methods in medical image can generate abnormal sample, they also have some limits. For example, previous works need masks to lead the generator to generate tumors in the proposed places, also, it is hard for GANs-based model to generate a large-size medical image. Most generated abnormal samples are small ($32px \times 32px$). The data augmentation method we proposed can generate abnormal images of a real medical image size (In this case, $240px \times 240px$). We expect to get significant improvements in the quality of generated abnormal images by incorporating an attention module into both generator and discriminator. However, then we found that the attention mapping is not robust since the shape of abnormal lesion is changing between images. So we add a semi-supervised mechanism to stabilize this training procedure. The result shows that this approach improves the robust of attention modules. Experimental results on three datasets demonstrate that our data augmentation method can generate striking results with convincing details than the state-of-the-art models.

There are several limitations to this work. One possible extension could be the evaluation from the classification task to the segmentation task. Data insufficient happens more in the field of tumor/lesion segmentation. In the future, we plan to extend our work to other medical domains that can benefit from generated abnormal images to improve training performance.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under grants 61906063 and 51737003, in part by the Natural Science Foundation of Tianjin, China, under grant 19JCQNJC00400, and in part by the Yuanguang Scholar Fund of Hebei University of Technology, China.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[2] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[3] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the International Conference on Machine Learning*, 2017, pp. 214–223.

[4] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[5] S. K. Lim, Y. Loo, N.-T. Tran, N.-M. Cheung, G. Roig, and Y. Elovici, "Doping: Generative data augmentation for unsupervised anomaly detection with gan," in *Proceedings of the IEEE International Conference on Data Mining*, 2018, pp. 1122–1127.

[6] C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. Gunn, A. Hammers, D. A. Dickie, M. V. Hernández, J. Wardlaw, and D. Rueckert, "Gan augmentation: augmenting training data using generative adversarial networks," *arXiv preprint arXiv:1810.10863*, 2018.

[7] A. Madani, M. Moradi, A. Karargyris, and T. Syeda-Mahmood, "Chest x-ray generation and data augmentation for cardiovascular abnormality classification," in *Proceedings of the Medical Imaging 2018: Image Processing*, vol. 10574, 2018, p. 105741M.

[8] A. Ben-Cohen, E. Klang, S. P. Raskin, M. M. Amitai, and H. Greenspan, "Virtual pet images from ct data using deep convolutional networks: initial results," in *Proceedings of the International Workshop on Simulation and Synthesis in Medical Imaging*, 2017, pp. 49–57.

[9] Y. Pan, M. Liu, C. Lian, T. Zhou, Y. Xia, and D. Shen, "Synthesizing missing pet from mri with cycle-consistent generative adversarial networks for alzheimer's disease diagnosis," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 455–463.

[10] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. Michalski, "Medical image synthesis for data augmentation and anonymization using generative adversarial networks," in *Proceedings of the International Workshop on Simulation and Synthesis in Medical Imaging*, 2018, pp. 1–11.

[11] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.

[12] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," *arXiv preprint arXiv:1802.05957*, 2018.

[13] Y. A. Mejjati, C. Richardt, J. Tompkin, D. Cosker, and K. I. Kim, "Unsupervised attention-guided image-to-image translation," in *Advances in Neural Information Processing Systems*, 2018, pp. 3693–3703.

[14] X. Chen, C. Xu, X. Yang, and D. Tao, "Attention-gan for object transfiguration in wild images," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 164–180.

[15] D. Kastaniotis, I. Ntinou, D. Tsourounis, G. Economou, and S. Fotopoulos, "Attention-aware generative adversarial networks (ata-gans)," in *Proceedings of the IEEE Workshop on Image, Video, and Multidimensional Signal Processing*, 2018, pp. 1–5.

[16] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," *arXiv preprint arXiv:1805.08318*, 2018.

[17] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7794–7803.

[18] X. Liang, H. Zhang, and E. P. Xing, "Generative semantic manipulation with contrasting gan," *arXiv preprint arXiv:1708.00315*, 2017.

[19] R. Sun, C. Huang, J. Shi, and L. Ma, "Mask-aware photorealistic face attribute manipulation," *arXiv preprint arXiv:1804.08882*, 2018.

[20] C. Han, K. Murao, T. Noguchi, Y. Kawata, F. Uchiyama, L. Rundo, H. Nakayama, and S. Satoh, "Learning more with less: conditional pggan-based data augmentation for brain metastases detection using highly-rough annotation on mr images," *arXiv preprint arXiv:1902.09856*, 2019.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[22] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.