Multimodal Brain MRI Translation Focused on Lesions

Yili Qu, Chufu Deng, Wanqi Su, Ying Wang, Yutong Lu and Zhiguang Chen
School of Data and Computer Science,
Sun Yet-Sen University,
Guangzhou, China
quyli@mail2.sysu.edu.cn; dengchf3@mail2.sysu.edu.cn; suwq7@mail2.sysu.edu.cn

ABSTRACT
Registered multimodal images are lacking in many medical image processing tasks. To obtain sufficient registered multimodal data, in this paper, we propose a new unsupervised scheme for medical image translation based on cycle generative adversarial networks (CycleGAN), which can generate registered multimodal from single modality and retain the lesion information. We improve parameter initialization method, upsampling method and loss items to speed up model training and improve translation quality. Compared with previous studies that focus only on the overall quality of translation, we attach more importance to the lesions information in medical images, so we propose a method for the preservation of lesions information in the translation process. We perform a series of multimodal translation experiments on the BRATS2015 dataset, verify the effect of each of our improvements as well as the consistency of the lesions information between translation images and original images. And we also verify the effectiveness and availability of the lesions information in translation images.

CCS Concepts
• Computing methodologies→Computer vision • Computing methodologies→Unsupervised learning

Keywords
GAN; multimodal; medical images; image translation; brain MRI.

1. INTRODUCTION
Medical images have many modalities, such as MRI with different contrast and CT with different doses. For doctors, medical images of different modalities have different reference values. The more modal data can bring more evidence to doctors for diagnosis and treatment. It is natural to expect more multimodal medical data in the study of medical image intelligent processing. However, researchers can only obtain a small number of sample data due to the difficulty of medical image acquisition. And there are few samples and scarce registered multimodal data in the current popular medical images open dataset. Therefore, some studies[1, 2, 3] are devoted to image translation to get other modality images. This idea makes an important contribution to reducing the cost for doctors and patients in diagnosis(e.g. reduction of radiation dose) and improving the feasibility of treatment[4], Meanwhile, it is also an effective solution to the scarcity of data samples problem[5].

ICMLC 2020, February 15–17, 2020, Shenzhen, China

ACM ISBN 978-1-4503-7642-6/20/02…$15.00
DOI: https://doi.org/10.1145/3383972.3384024

The cross-modal translation of medical images refers to the process of transforming images of one modality into images of other modalities, which is similar to style transfer[6, 7] of transforming one style image into another in natural image processing.

The wider image translation includes the translation and style migration of medical image modalities, as well as scenes such as image denoising[8] and portrait transformation[9, 10]. In recent studies, the image-to-image translation task is formulated as a pixel-to-pixel mapping using an encoder and decoder CNN[11, 12, 13]. Some studies explored cross-modal medical images translation prior to GAN by using random forests[12], graph dictionary mapping[4], sparse coding[15, 16], and CNN[17]. Since then, many studies used GAN to generate higher quality translation results[9, 18, 19, 20].

Owe to the enormous capabilities, GAN has become the mainstream to achieve multimodal medical image translation[1, 2, 3, 17, 21]. The general translation is based on paired data. After the implementation of unpaired image-toimage translation by CycleGAN[18], some studies have also learned from unpaired cross-modal data[1]. Recent studies have realized brain MRI to CT image translation with pixelto-pixel GAN[3, 2], retinal vascular annotation to image translation[22], cardiac MRI to CT image translation and segmentation[3]. In these medical image translation studies, we found most of them are two-modal translation[1, 2, 3, 4, 16, 17, 21], and the study of multimodal translation is very rare. In the field of natural image processing, the development of many-to-many translation has recently made great progress[9, 19, 23, 24]. Therefore, the application of them in medical images to achieve more efficient and accurate multimodal translation is an attractive problem.

Thomas et al.[25] and Agisilaos et al.[26] made the initial attempt of multimodal MRI synthesis with deep learning method, then Salman et al.[27] realized the two-modal translation of MRI T1/T2 based on CycleGAN, and achieved the best current translation quality. In addition to the overall translation quality, lesions information is the most important part of medical images, some researches[5, 25] therefore conducted experiments to generate tumors by using labels, and Hoochang et al.[5] further carried out the availability verification experiments on synthetic data. However, in these researches for medical image translation, there is no exploration of retaining the original image lesions information in the process of translation.

For the translation of medical images, it is not only required to be sufficiently good in visual effect and quantitative evaluation but also required high fidelity of pathological information. There are many tiny lesions in medical images, and they are the key to a doctor's diagnosis and treatment. Whether the original modal

lesions are successfully retained and translated during the translation process determines whether the translation modality can be used for medical diagnosis. Based on the above research work, we improve CycleGAN and propose a new multimodal translation method for medical images using modular CycleGAN (MCGAN) to achieve unsupervised multimodal MRI translation, and the lesions information is well preserved in translation. In this paper, we perform lesions consistency and lesions effectiveness validation experiments on the translation results to show that our translation process is non-destructive to lesions information. Moreover, we conduct data availability verification experiments to verify that our synthetic data can be applied to the training of multimodal intelligent medical image learning. The translation data can improve the generalization ability of the model, and even replace the real data to some extent. Specifically, our contributions include:

- We design an unsupervised translation scheme MCGAN for four-modal translation on the BRATS2015, and the scheme is easy to expand to a supervised situation. We use deconvolution+bilinear-resize as our upsampling method, which can effectively learn and eliminate the checkerboard artifacts. And we use the arithmetic mean filter to perform special initialization to speed up learning according to the special needs of the image translation task. We specially design a series of loss items, including specialized SSIM loss and Sobel loss to promote translation image detail and texture, so that the translation results are improved in various indexes.
- We use an independently trained tumor segmentation network to provide lesions consistency loss for translation training, thus ensuring good preservation of tumor lesions during the translation process.
- We perform lesions consistency and lesions effectiveness validation experiments on the translation results to show that our translation process is non-destructive to lesions information. We also conduct data availability verification experiments to verify that our synthetic data can be applied to the training of multimodal intelligent medical image learning. Experiment results show that registered multimodal synthetic data obtained by MCGAN from a single modal can replace the real data and use in medical image intelligent processing tasks.

2. METHOD

2.1 Modular and Network

When we extend two-modal translation to multimodal, our ideas coincide with ComboGAN[23] and XGAN[24]. Especially in medical image multimodal translation, independent single-function modular components are more reliable than conditional multi-function components or translators. As shown in Fig. 1, our modules include encoders, decoders, discriminators, our modularization is similar to ComboGAN and XGAN. However, considering the similarity between multimodal of medical images, only one encoder is used to encode different modalities images into the same semantic space. We use a discriminator for true/false identification and modality identification. Compared with the former two, our method greatly reduces the cost of parameters.

The current excellent model structure is almost uncountable, and our approach is network-independent. In our experiments, the structure of various components is a simple multi-layer full convolution network, the specific structure can be seen in our open-source code[1].

We apply a special upsampling method in our network. There are two popular methods of upsampling: Deconvolution is the most commonly used upsampling method, which has a high learning efficiency but an inevitable checkerboard artifacts[28]. Zhang et al.[1] replaced the deconvolution in CycleGAN with the method of resize upsampling in the translation of cardiac CT and MRI, thereby avoiding the checkerboard artifacts. In practice, we find that resize upsampling method is extremely sensitive to parameter initialization and difficult to learn. Therefore, we combine the advantages of both, using deconvolution+bilinear-resize as our upsampling method, which can retain the learning ability of deconvolution as well as to use resize for a smooth transition, thus to avoid the checkerboard artifacts. In our experiment, the addition upsampling result is equal to 0.2 ×deconvolution result+0.8 ×bilinear-resize result on each pixel. Compared to other learnable upsampling methods or data-related upsampling methods [29], our method is simple and convenient, without complicated coding implementation and training.

Since a large proportion of bilinear-resize upsampling is used to balance the checkerboard artifacts, to speed up the learning, our generator adopts a special parameter initialization method. Our generator is a full convolution structure. Considering that our task is image translation, which means the input is very similar to the expected output. Therefore, we take the arithmetic mean filter parameters as the initial convolution kernels parameters to make the output obtained by the generator after network calculation as similar as possible to the input, so as to reduce the learning difficulty. In practice, when initializing parameters for a convolutional layer with a [k,k,f] size kernel, we use the random normal initializer to set the mean value of $1/(k \times k \times f)$, the standard deviation is 0 and bias is 0.

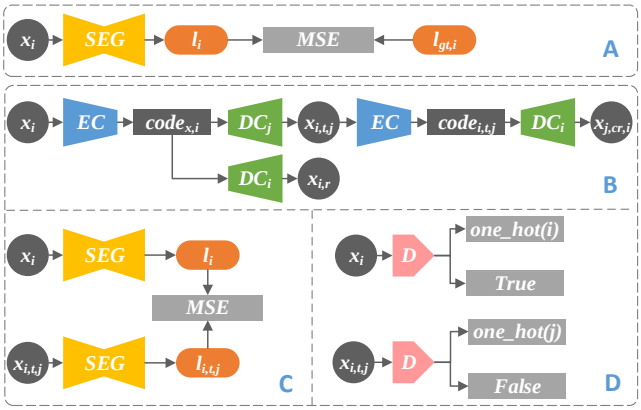

Figure 1: Training architecture.

2.2 Training Architecture

The training process of each module is shown in Fig. 1, Part A is the training of the tumor segmentor. After the segmentor training is completed, as shown in part C, it is used to provide loss guidance for multimodal translation training. Part B, C and D constitute a multimodal translation training. Part B is the generator training, multimodal sharing an encoder, and an additional reconstruction training process is added in translation training. Part D is the discriminator training. And the whole training process can be expressed as:

[1] https://github.com/tofudofu/MCGAN

Where EC,DC_i denote the encoder and decoder of modality i, D denotes the discriminator, SEG denotes the tumor segmentor. $x_i,x_{i,r},x_{i,t,j},x_{j,cr,i}$ denote the real image of modality i, reconstruction image of modality i, translation image translate from modality i to j, cycle-reconstruction image translate from modality i to j and then translate back to modality i.

$code_i,code_{i,t,j}$ denote the last encoder layer output after the real image x_i and translation image $x_{i,t,j}$ encode by EC respectively. d_i,c_i denote the true/false identification results and category identification results of real image x_i discriminate by D. $d_{i,t,j},c_{i,t,j}$ denote the true/false identification results and category identification results of translation image $x_{i,t,j}$ discriminate by D. $l_i,l_{i,r},l_{i,t,j},l_{j,cr,i}$ denote the SEG segmentation results of $x_i,x_{i,r},x_{i,t,j},x_{j,cr,i}$ respectively.

2.3 Loss Items

According to the training architecture, we design a series of loss items.

Segmentor Training Loss The segmentor is trained separately in advance, and its loss function is as follow, where $\mathbb{E}$ denotes the expectation function.

$$\Lambda_{seg}\ (SEG) = \mathbb{E}_{x,l}\left[\sum_{i=0} \parallel l_i - l_{gt,i} \parallel_2^2\right]$$

where $l_{gt,i}$ is the ground truth label of x_i.

Discriminant Loss The discriminator is synchronized training with the generator, but updated separately, and its loss function is

$$\Lambda_d\ (D) = \mathbb{E}_x[\sum_{i=0}(\parallel d_i - 1 \parallel_2^2 + \parallel c_i - i \parallel_2^2]) + \sum_{j=0}\sum_{i=0}(\parallel d_{i,t,j} \parallel_2^2 + \parallel c_{i,t,j} - j \parallel_2^2)]$$

Adversarial Loss and Category Loss

$$\Lambda_a\ (EC, DCs) = \mathbb{E}_x[\sum_{j=0}\sum_{i=0}(\parallel d_{i,t,j} - 1 \parallel_2^2 + \parallel c_{i,t,j} - j \parallel_2^2)]$$

Reconstruction Loss

$$\Lambda_r\ (EC, DCs) = \mathbb{E}_x[\sum_{i=0}(\parallel x_i - x_{r,i} \parallel_2^2)]$$

Cycle Consistency Loss

$$\Lambda_{cr}\ (EC, DCs) = \mathbb{E}_x\left[\sum_{j=0}\sum_{i=0}(\parallel x_i - x_{j,cr,i} \parallel_2^2) + \sum_{k=0}\sum_{j=0}\sum_{i=0}(\parallel x_{j,cr,i} - x_{k,cr,i} \parallel_2^2)\right]$$

Semantic Consistency Loss

$$\Lambda_{code}\ (EC, DCs) = \mathbb{E}_x\left[\sum_{i=0}(\parallel code_i - code_{r,i} \parallel_2^2]) + \sum_{j=0}\sum_{i=0}(\parallel code_i - code_{i,t,j} \parallel_2^2)\right]$$

SSIM Loss Some studies[30] found that the optimization direction of the mean squared loss (MSE) has limited improvement on structural similarity index measure (SSIM). So we add structural similarity loss to improve the structural similarity of the translation results. Based on the structural similarity function $ssim(\cdot)$[31], we design function:

$$\Lambda_{a,b}^{ssim} = 1 - ssim(a,b)$$

Then our structural similarity loss is:

$$\Lambda_{ssim}\ (EC, DCs) = \mathbb{E}_x[\sum_{j=0}\sum_{i=0}(\Lambda_{x_i,x_{j,cr,i}}^{ssim})]$$

Sobel Loss In the field of traditional digital image processing, people use the excellent edge detection operator such as Roberts operator [32], Prewitt operator[33], Sobel operator[34] to extract texture features in biomedical images. Both the Prewitt operator and the Sobel operator are templates for 3×3, which have a more accurate partial derivative than the Roberts operator with 2×2 templates. Compared with the Prewitt operator, the Sobel operator weights the influence of the pixel position, which can better suppress noise and reduce edge blur, so the effect is better[35]. Learn from this, we design a loss function based on the Sobel operator function $sobel(\cdot)$ to emphasize the reconstruction of edge details:

$$\Lambda_{a,b}^{sobel} = \parallel reduce\_max(sobel(a)) - reduce\_max(sobel(b)) \parallel_2^2$$

Then our sobel loss is:

$$\Lambda_{sobel}\ (EC, DCs) = \mathbb{E}_x[\sum_{j=0}\sum_{i=0}(\Lambda_{x_i,x_{j,cr,i}}^{sobel})]$$

Lesions Supervise Loss

$$\Lambda_{lesions}\ (EC, DCs) = \mathbb{E}_{x,l}\left[\sum_{i=0}(\parallel l_i - l_{i,r} \parallel_2^2) + \sum_{j=0}\sum_{i=0}(\parallel l_i - l_{i,t,j} \parallel_2^2) + \sum_{k=0}\sum_{j=0}\sum_{i=0}(\parallel l_{i,t,j} - l_{i,t,k} \parallel_2^2)\right]$$

Therefore, our total generator loss is as follows. In backforward, each loss will calculate gradients to update corresponding module parameters:

$$\Lambda_g\ (EC, DCs) = \Lambda_a + \Lambda_r + \Lambda_{cr} + \Lambda_{code} + \Lambda_{ssim} + \Lambda_{sobel} + \Lambda_{lesions}$$

3. EXPERIMENTS

3.1 Dataset and Training Settings

BRATS2015[36] is an open dataset with four modalities of T1/T2/T1c/Flair. The training dataset contains 274 subjects, each subject has registered four-modal 3D MRI with a size of 155×240×240 and a tumor segmentation label. We take 20% subjects as the real testing dataset in experiments. During data preprocessing, we normalize each 3D image and divide it into 2D axial-plane slices.

We train 100 epochs for each experiment using Adam optimizer with beta1 of 0.5. The learning rate is 1e-4 without weight decay. We perform a Dropout of 0.1 on the input layer, Batch size is 1. In generator components, the mean kernel filter parameter is used to initialize the convolution kernel parameter. In discriminator, we use a random normal initializer with a mean of 0 and a standard deviation of 0.2, and the bias is 0. The evaluation result is the best result calculated on full-size images after four times of training. All training is carried out on a Tesla V100 GPU cluster with 64 nodes based on the TensorFlow1.9 and PyThon2.7 environment.

3.2 Verification Experiment for Translation Quality Improvement

As shown in Table 1, we perform various translation experiments on BRATS2015 to verify the quality promotion of the upsampling method, ssim loss, sobel loss, semantic consistency loss, lesions supervise loss and reconstruction training process. Finally, we compare T1↔T2 translation results of MCGAN with the current best baseline[27]. In addition, we perform T1→T2, T1→Flair translation experiments according to the dataset processing method of baseline[26], and the translation results are compared with baseline[26].

3.3 Verification Experiment for Lesions Effectiveness

As shown in Table 4, we generate four-modal synthetic data by random unimodal, then we combine the synthetic data with real data, use them as training data and testing data respectively to verify the medical effectiveness of synthetic data. Among them, to verify the role of lesions loss Λlesions we proposed, we design a group of comparative experiments with and without lesions loss.

Table 1: Verification Experiment Setting for Translation Quality Improvement

	Upsampling	Λ_{ssim}	Λ_{sobel}	Λ_{code}	Λ_r
MCGAN with transpose	transpose	√	√	√	√
MCGAN with resize	resize	√	√	√	√
MCGAN without Λ_{ssim}	transpose + resize	√	√	√	√
MCGAN without Λ_{sobel}	transpose + resize	×	×	√	√
MCGAN without Λ_{code}	transpose + resize	√	√	×	√
MCGAN without Λ_r	transpose + resize	√	√	√	×
MCGAN	transpose + resize	√	√	√	√

Table 2: Translation Quality Evaluation for Improvements

	*→T1 PSNR	*→T2 PSNR	*→T1c PSNR	*→Flair PSNR	*→T1 SSIM	*→T2 SSIM	*→T1c SSIM	*→Flair SSIM
MCGAN with transpose	26.62	25.70	25.86	25.30	0.933	0.924	0.920	0.921
MCGAN with resize	27.76	26.86	27.20	26.88	0.936	0.927	0.929	0.927
MCGAN without Λ_{ssim}	28.07	27.14	27.38	26.70	0.936	0.928	0.932	0.928
MCGAN without Λ_{sobel}	27.25	26.64	27.37	26.17	0.938	0.929	0.931	0.927
MCGAN without Λ_{code}	23.65	23.26	23.15	23.56	0.846	0.843	0.844	0.841
MCGAN without Λ_r	27.72	26.71	26.81	26.68	0.935	0.926	0.929	0.926
MCGAN	**28.09**	**27.15**	**27.76**	**27.37**	**0.940**	**0.931**	**0.932**	**0.929**

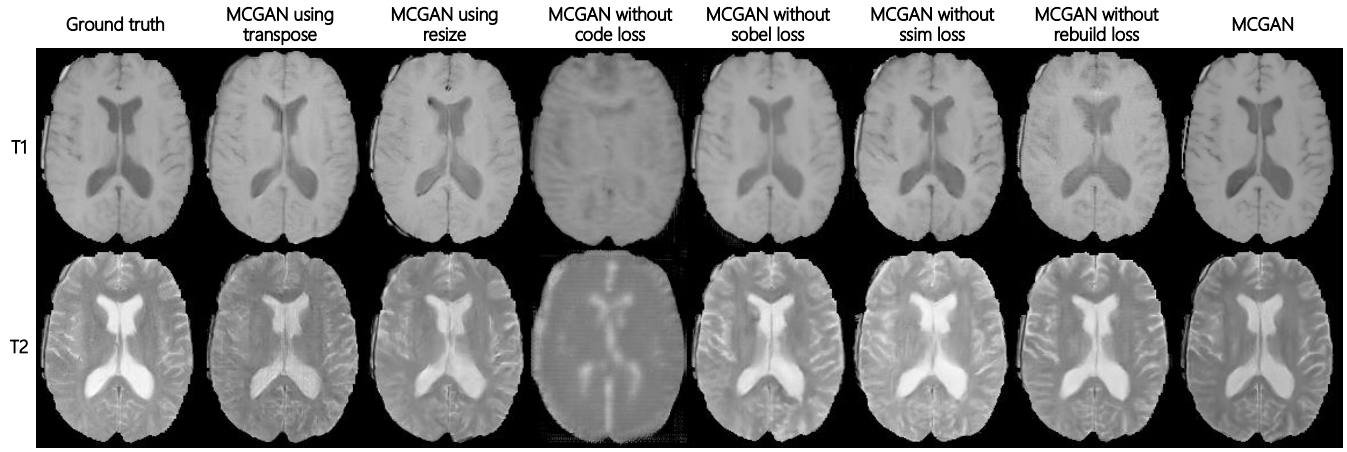

Figure 2: T1↔T2 Visual effect comparison of each improvement on BRATS2015.

Table 3: Comparison of T1↔T2 Test Results on BRATS2015. "Local" means test on the circumscribed rectangular area of the lesion

	Data Type and Setting	T1→T2 PSNR	T2→T1 PSNR	T1→T2 SSIM	T2→T1 SSIM
REPLICA [14] [27]	registered	24.64	24.49	0.924	0.917
Multimodal [26] [27]	registered	25.09	23.78	0.939	0.935
pGAN[27]	registered	27.19	25.80	0.946	0.940
MCGAN without Λ_{lesion}	unregistered	27.56	26.97	0.942	0.937
MCGAN	unregistered	27.78	27.13	0.944	0.939
MCGAN	registered	**31.29**	**29.03**	**0.963**	**0.957**
MCGAN without Λ_{lesion}	unregistered, Local	27.16	26.77	0.942	0.939
MCGAN	unregistered, Local	28.42	28.18	0.954	0.951

Table 4: Verification Experiment for Lesions Effectiveness

Source Modality	Training Data	Testing Data	Dice Score	MSE
-	real	real	**0.921**	**0.012**
random	real	synthetic without $\Lambda_{lesions}$	0.544	0.101
random	synthetic without $\Lambda_{lesions}$	real	0.413	0.160
random	real	synthetic	0.917	0.013
random	synthetic	real	0.907	0.028

3.4 Verification Experiment for Data Availability

As shown in Table 5, firstly, we take synthetic data obtained from different source modalities and real data as training data separately, and test with real data to verify if there are differences among the synthetic data obtained by different unimodal. Then, we mix synthetic data obtained by random modality and real data in different proportions and different ways as training data, also test with real data to verify whether the translated synthetic data can replace the real data in intelligent lesions processing task.

Table 5: Verification Experiment for Data Availability

Source Modality	Training Data	Mix Method	Dice Score	MSE
-	100%real	-	**0.921**	**0.012**
-	50%real	-	0.906	0.029
T1	100%synthetic	-	0.904	0.031
T2	100%synthetic	-	0.911	0.017
T1c	100%synthetic	-	0.906	0.022
Flair	100%synthetic	-	0.910	0.018
random	100%synthetic	-	0.907	0.028
random	20%synthetic+80%real	random	0.920	0.012
random	80%synthetic+20%real	random	0.909	0.026
random	50%synthetic+50%real	random	0.914	0.015
random	50%synthetic+50%real	synthetic 1st	0.917	0.013
random	50%synthetic+50%real	real 1st	0.908	0.026

4. RESULTS

We use the peak signal-to-noise ratio (PSNR) and SSIM for quantitative analysis the quality of our translation results, use MSE and Dice Score as evaluation indicators of segmentation quality. And we make qualitative evaluation through visual effect.

4.1 Analysis of Translation Quality Improvement

Table 2 shows the quality evaluation of four-modal translation. The evaluation score for each modal in the table is the average score for the transition MRIs from all other modes. From the results, our improvement points can improve the translation quality to a certain extent, and Λ_{code} has the greatest promotion on translation. Since our method is not required to be registered, we do not have direct constraints on the translation result, but improve the translation quality indirectly by constraining the cycle-reconstruction image, thus Λ_r, Λ_{ssim} and Λ_{sobel} have relatively small improvement effects. From the experimental situation, the translation quality has been greatly improved when all the improvements are applied. Fig. 2 shows the visual improvement effect of our various improvements on BRATS2015 translation results. And Fig. 3 shows the visual effect of multimodal translation on BRATS2015.

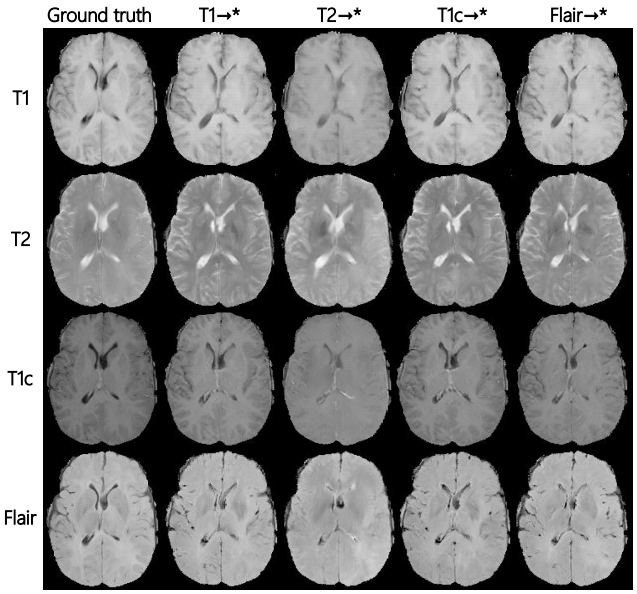

Figure 3: Example for multimodal translation on BRATS2015.

As shown in Table 3, to compare with the previous methods based on supervised learning, we additionally train MCGAN on registered images. Specifically, we obtained the consistency loss by calculating the mean square error between the original images and translation images, At the same time, add supervision information to the ssim loss and sobel loss. The supervision loss is as follow:

$$\Lambda_{supervision}\ (EC, DCs) = \mathbb{E}_x[\sum_{j=0}\sum_{i=0}(\| x_i - x_{j,t,i} \|_2^2 + \Lambda_{x_i,x_{j,t,i}}^{sobel} + \Lambda_{x_i,x_{j,t,i}}^{ssim})]$$

The results show that our MCGAN trained on registered images get the highest PSNR and SSIM. even with training on unregistered images, our method can compete with state-of-the-art methods pGAN[27] on SSIM and better than the other methods. Compare the PNSR, our results are higher. This shows that our method can achieve a competitive overall transformation quality compared with current methods that rely on registration data. To verify the translation quality of our lesions loss on the lesions area, we evaluated the PSNR and SSIM for the circumscribed rectangular area of the lesions (the red dotted frames in Fig. 4). Because pGAN did not specifically focus on the preservation of lesions information, we assume that its assessment of the lesions area is equal to the assessment of the entire image. The results showed that, without lesion loss, the differences of translation quality between the entire image and the lesions area are not significant; in terms of the translation quality of the lesions area, the PSNR and SSIM of the experiments with lesions loss are the best, this shows that our lesions loss does improve the translation quality of the lesions area.

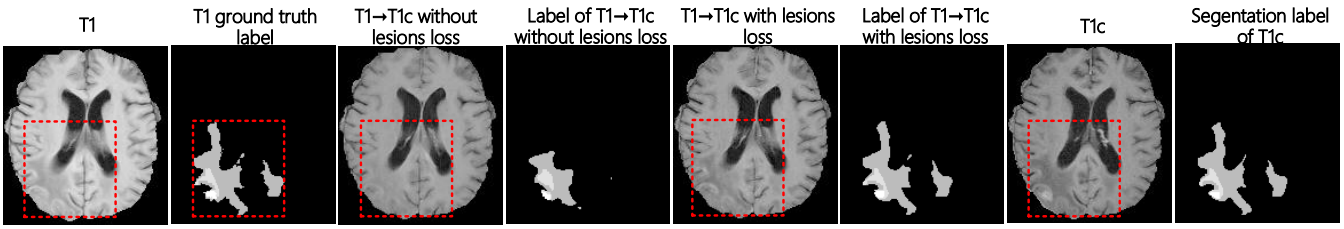

Figure 4: Comparison of multimodal synthetic MRI and tumor label with or without Llesions.The red dotted frames denote the circumscribed rectangular area of the lesion. The synthetic images with or without Llesions shown in figure are very similar to the real image, but the synthetic images without Llesions destroy the potential lesions information in translation process, making the segmentation results of the segmentor inaccurate, and the synthetic images with Llesions retain the lesions information very well.

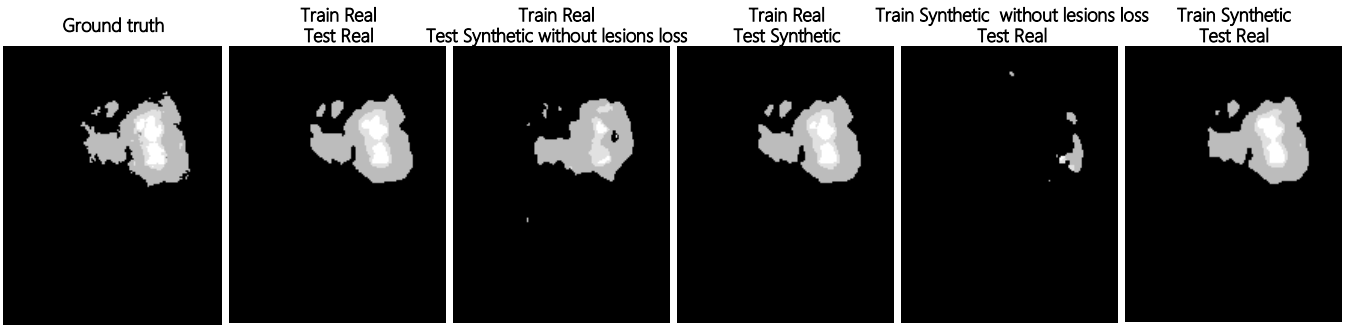

Figure 5: Comparison of segmentation effects in lesions effectiveness verification experiment.

4.2 Analysis of Lesions Effectiveness

As shown in Table 4, we use segmentor trained with real data to segment the synthetic data without $\Lambda_{lesions}$. From the results, although synthetic data without $\Lambda_{lesions}$ overall similar to real data, potential details such as lesions information have changed. We use segmentor trained with real data to segment the synthetic data with $\Lambda_{lesions}$, and the segmentation results are consistent with the real data, which indicates that $\Lambda_{lesions}$ effectively retains the lesions information of translation process. Further, we use synthetic data for training and then test on real data. We found that the performance of synthetic data guided by $\Lambda_{lesions}$ in the lesions segmentor training is close to real data, indicating that the synthetic data obtained by our proposed scheme could be used as real data to some extent. In data availability verification experiment, we further validate the availability of synthetic data in the segmentation experiment. The visual comparing is shown in Fig. 4 and Fig. 5.

4.3 Analysis of Data Availability

As shown in Table 5, we use different data as training dataset of segmentor, and real data as testing dataset. The test results obtained from the individual unimodal are comparable, indicating that our model is robust to different source modalities. In the segmentation experiment which mixes synthetic data and real data in different proportions, we find that the higher the real data ratio, the better the segmentation effect while the effect of 20% synthetic data and 50% synthetic data are not much different. It shows that the synthetic data can not fully achieve the effect of real data but can be much closer to the real data. The experimental results of different mix methods indicate that the synthetic data has the following uses: (i) mix synthetic data with real data and use as enhanced data in training; (ii) use synthetic data in pre-training and fully trained using real data; (iii) train on real data then use synthetic data in supplement training to improve model generalization ability. The best way to use the synthetic data is as a pre-training dataset. In general, the results show that synthetic data obtained by MCGAN can replace the real data in intelligent lesions processing tasks, and in the absence of sufficient real data, the synthetic data can be used as a substitute for real data. The direct medical alternative ability of synthetic data is not yet validated. However, from the potential presented in intelligent lesions processing task, we are cautiously conservative in the direct medical alternatives to the current synthetic data, but optimistic about the future.

5. CONCLUSIONS

We propose a new scheme for multimodal translation of medical images based on CycleGAN. Compared with the current state-of-the-art translation methods, we have achieved promotion on a number of translation quality evaluation indicators. We use a number of improvements to make the texture details of the translation results more realistic. We verify the effectiveness of the proposed improvements by a series of comparative experiments. According to the characteristics of medical images, we design a method of lesions retention during the translation process, and show that the lesions information in our translation image is well preserved by the lesions effectiveness verification experiment. In addition, we conduct sufficient data availability verification experiments to verify that our translation data can replace real data in intelligent lesions processing tasks.

6. ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China under Grant NO. 2018YFB0203904; the Nature Science Foundation of China under Grant NO. U1611261, NO. 61872392, and NO. U1811461; the Program for Guangdong Introducing Innovative and Entrepreneurial Teams under Grant NO. 2016ZT06D211; the Pearl River S&T Nova Program of Guangzhou under Grant NO. 201906010008; and the Natural Science Foundation of Guangdong Province, China, under Grant NO. 2018B030312002.

Additional authors: Nong Xiao (School of Data and Computer Science, Sun Yet-Sen University).

7. REFERENCES

[1] Zizhao Zhang, Lin Yang, and Yefeng Zheng. Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network. computer vision and pattern recognition, pages 9242–9251, 2018.

[2] Konstantinos Kamnitsas, Christian F Baumgartner, Christian Ledig, Virginia Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Aditya V Nori, Antonio Criminisi, Daniel Rueckert, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. information processing in medical imaging, pages 597–609, 2017.

[3] Dong Nie, Roger Trullo, J Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. medical image computing and computer assisted intervention, pages 417–425, 2017.

[4] Ninon Burgos, M J Cardoso, Filipa Guerreiro, Catarina Veiga, Marc Modat, J Mcclelland, Antjechristin Knopf, Shonit Punwani, D Atkinson, Simon R Arridge, et al. Robust ct synthesis for radiotherapy planning: Application to the head and neck region. medical image computing and computer assisted intervention, pages 476–484, 2015.

[5] Hoochang Shin, Neil A Tenenholtz, Jameson K Rogers, Christopher G Schwarz, Matthew L Senjem, Jeffrey L Gunter, Katherine P Andriole, and Mark Michalski. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. arXiv: Computer Vision and Pattern Recognition, pages 1–11, 2018.

[6] Leon Gatys, Alexander S Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, Advances in Neural Information Processing Systems 28, pages 262–270. Curran Associates, Inc., 2015.

[7] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. Journal of Vision, 16(12):326–326, 2016.

[8] Xiaojiao Mao, Chunhua Shen, and Yubin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. Neural Information Processing Systems, pages 2810–2818, 2016.

[9] Yunjey Choi, Minje Choi, Munyoung Kim, Jungwoo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. computer vision and pattern recognition, pages 8789–8797, 2018.

[10] Guim Perarnau, Joost Van De Weijer, Bogdan Raducanu, and Jose M Alvarez. Invertible conditional gans for image editing. arXiv: Computer Vision and Pattern Recognition, 2016.

[11] Phillip Isola, Junyan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. computer vision and pattern recognition, pages 5967–5976, 2017.

[12] Mingyu Liu, Thomas M Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. neural information processing systems, pages 700–708, 2017.

[13] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. international conference on machine learning, pages 1857–1865, 2017.

[14] Amod Jog, Aaron Carass, Snehashis Roy, Dzung L Pham, and Jerry L Prince. Random forest regression for magnetic resonance image synthesis. Medical Image Analysis, 35:475–488, 2017.

[15] Yawen Huang, Ling Shao, and Alejandro F Frangi. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. computer vision and pattern recognition, pages 5787–5796, 2017.

[16] Raviteja Vemulapalli, Hien Van Nguyen, and Shaohua Kevin Zhou. Unsupervised cross-modal synthesis of subject-specific scans. international conference on computer vision, pages 630–638, 2015.

[17] Hien Van Nguyen, Kevin S Zhou, and Raviteja Vemulapalli. Cross-domain synthesis of medical images using efficient location-sensitive deep network. medical image computing and computer assisted intervention, pages 677–684, 2015.

[18] Junyan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. international conference on computer vision, pages 2242–2251, 2017.

[19] Bo Zhao, Bo Chang, Zequn Jie, and Leonid Sigal. Modular generative adversarial networks. european conference on computer vision, pages 157–173, 2018.

[20] Xiaodan Liang, Hao Zhang, Liang Lin, and Eric P Xing. Generative semantic manipulation with mask-contrasting gan. european conference on computer vision, pages 574–590, 2018.

[21] Anton Osokin, Anatole Chessel, Rafael E Carazo Salas, and Federico Vaggi. Gans for biological image synthesis. international conference on computer vision, pages 2252–2261, 2017.

[22] Pedro Costa, Adrian Galdran, Maria Ines Meyer, Michael D Abramoff, Meindert Niemeijer, Ana Maria Mendonca, and Aurelio Campilho. Towards adversarial retinal image synthesis. IEEE Transactions on Medical Imaging, 2017.

[23] Asha Anoosheh, Eirikur Agustsson, Radu Timofte, and Luc Van Gool. Combogan: Unrestrained scalability for image domain translation. computer vision and pattern recognition, pages 783–790, 2018.

[24] Amelie Royer, Konstantinos Bousmalis, Stephan Gouws, Fred Bertsch, Inbar Mosseri, Forrester Cole, and Kevin P Murphy. Xgan: Unsupervised image-to-image translation for many-to-many mappings. arXiv: Computer Vision and Pattern Recognition, 2018.

[25] Thomas Joyce, Agisilaos Chartsias, and Sotirios A Tsaftaris. Robust multi-modal mr image synthesis. pages 347–355, 2017.

[26] Agisilaos Chartsias, Thomas Joyce, Mario Valerio Giuffrida, and Sotirios A Tsaftaris. Multimodal mr synthesis via modality-invariant latent representation. IEEE Transactions on Medical Imaging, 37(3):803–814, 2018.

[27] Salman Ul Hassan Dar, Mahmut Yurt, Levent Karacan, Aykut Erdem, and Tolga ˜ A˘Gukur. Image synthesis in multi-

contrast mri with conditional generative adversarial networks. IEEE Transactions on Medical Imaging, PP(99):1–1, 2018.

[28] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. Distill, 2016.

[29] Zhi Tian, Chunhua Shen, Tong He, and Youliang Yan. Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation. 2019.

[30] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Peter Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. computer vision and pattern recognition, pages 105–114, 2017.

[31] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4):600–612, 2004.

[32] Roberts Lawrence G. Machine perception of three-dimensional solids. Optical and Electro-Optical Infomation Processing, 1965.

[33] J.M.S. Prewitt. Object enhancement and extraction. Picture processing and Psychopictorics, 1970.

[34] Irwin Sobel. Camera models and machine perception. Ph.D.dissertation, 1970.

[35] Richard E. Woods Rafael C. Gonzalez. Digital image processing, 3th edition. 2008.

[36] Bjoern Menze and Jakab et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging, page 33, 2014.