Research Article

Study on Optimal Generative Network for Synthesizing Brain Tumor-Segmented MR Images

Hyunhee Lee ¹, Jaechoon Jo ², and Heuseok Lim ¹

¹Department of Computer Science and Engineering, Korea University, Seoul 02841, Republic of Korea

²Division of Computer Engineering, College of Information Technology, Hanshin University, Osan-si 18101, Republic of Korea

Correspondence should be addressed to Heuseok Lim; limhseok@korea.ac.kr

Received 24 February 2020; Accepted 13 April 2020; Published 20 May 2020

Guest Editor: Sanghyuk Lee

Copyright © 2020 Hyunhee Lee et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to institutional and privacy issues, medical imaging researches are confronted with serious data scarcity. Image synthesis using generative adversarial networks provides a generic solution to the lack of medical imaging data. We synthesize high-quality brain tumor-segmented MR images, which consists of two tasks: synthesis and segmentation. We performed experiments with two different generative networks, the first using the ResNet model, which has significant advantages of style transfer, and the second, the U-Net model, one of the most powerful models for segmentation. We compare the performance of each model and propose a more robust model for synthesizing brain tumor-segmented MR images. Although ResNet produced better-quality images than did U-Net for the same samples, it used a great deal of memory and took much longer to train. U-Net, meanwhile, segmented the brain tumors more accurately than did ResNet.

1. Introduction

General characteristics of medical imaging data are as follows. It is difficult to obtain a large volume of data, and it is more difficult to acquire labelled data necessary for supervised learning. As shown in Figure 1, since the picture archiving and communicating system (PACS) was introduced in hospitals, vast amounts of multimedia data in the medical imaging field have been stored. However, due to various institutional and privacy issues, external institutions have difficulty gaining access to such data. Additionally, utilizing the accumulated data for learning requires preprocessing the data, which consequentially takes considerable time and effort.

In addition, medical imaging data have the following characteristics. A typical chest X-ray image contains 2,000 pixels horizontally and 2,500 vertically, which results in a total of five million pixels. Meanwhile, the lesions usually occupy a relatively small part of the whole image. Magnetic resonance imaging (MRI) scans provide more detailed information about the inner organs such as the brain, skeletal system, and other organ systems than do computerized

tomography (CT) scans. Although MRI has many advantages, it has some disadvantages such as prolonged acquisition time (about 45 min), high costs, and limiting patient factors such as claustrophobia or metal devices in their bodies [3]. Because MRI scanners use strong magnetic fields and magnetic field gradients, MRI scans could be dangerous especially for a patient with nonremovable metal inside the body [4], and therefore the acquired images could be blurred and abnormal. CT scans are combinations of X-ray images taken from different angles, and they are fast, painless, and noninvasive. However, they do expose the patient to radiation, albeit at a relatively low dose. To minimize radiation exposure, CT scans produce low-dose images, and as a result, they unfortunately tend to be severely degraded by excessive noise and streak artifacts. For these reasons, the medical imaging field contains a small number of available medical data.

Generative adversarial networks (GANs) [5] provide a generic solution to the lack of medical imaging data. As shown in Figure 2, they can be applied to diverse tasks such as image synthesis [7, 9], segmentation [6, 11, 12], reconstruction, and classification [8]. Figure 3 shows the statistics

FIGURE 1: The architecture of PACS [1]. It is a medical imaging technology that provides economical storage and convenient access to images from multiple modalities such as MR, CT, and X-ray [2].

FIGURE 2: Example applications using GANs. (a) Organ (lung and heart) segmentation on chest X-ray of adult [6]. (b) Input MR image, synthesized CT image, and real CT image [7]. (c) Randomly generated skin lesions from random noise (a mixture of malignant and benign) [8]. (d) Ki67 synthetic image by using the segmentation/annotation [9]. (e) Generated retinal fundus images and vessel maps [10].

of GAN-related papers categorized by tasks and imaging modalities. These statistics are based on the databases of PubMed, arXiv, proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), SPIE Medical Imaging, IEEE International Symposium on Biomedical Imaging (ISBI), and International conference on Medical Imaging with Deep Learning (MIDL) [13]. The cutoff date of the search was set as July 30, 2018, and the number of GAN-related papers

increased significantly in 2017 and 2018. As Figure 3 shows, about 70% of these papers studies on image synthesis and segmentation, and MR is the most-studied imaging modality in the GAN-related publications.

In this paper, we use a CycleGAN [14] which has the significant performance in medical imaging so as to synthesize brain tumor-segmented MR images. Generating brain tumor-segmented MR images consists of two tasks, namely, synthesis and segmentation. One entails image-to-image

FIGURE 3: GAN-related publications published on or before July 30, 2018, categorized according to (a) canonical tasks and (b) imaging modalities [13].

translation, which synthesizes a novel image by combining the content of one image with the style of another. The other involves locating and marking lesions on the images. To perform these two tasks, we conduct two experiments using two different generative networks. In the first experiment in this paper, we use a ResNet [15] model, which has significant advantages in style transfer. In the second, we use a U-Net [16] model, one of the most commonly used segmentation techniques. We compare the performance of each model and propose a more robust model for synthesizing brain tumor-segmented MR images, which could lead to high-quality multimedia data augmentation in the medical imaging field.

2. Related Work

In a study on brain CT image synthesis from MR images by Wolterink et al. [7], training using unpaired images was even better than using paired images [13]. That is, the performance of CycleGAN is greater than that of Pix2pix [17] in cross-modality synthesis of medical images. In addition, as shown in Table 1, the use of CycleGAN in the recent publications for medical image synthesis is increasing. In this section, we discuss CycleGAN, ResNet, and U-Net used in our experiments.

2.1. CycleGAN. The problem that Pix2pix or CycleGAN has to solve is the interpretation of images from one domain to another. In Pix2pix, there must be data pairs corresponding to both domains, whereas CycleGAN can solve this problem without these data pairs [14]. CycleGAN is a GAN using two generators and two discriminators. We call one generator G , and it converts images from the X domain to the Y . The other generator is called F , and it converts images from the Y domain to the X . Each generator has a corresponding discriminator that attempts to tell apart its synthesized images from real ones. CycleGAN is not just a simple mapping technique. It considers the returning mapping and puts a constraint on coming back

TABLE 1: Published works on cross-modality image synthesis.

Modality	Publication	Method	Remarks
CT $\rightarrow$ MR	Jin et al. [18]	CycleGAN	Brain
MR $\leftrightarrow$ CT	Zang et al. [19]	CycleGAN*	(3D) Cardiac
MR $\leftrightarrow$ CT	Hiasa et al. [20]	CycleGAN*	Musculoskeletal
MR $\leftrightarrow$ CT	Wolterink et al. [7]	CycleGAN	Brain
T1 $\leftrightarrow$ T2 MR	Dar et al. [21]	CycleGAN	Brain
T1 $\leftrightarrow$ T2 MR	Welander et al. [22]	CycleGAN	Brain

In the third column, *following the method denotes some modifications either on the architecture of the network or on the employed losses.

to its original state. As shown in Figure 4, not only mapping from X to Y but also mapping back to original X from Y should be defined, and this applies to the opposite mapping as well. The reason for doing this is X and Y are unpaired domains. When X goes to Y , it is checked that it looks like Y , and the actual constraint is to keep the original shape when it returns to X again. That is, the shape of X does not change much, only its style is changed to that of Y , and therefore it looks as if only the style has been transferred.

There are two components in the CycleGAN objective function, an adversarial loss and a cycle-consistency loss. Both are essential for successful results. The adversarial loss alone is not sufficient to produce high-quality images, which leaves the model underconstrained [14]. In other words, it forces the generated output to be of the appropriate domain but does not force the input and output to be recognizably the same. The cycle-consistency loss addresses this underconstrained problem. The full objective function by putting these loss terms together and by weighting the cycle-consistency loss with a hyper parameter λ is defined as equation (1). Pix2pix can rely heavily on L1 loss; therefore, adversarial loss may play a supplementary role and it may be better to subtract it. In contrast, the CycleGAN does not learn at all except for the adversarial loss.

$$\mathcal{L}_{\text{full}} = \mathcal{L}_{\text{adv}} + \lambda \mathcal{L}_{\text{cyc}}. \quad (1)$$

FIGURE 4: CycleGAN with mapping functions $G: X \rightarrow Y$ and $F: Y \rightarrow X$, and associated adversarial discriminators D_Y and D_X . Forward cycle-consistency loss (red): $X \rightarrow G(X) \rightarrow F(G(X)) \approx \hat{X}$, and backward cycle-consistency loss (blue): $Y \rightarrow F(Y) \rightarrow G(F(Y)) \approx \hat{Y}$.

2.2. ResNet. The general CNN network in Figure 5(a) receives the input x and yields the output $H(x)$ through the two weighted layers. This output is the input to the next layer. Figure 5(b) shows the architecture of the ResNet, and it uses a *shortcut connection* that connects the input of the layer directly to the output [15]. It is simple network, but its performance is significantly high. As can be seen in Figure 5(b), the output $H(x)$ has been changed to $F(x) + x$, and consequently $H(x) = F(x) + x$ is derived as $F(x) = H(x) - x$. Learning $F(x)$ can be seen as learning residual, and therefore it is called ResNet. It adds the results from the weight layer and the previous results and uses ReLU. ResNet learns in the direction that $F(x)$ becomes zero. In addition, since x is directly connected to *shortcut connection*, there is no increase in computation, and it is possible to select which layer to include. For example, a fully connected layer as well as a convolution layer can be added [15].

2.3. U-Net. Segmentation is the process of partitioning an image into different meaningful segments [23]. In medical imaging, these segments often correspond to different tissue classes, organs, pathologies, or other biologically relevant structures [24]. In the past, there were few medical images, and therefore experts could segment images directly. However, more needs for the automation of the segmentation have arisen as the volume of the medical images has increased exponentially. Analysing medical images can often be difficult and time consuming, and therefore deep neural networks can help doctors make more rapid and more accurate diagnoses [25].

U-Net is one of the most preferred models when segmenting images. Figure 6(a) shows the general encoding and decoding process, and Figure 6(b) shows a U-Net model with a *skip connection* added to the encoder-decoder structure. If the image size is reduced (down sampling) and then reraised (up sampling), sophisticated pixel information disappears. This is a big problem for image segmentation which requires dense prediction on a pixel-by-pixel basis. The *skip connection*, which passes important information directly from the encoder to the decoder, results in a much clearer image at the decoder section, allowing for more accurate prediction.

3. Materials and Methods

3.1. Materials. The first CycleGAN used the U-Net model. The advantage of *skip connection* is that it has much more detail, but the disadvantage is that the performance is not good when the two contents are similar. On the other hand, the last CycleGAN used ResNet model, which is good for image quality but has the disadvantage of using a great deal of memory.

In this paper, we synthesize high-quality brain tumor-segmented MR images. It consists of two tasks, namely, synthesis and segmentation, and therefore we conduct experiments from two perspectives. One is to perform image-to-image translation, which synthesizes a novel image with the style of another image. The other is to locate and mark tumors in brain MR image. Therefore, we perform two experiments with two different generative networks. In the first experiment, we use a ResNet model, which has significant advantages in style transfer. In the second experiment, we use a U-Net model, which is one of the most commonly used segmentation techniques. In this paper, we compare the performance of each model and propose a more powerful model for synthesizing brain tumor-segmented MR images.

Figure 7 shows the datasets of source and target domains used in our experiments. Figure 7(a) represents the brain lesion images in the source domain, and Figure 7(b) is the segmentation mask images of the brain lesion in the target domain. We used a training set of 765 unaligned image pairs and a test set of 92 unaligned image pairs in our experiments. The size of each image is on average 30 to 50 KB. The environments for our experiments are as follows: Ubuntu 16.04.4 LTS for OS, GeForce GTX 1080Ti and CUDA Toolkit for high-performance GPU-accelerated application, and TensorFlow library for deep learning framework.

3.2. Architecture of Our Discriminative Model. The configuration of the discriminative model in our experiments is shown in Figure 8. It consists of four convolution layers, and we use leaky ReLU as an activation function for each layer. In the first step, we extract the features from the image, and in the last, we decide which specific category these features belong to. For that, we add a final convolution layer that

FIGURE 5: The architecture of the networks: (a) CNN and (b) ResNet. The output in CNN is the input of the next layer. The input in ResNet connects directly to the output of the layer [15].

FIGURE 6: The architecture of the networks: (a) Encoder-decoder and (b) U-Net. U-Net is a model with a *skip connection* added to the encoder-decoder structure [17].

FIGURE 7: Our unpaired datasets: (a) brain tumor; (b) segmentation mask.

produces a one-dimensional output. Both ResNet and U-Net generative models use this model as a discriminator in our experiments.

3.3. Architecture of Our Generative Model Using ResNet. The generator has the job of taking an input image and performing the transformation to produce the target image. The architecture of our generative model using ResNet can be viewed in Figure 9. First, the encoding process consists of three convolution layers, and ReLU is used as an activation function for each layer. In the transformation process, nine

residual blocks are constructed, and each block consists of convolution layer-ReLU-convolution layer. The decoding process consists of two deconvolution layers, and each uses ReLU as an activation function. In the last decoding step, we add a final convolution layer.

3.4. Architecture of Our Generative Model Using U-Net. As shown in Figure 10, the encoder-decoder structure of our generative model using U-Net is as follows. First, the encoding process consists of eight convolution layers, and leaky ReLU is used as an activation function for each layer.

FIGURE 8: The architecture of our discriminative model.

FIGURE 9: The architecture of our generative model using ResNet. It consists of encoding, transformation, and decoding processes.

The decoding process consists of eight deconvolution layers. ReLU is used as an activation function for each layer, and 50% dropout is performed in the first to third decoding processes. We also use *concat* function in the decoding process to perform a *skip connection* that passes important information directly from the encoder to the decoder.

3.5. Methods for Synthesizing Brain Tumor-Segmented MR Images. The models in our work includes forward and backward cycles, just like the CycleGAN model proposed by Zhu et al. [14]. With these cycles, the novel synthesized image can only obtain the style of the target image while retaining the shape of the original image.

As shown in Figure 11, the architecture of our forward and backward cycles is composed of two generators ($\text{Gen}_{A \rightarrow B}$ and $\text{Gen}_{B \rightarrow A}$) and two discriminators (Dis_B and Dis_A). Forward process is as follows: $A \rightarrow \text{Gen}_{A \rightarrow B}(A) \rightarrow \text{Gen}_{B \rightarrow A}(\text{Gen}_{A \rightarrow B}(A)) \approx \hat{A}$. More specifically, our forward process can be explained in three steps. First, generator $\text{Gen}_{A \rightarrow B}$ is trained to translate an input *brain tumor* domain (A) into a *segmentation mask* domain ($\hat{B}$). Second, Dis_B is trained to discriminate the

generated image $\hat{B}$ ($\text{Gen}_{A \rightarrow B}(A) \approx \hat{B}$) from the real image B . Third, $\text{Gen}_{B \rightarrow A}$ is trained to translate the generated image $\hat{B}$ into the brain tumor MR image $\hat{A}$ ($\text{Gen}_{B \rightarrow A}(\hat{B}) \approx \hat{A}$). Likewise, backward process is as follows:

$B \rightarrow \text{Gen}_{B \rightarrow A}(B) \rightarrow \text{Gen}_{A \rightarrow B}(\text{Gen}_{B \rightarrow A}(B)) \approx \hat{B}$. More specifically, our backward process can be also defined in three steps. First, $\text{Gen}_{B \rightarrow A}$ is trained to translate an input *segmentation mask* domain (B) into a *brain tumor* domain ($\hat{A}$). Second, Dis_A is trained to discriminate the generated image $\hat{A}$ ($\text{Gen}_{B \rightarrow A}(B) \approx \hat{A}$) from the real image A . Third, $\text{Gen}_{A \rightarrow B}$ is trained to translate the generated image $\hat{A}$ into the segmentation mask MR image $\hat{B}$ ($\text{Gen}_{A \rightarrow B}(\hat{A}) \approx \hat{B}$).

The goal of the discriminator is to distinguish the novel image generated by the two generators from the real one, and therefore the discriminative neural network is trained to minimize the final classification error. On the other hand, the goal of the generator is to fool the discriminator, and therefore the generative neural network is trained to maximize the final classification error. Both networks attempt to beat each other, and this competition between them makes them evolve with respect to their respective goals. Hence, the adversarial loss function that discriminator Dis_B aims to

FIGURE 10: The architecture of our generative model using U-Net. We use *concat* in the decoding process to perform a *skip connection* that passes important information directly from the encoder to the decoder.

minimize and generator $\text{Gen}_{A \rightarrow B}$ aims to maximize is defined as

$$\mathcal{L}_B = (1 - \text{Dis}_B(B))^2 + \text{Dis}_B(\text{Gen}_{A \rightarrow B}(A))^2. \quad (2)$$

Next, the adversarial loss function that discriminator Dis_A aims to minimize and generator $\text{Gen}_{B \rightarrow A}$ aims to maximize is defined as

$$\mathcal{L}_A = (1 - \text{Dis}_A(A))^2 + \text{Dis}_A(\text{Gen}_{B \rightarrow A}(B))^2. \quad (3)$$

Adversarial loss alone cannot guarantee that the learned function can map an input domain A to a target domain B . To regularize the model and to transform source distribution into target and then back again, we should introduce the constraint of cycle-consistency into the model. An additional loss term, cycle-consistency loss, is defined as

$$\mathcal{L}_{\text{cyc}} = \|\text{Gen}_{B \rightarrow A}(\text{Gen}_{A \rightarrow B}(A)) - A\|_1 + \|\text{Gen}_{A \rightarrow B}(\text{Gen}_{B \rightarrow A}(B)) - B\|_1. \quad (4)$$

Our full objective function by putting these loss terms together and by weighting the cycle-consistency loss with a hyper parameter λ is defined as

$$\mathcal{L}_{\text{full}} = \mathcal{L}_B + \mathcal{L}_A + \lambda \mathcal{L}_{\text{cyc}}, \quad (5)$$

where we set λ to 10, which is the optimal value as introduced in CycleGAN.

3.6. Evaluation. We basically use metrics like mean-squared error and mean absolute error to evaluate the performance of the ResNet model as well as that of the U-Net model. We also evaluate the performance of the discriminator Dis_B in each model using the following metrics:

$$\mathcal{L}(\text{Dis}_{\text{real}B}) = \frac{1}{M} \sum_{i=1}^M (1 - \text{Dis}_B(B))^2, \quad (6)$$

which is the discriminator's loss for the real image B :

$$\mathcal{L}(\text{Dis}_{\text{fake}B}) = \frac{1}{M} \sum_{i=1}^M (\text{Dis}_B(\text{Gen}_{A \rightarrow B}(A)))^2, \quad (7)$$

which is the discriminator's loss for the fake image B synthesized by a generator $\text{Gen}_{A \rightarrow B}$, and

$$\mathcal{L}(\text{Dis}_B) = \mathcal{L}(\text{Dis}_{\text{real}B}) + \mathcal{L}(\text{Dis}_{\text{fake}B}), \quad (8)$$

which is the sum of the discriminator's losses for both $\text{Dis}_{\text{real}B}$ and $\text{Dis}_{\text{fake}B}$.

We also evaluate the performance of the generator $\text{Gen}_{A \rightarrow B}$ in each model using the following metrics:

$$\mathcal{L}(\text{Cycle}) = \lambda \frac{1}{M} \sum_{i=1}^M |B(i) - \text{Gen}_{A \rightarrow B}(A(i))| + \lambda \frac{1}{M} \sum_{i=1}^M |A(i) - \text{Gen}_{B \rightarrow A}(B(i))|, \quad (9)$$

which is the forward and backward cycle-consistency loss, where the hyper parameter λ is set to 10, and

$$\mathcal{L}(\text{Gen}_{A \rightarrow B}) = \mathcal{L}(\text{Cycle}) + \mathcal{L}(\text{Dis}_{\text{fake}B}), \quad (10)$$

which is the generator's loss for the fake image B synthesized by the generator $\text{Gen}_{A \rightarrow B}$, and vice versa for Dis_A and $\text{Gen}_{B \rightarrow A}$. In addition to the above-mentioned evaluation metrics, we use human perception to judge the visual quality of samples. We evaluate the quality of the generated images and whether they can segment the brain tumors well.

4. Results and Discussion

Our training process is as follows. We used 20 epochs, 100 steps, and Adam as an optimizer with initial learning rate of 0.0002 and Adam's momentum term of 0.5. The training took 28.6 hours for the ResNet model and 6.9 hours for the U-Net model. That is, the training of the ResNet took four times as long as that of the U-Net.

FIGURE 11: The architecture of our forward and backward cycles, composed of two generators ($Gen_{A \rightarrow B}$ and $Gen_{B \rightarrow A}$) and two discriminators (Dis_B and Dis_A). Forward cycle process: $A \rightarrow Gen_{A \rightarrow B}(A) \rightarrow Gen_{B \rightarrow A}(Gen_{A \rightarrow B}(A)) \approx \hat{A}$. Backward cycle process: $B \rightarrow Gen_{B \rightarrow A}(B) \rightarrow Gen_{A \rightarrow B}(Gen_{B \rightarrow A}(B)) \approx \hat{B}$. Dis_B is trained to discriminate the generated image $\hat{B}$ from the real image B. Dis_A is trained to discriminate the generated image $\hat{A}$ from the real image A.

TABLE 2: Our discriminator's and generator's losses for each generative model: ResNet, and U-Net.

Loss	ResNet	U-Net	Remarks
Dis	0.430	4.890e-3	Sum of discriminator's losses for A, B
Dis_A	0.206	2.358e-3	Sum of discriminator's losses for both real A and fake A
Dis_{fakeA}	0.282	8.767e-4	Discriminator's loss for fake A
Dis_{realA}	0.131	3.840e-3	Discriminator's loss for real A
Dis_B	0.223	2.532e-3	Sum of discriminator's losses for both real B and fake B
Dis_{fakeB}	0.252	3.000e-3	Discriminator's loss for fake B
Dis_{realB}	0.195	1.027e-3	Discriminator's loss for real B
Gen	1.338	2.572	Sum of generator's losses for $A \rightarrow B$, $B \rightarrow A$
$Gen_{A \rightarrow B}$	1.185	1.591	Generator's loss for $A \rightarrow B$
$Gen_{B \rightarrow A}$	1.125	1.549	Generator's loss for $B \rightarrow A$

A is the brain tumor domain and B is the segmentation mask domain.

FIGURE 12: Images generated by each generative model, ResNet, and U-Net, for the same samples ($A \rightarrow B$). The U-Net model locates and marks tumors in the brain MR image, whereas ResNet does not segment the exact location of the brain lesions: (a) input, (b) ResNet, (c) U-Net, and (d) ground truth.

FIGURE 13: Images generated by each generative model, ResNet, and U-Net, for the same samples ($B \rightarrow A$). Both networks produced high-quality images. However, the product's quality of ResNet is better than that of U-Net, (a) input, (b) ResNet, (c) U-Net, and (d) ground truth.

FIGURE 14: Images generated by each generative model, ResNet, and U-Net, for the same samples ($A \rightarrow B$). U-Net locates and marks tumors in the brain MR image. In contrast, ResNet does not segment the exact location of the brain lesions: (a) input, (b) ResNet, (c) U-Net, and (d) ground truth.

FIGURE 15: Images generated by each generative model, ResNet, and U-Net, for the same samples ($B \rightarrow A$). Both networks produced high-quality images. However, the product's quality of ResNet is better than that of U-Net: (a) input, (b) ResNet, (c) U-Net, and (d) ground truth.

Table 2 shows all losses of the discriminator and the generator in each training process of the ResNet and the U-Net. Full loss means the sum of all losses not only from the *brain tumor* domain to the *segmentation mask* domain but also from the *segmentation mask* domain to the *brain tumor* domain. As shown in Table 2, the full loss of the discriminator is 0.430 for the ResNet and $4.890e-3$ for the U-Net. It indicates that the performance of the discriminator is significantly higher when using U-Net than using ResNet. In contrast, the full loss of the generator is 1.338 for the ResNet and 2.572 for the U-Net. That is, the performance of the generator is slightly higher when using ResNet than using U-Net.

Additionally, Figure 12 ($A \rightarrow B$) and Figure 13 ($B \rightarrow A$) show the novel images generated by the ResNet and the U-Net model for the same samples. As shown in Figure 12 ($A \rightarrow B$), U-Net, which is a more robust model

for segmentation, located and marked tumors in the brain MR image and produced brain tumor-segmented MR image synthesis similar to ground truth. ResNet, in contrast, did not segment the exact location of the brain lesion. Figure 13 ($B \rightarrow A$) shows the novel brain images generated from the segmentation mask domain. Both networks produced high-quality images, whereas the synthesis-quality of ResNet was higher than that of U-Net (see Figures 14 and 15 for more details about the novel images generated by the ResNet and the U-Net model for the same samples).

5. Conclusions

In our work, we augmented brain tumor-segmented MR images, which consists of two tasks: synthesis and

segmentation. Therefore, we conducted two experiments, one to perform image-to-image translation, namely, image style transfer, and the other to locate and mark tumors in brain MR image, that is to say image segmentation. We performed experiments with two different generative networks, the first one using the ResNet model, which has great advantages in style transfer, and the second one, the U-Net model, one of the highly robust models for segmentation.

The performance comparison between the ResNet and the U-Net generative model is as follows. When the generator used ResNet, its training loss was slightly less than that of U-Net, and moreover it produced better-quality images than U-Net. However, it was a memory-intensive network and took much longer to train, and it did not segment the brain tumors better than U-Net. On the other hand, when the generator was with U-Net, the discriminator performed the better discrimination of whether the generated image was real or fake. Additionally, for the same samples, U-Net segmented the brain tumors more accurately than did ResNet, i.e., the segmented images generated from the brain tumor domain marked the exact location of the brain lesions.

The generative networks proposed in our paper will enable the synthesis of not only brain tumor-segmented images, but also medical images in Figure 2, as well as the novel images of segmenting tumors from the breast, uterus, and other organs, depending on the intended application. In future work, we will apply a network that combines the advantages of two networks. If we merge two models, it will be possible to generate a high-quality synthetic image with accurate segmentation. We will also increase the number of epochs and adjust hyper parameters such as initial learning rate. High-quality multimedia data augmentation using GANs has a direct impact on radiology workflow and patient care improvement. Although promising results have been reported, the adoption of GANs in medical imaging is still in its infancy and there are no clinically adopted breakthrough applications yet. Therefore, more studies and more diverse attempts are needed.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program (IITP-2020-2018-0-01405) supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP). This work was supported by the National Research

Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2017M3C4A7068189).

References

- [1] O. Ratib, N. Roduit, D. Nidup, G. De Geer, A. Rosset, and A. Geissbuhler, "PACS for Bhutan: a cost effective open source architecture for emerging countries," *Insights Into Imaging*, vol. 7, no. 5, pp. 747–753, 2016.
- [2] R. H. Choplin, J. M. Boehme II, and C. D. Maynard, "Picture archiving and communication systems: an overview," *Radiographics*, vol. 12, no. 1, pp. 127–129, 1992.
- [3] S. Lütje, J. W. J. de Rooy, S. Croockewit, E. Koedam, W. J. G. Oyen, and R. A. Raymakers, "Role of radiography, MRI and FDG-PET/CT in diagnosing, staging and therapeutic evaluation of patients with multiple myeloma," *Annals of Hematology*, vol. 88, no. 12, pp. 1161–1168, 2009.
- [4] F. G. Shellock and J. V. Cruess, "MR procedures: biologic effects, safety, and patient care," *Radiology*, vol. 232, no. 3, pp. 635–652, 2004.
- [5] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial nets," in *Proceedings of the Paper presented at the Advances in Neural Information Processing Systems*, Montreal, Canada, December 2014.
- [6] W. Dai, J. Doyle, X. Liang, Scan: structure correcting adversarial network for chest X-rays organ segmentation, arXiv preprint arXiv:1703.08770 1, 2017, 2.
- [7] J. M. Wolterink, A. M. Dinkla, and M. H. F. Savenije, "Deep MR to CT synthesis using unpaired data," in *Proceedings of the Paper presented at the International Workshop on Simulation and Synthesis in Medical Imaging*, Québec, Canada, September 2017.
- [8] X. Yi, E. Walia, P. Babyn, Unsupervised and semi-supervised learning with categorical generative adversarial networks assisted by wasserstein distance for dermoscopy image classification, arXiv preprint arXiv:1804.03700, 2018.
- [9] C. Senaras, M. K. K. Niazi, and B. Sahiner, "Optimized generation of high-resolution phantom images using cgan: application to quantification of Ki67 breast cancer images," *PLoS One*, vol. 13, no. 5, 2018.
- [10] P. Costa, A. Galdran, and M. I. Meyer, "End-to-end adversarial retinal image synthesis," *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 781–791, 2017.
- [11] F. Pollastri, F. Bolelli, R. Paredes, and C. Grana, "Augmenting data with gans to segment melanoma skin lesions," *Multimedia Tools and Applications*, pp. 1–18, 2019.
- [12] M. Rezaei, H. Yang, and C. Meinel, "Recurrent generative adversarial network for learning imbalanced medical image semantic segmentation," *Multimedia Tools and Applications*, pp. 1–20, 2019.
- [13] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: a review," *Medical Image Analysis*, vol. 58, p. 101552, 2019.
- [14] 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 Paper presented at the IEEE International Conference on Computer Vision*, Venice, Italy, October 2017.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 2016.
- [16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in

- Proceedings of the Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, October 2015.
- [17] P. Isola, J. Y. Zhu, and T. Zhou, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 2017.
 - [18] C.-B. Jin, H. Kim, M. Liu et al., "Deep CT to MR synthesis using paired and unpaired data," *Sensors*, vol. 19, no. 10, p. 2361, 2019.
 - [19] Z. Zhang, L. Yang, and Y. Zheng, "Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network," in *Proceedings of the Paper Presented at the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake, UT, USA, June 2018.
 - [20] Y. Hiasa, Y. Otake, and M. Takao, "Cross-modality image synthesis from unpaired data using cyclegan," in *Proceedings of the Paper presented at the International workshop on simulation and synthesis in medical imaging*, Shenzhen, China, October 2018.
 - [21] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Cukur, "Image synthesis in multi-contrast mri with conditional generative adversarial networks," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.
 - [22] P. Welander, S. Karlsson, A. Eklund, "Generative adversarial networks for image-to-image translation on multi-contrast MR images-a comparison of cyclegan and unit, arXiv preprint arXiv:1806.07777, 2018.
 - [23] H. Ji, D. Hooshyar, K. Kim, and H. Lim, "A semantic-based video scene segmentation using a deep neural network," *Journal of Information Science*, vol. 45, no. 6, pp. 833–844, 2019.
 - [24] M. Forouzanfar, N. Forghani, and M. Teshnehlab, "Parameter optimization of improved fuzzy C-means clustering algorithm for brain mr image segmentation," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 2, pp. 160–168, 2010.
 - [25] A. So, D. Hooshyar, K. Park, and H. Lim, "Early diagnosis of dementia from clinical data by machine learning techniques," *Applied Sciences*, vol. 7, no. 7, p. 651, 2017.