Synthesis of Registered Brain Multimodal MRI with Lesions

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

Abstract

In a large number of data-driven medical images intelligent processing tasks, the collection and acquisition of medical image data is very difficult, especially the registered multimodal medical images data. Synthetic medical image data can well alleviate the problem of insufficient data. In this paper, based on the unsupervised conditional GAN model, we achieve the generation of registered multimodal medical images from random normal distribution matrix and corresponding lesion information can be efficiently generated based on the freely selected lesion label. We conducted a number of validation experiments on BRATS2015 to verify that our synthetic MRI can be used as pre-trained data or enhanced data in medical image intelligent processing tasks, and can greatly improve the generalization ability of the model.

1 Introduction

Magnetic resonance imaging (MRI) is a common medical image that can have multiple modalities depending on imaging parameters, such as T1, T2, T1c and so on. Different modalities have different reference values for doctors. To make accurate judgments, doctors often need multiple modal images to compare with each other. In training and learning of medical image intelligent processing tasks, we often expect to obtain more modal images, such as medical image processing tasks based on Convolutional Neural Networks (CNN)(Krizhevsky, Sutskever, and Hinton 2012) or Generative Adversarial Networks (GAN)(Goodfellow et al. 2014).

When obtaining different modalities from the same part of the same patient through different imaging techniques, these modalities are considered to be registered if the imaging position and the viewing angle are identical. Compared with unimodal data, the registered multimodal image data can provide more information, can support more complex application scenarios, meet the training data requirements of deep neural networks, and help to provide intelligent diagnostic services more efficient and more reliable. For doctors, it takes longer to acquire images of different modalities and requires patient patience. For researchers of medical image intelligent processing tasks, multimodal MRI datasets are scarce, and the collection is very difficult, especially rare disease data, and the registered data is even rarer, which makes many training tasks impossible. Therefore, the application of image synthesis technology to extend datasets, translate existing unimodal images to registered multimodal images, or generate registered multimodal images from random noise matrix, has a wide range of uses and far-reaching significance.

Before the popularity of GAN, researchers focused on existed medical images modal translate to other modals(Burgos et al. 2015),(Huang, Shao, and Frangi 2017),(Vemulapalli, Van Nguyen, and Zhou 2015),(Van Nguyen, Zhou, and Vemulapalli 2015).After GAN has shown its powerful generating ability, there are many researches on medical image translation based on GAN(Zhang, Yang, and Zheng 2018),(Nie et al. 2017),(Osokin et al. 2017),(Van Nguyen, Zhou, and Vemulapalli 2015),(Kamnitsas et al. 2017).Many studies use GAN to achieve higher quality translation results(Zhao et al. 2018),(Liang et al. 2018),(Zhu et al. 2017),(Choi et al. 2018).Recent studies have implemented the translation on unregistered multimodal data(Zhang, Yang, and Zheng 2018),(Joyce, Chartsias, and Tsaftaris 2017). GAN has been gradually applied to various organs, such as brain MRI to CT translation(Nie et al. 2017),(Kamnitsas et al. 2017), retinal vascular annotation to image translation(Costa et al. 2017), cardiac MRI to CT translation and segmentation using unsupervised CycleGAN(Zhu et al. 2017),(Nie et al. 2017). For multimodal image synthesis, (Chartsias et al. 2018) implements multi-input and multi-output MRI synthesis, but registration is required for input multi-modal data; (**?**) further implements an unregistered multi-input synthesis model, which can perform image synthesis from any subset of its input, but limits the output to single modal;(**?**) studies for medical image registration in-depth;(Shin et al. 2018) uses GAN to synthesize brain tumor images to enhance data and anonymize data, but requires additional training of anatomical structure segmentation network. Besides,(Costa et al. 2017) implements the random generation of vascular annotation maps based on the idea of variational auto-encoder (VAE)(Kingma and Welling 2014), and then synthesizes color retinal images

based on the academic annotation map. Beyond the field of medical image processing, the research of multi-domain translation is very active(Zhao et al. 2018),(Liang et al. 2018),(Choi et al. 2018),(Isola et al. 2017). But in the field of medical image synthesis, there still are many researches of translation synthesis on two modalities(Zhang, Yang, and Zheng 2018),(Nie et al. 2017),(Burgos et al. 2015),(Vemulapalli, Van Nguyen, and Zhou 2015),(Osokin et al. 2017),(Van Nguyen, Zhou, and Vemulapalli 2015),(Kamnitsas et al. 2017), and rare research on multimodal(Chartsias et al. 2018),(Joyce, Chartsias, and Tsaftaris 2017),(Shin et al. 2018).

At present, there are various problems in the research of medical image synthesis, such as the difficulty of expanding the number of modalities, the need to registered training data, relying on complex large networks, the inability to add or retain lesions, the inability to generate from random matrices, and the need of additional training data, etc. Moreover, the evaluation of synthetic data in most studies relies on the evaluation of artificial visual effects by experienced physicians, without objective quantitative evaluation. Therefore, we design a registered multimodal MRI generation scheme based on the Conditional Generative Adversarial Networks (CGAN)(Mirza and Osindero 2014). With unsupervised learning method, training data does not need to be registered. Our solution can receive a random normal distribution as input to generate a set of multimodal registration MRI with setted lesions. We perform multimodal brain MRI generation experiments with tumor segmentation labels on BRATS2015, and verify the effectiveness of lesion information and the availability of synthetic data in tumor lesion segmentation experiments. We will open all the code. In general, our contribution is in the following three areas:

- **Extraction and Generation of Structural Feature Maps** We propose an extraction method for anatomical features for medical images. Without additional anatomical segmentation labels or label extraction training, structural feature maps can be extracted directly from real images of arbitrary modalities . The extraction method can directly obtain the anatomical features of real images, and improve the quality of synthetic images without introducing additional parameters and computation overhead. We also train a structural feature map generator to generate structural feature maps from multidimensional normal distribution. The random generation method can generate rich and diverse structural feature maps indefinitely.
- **Synthesis of Registered Multimodal MRI with Lesion** We use randomly generated structural feature maps to fuse with random lesion labels and then synthesize the registered multimodal MRIs through the generator. We explore a variety of lesion generation guidance methods to achieve effective mapping of lesions based on input lesion labels during multimodal MRI synthesis. In training, no registration data is required except for the lesion label. The synthetic data is registered, and the input lesion label is the lesion label of the synthetic data. Our solution enables fast and easy construction of registered multimodal MRI datasets with lesions label.

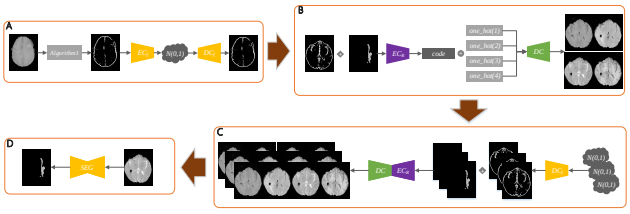

Figure 1: Architecture.

- **Objective Verification Method for Synthetic Data Availability** We construct datasets with different amount of synthetic and real data to train the lesion segmentor, and verify that the synthetic data can be used as pre-training data and enhanced data in the medical image intelligent processing tasks to improve the generalization ability of the model and improve the segmentation precision. Compared with the traditional subjective evaluation method of synthetic image quality, we present the availability of synthetic data in intelligent lesion processing tasks more objectively.

2 Method

We perform our scheme on multimodal brain MRI synthesis task , in which the synthetic lesion is tumors , the lesion processing task is lesion segmentation , and the lesion label is a tumor segmentation label . Our scheme does not limit the human parts , lesion types , lesion processing tasks , specific modality and the number of modalities . We can easily apply this method to other similar tasks through the brain MRI synthesis task.

2.1 Architecture

As shown in Fig. 1, our scheme includes four main stages: structural feature map extraction and generation, multimodal MRI generation, construction of synthetic datasets, and synthetic data availability verification.

In the structural feature map extraction and generation stage, we will obtain a structural feature map generator that can generate structural feature maps from the random normal distribution matrix. Models we train at this stage includes a structural feature map encoder, a structural feature map decoder, a structural feature map discriminator, a code matrix distribution discriminator and a structural feature mask generator.

In the multimodal MRI generation stage, we generate a conditional generator with input of structural feature maps, which can generate MRIs of different modalities according to different one-hot conditional matrixes, and can add lesion labels on the structural feature maps to generate MRI has corresponded lesion information. At this stage, we train a fusion map encoder for structural feature map and lesion label , a lesion segmentor for each modality, an MRI encoder, an MRI decoder, an MRI discriminator and an MRI code discriminator.

In the stage of constructing the synthetic datasets, we use the model produced in the first two stages to generate a sufficient number of structural feature maps from the random

normal distribution matrix and then randomly fuse with real lesion labels, and finally generate the registered multimodal MRI to construct synthetic datasets.

In the synthetic data availability verification stage, we train a lesion segmentor for each MRI modality based on real data, and perform segmentation ability tests on real dataset. Then these segmentors are used to perform segmentation tests on the synthetic data generated by different lesion generation guidance methods. In addition, we use datasets constructed from synthetic data and real data of different amounts to train the lesion segmentor. After training, the segmentation ability test is performed on the real dataset.

2.2 Structural Feature Map Extraction Method

Medical images generated directly from random noise by GAN are often difficult to train and difficult to generate real structural information. We call image that provids basic contour and structure information as structural feature map. For example, a retinal blood vessel distribution map can be regarded as a structural feature map of a retinal image(Costa et al. 2017). Structural feature maps can provide necessary basic guidance for the synthesis of medical images. When synthesizing brain MRI, some studies obtains basic structural information from the brain segmentation label(Shin et al. 2018). However, general structural features such as retinal vascular maps and brain segmentation label require additional data and training before extracting from the original image. To this end, we first design a method for extracting structural feature maps directly from brain MRI, which has the advantages of fast operation, no training, no additional data.

In the traditional digital image processing methods, Roberts operator, Prewitt operator, Sobel operator, etc. are excellent edge detection operators. Sobel operator is often used in processing of brain medical images. As shown in Algorithm 1, We explore a method for further extracting structural feature maps from the edge detection maps generated by Sobel operator.

Algorithm 1 Structural Feature Extraction

Input a real image x and pixel threshold $beta$
$f1 = reduce\_min(sobel(x))$
$f2 = reduce\_max(sobel(x))$
$f1 = mean(f1) - f1$
$f2 = f2 - mean(f2)$
$f1 = ones \times (f1 > beta)$
$f2 = ones \times (f2 > beta)$
$f = ones \times ((f1 + f2) > 0.)$

In Algorithm 1 we use Sobel operator to extract the horizontal and vertical edges detection maps from a real image, each performs maximum reduce and minimum reduce to obtain two edge detection fusion maps, then each fusion map calculate the difference with average pixel value, the two difference maps are binarized according to the set pixel threshold, and the two binary images are summed and then completely binarized. The final result is the structural feature map we need.

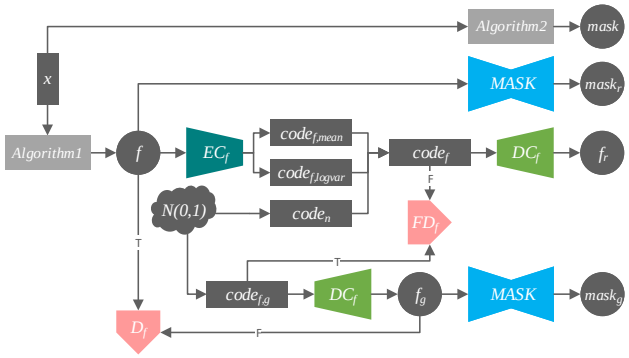

Figure 2: Training of Random Structural Feature Map Generation.

2.3 Structural Feature Map Generation Training

Algorithm 2 Mask Extraction

Input a real image x and expanded pixel value p
$mask = 1.0 - ones \times (x > 0.)$
$new\_size = [x.width() + p, x.length() + p]$
$mask = resize(mask, new\_size)$
$mask = crop\_padding(mask, p)$

When generating the structural feature map, (Shin et al. 2018) still needs to input the real modal image to get the generated structural feature map, which greatly reduces the diversity of generated data. (Costa et al. 2017) implements a method based on VAE for generating retinal blood vessel distribution maps from multidimensional normal distribution. Drawing on this, we design a hybrid network combining the characteristics of VAE and GAN for generating brain structural feature maps from random normal distribution matrixes, which has better diversity and no additional training data. In addition, we train a generator $MASK$ that acquires the brain area masks from the brain structure feature maps for later use to match lesion label. The generator is synchronized with the training of structural feature map generation. During the training of $MASK$, the mask extracted by the real brain MRI x through the mask extraction algorithm (Algorithm 2) is used as label data. As shown in Fig. 2, the specific training processing is as follows:

- The structural feature map f is obtained from x using Algorithm 1, and the mask $mask$ is obtained by the Algorithm 2;
- Encode f with VAE encoder EC_f to get $code_{f,mean}$ and $code_{f,logvar}$, then get a random noise $code_n$ from multidimensional normal distribution $\mathcal{N}(0, 1^2)$, so the approximate normal distribution matrix $code_f = code_{f,mean} + exp(0.5 \times code_{f,logvar}) \times code_n$;
- Decode $code_f$ with VAE decoder DC_f to obtain the reconstructed structural feature map f_r;
- Use mask generator $MASK$ to extract mask $mask_r$ from f;

- Randomly generate a matrix $code_{f,g}$ that obeys normal distribution $\mathcal{N}(0, 1^2)$;
- Decode $code_{f,g}$ with VAE decoder DC_f to get the generated random structure feature map f_g;
- Use mask generator $MASK$ to extract mask $mask_g$ from f;
- Structural feature discriminator D_f identifies f and f_g respectively, the former is a positive sample and the latter is a negative sample;
- Code matrix distribution discriminator FD_f discriminates between $code_{f,g}$ and $code_f$, the former is a positive sample and the latter is a negative sample.

Where, for the constraint of approximate normal distribution matrix, we do not use the original VAE encoder loss, but add a code matrix distribution discriminator to provide adversarial loss for VAE encoder.Meanwhile, we use $L2$ regular loss to guide the mean matrix with a mean of 0 and the variance deviation matrix with a mean of 1. The complete loss items are as follows, where $\omega_{i,j}$ is the weight of each loss item:

- **Code matrix distribution discriminator loss**

$$loss_{FD_f} = \|FD_f(code_{f,g}) - 1\|_2^2 + \|FD_f(code_f)\|_2^2$$

- **Structural feature map discriminator loss**

$$loss_{D_f} = \|D_f(f) - 1\|_2^2 + \|D_f(f_g)\|_2^2$$

- **Adversarial loss**

$$loss_{G_f} = \|FD_f(code_f) - 1\|_2^2 + \|D_f(f_g) - 1\|_2^2$$

- **Code matrix distribution loss**

$$loss_{normal} = \|mean(code_{f,mean})\|_2^2 + \|mean(exp(0.5 \times code_{f,logvar})) - 1\|_2^2$$

where $mean()$ is a mean function.

- **Structural feature map reconstruction loss**

$$loss_{sv} = \|f - f_r\|_2^2 + \|f_r \times mask\|_2^2$$

- **Mask reconstruction loss**

$$loss_{mask} = \|mask - mask_r\|_2^2 + \|f \times mask_r\|_2^2 + \|f_r \times mask_r\|_2^2 + \|f_g \times mask_g\|_2^2$$

2.4 Reconstruction and translation training

We perform reconstruction and transformation training on real data and synchronize with the multi-modal MRI generation training in subsequent chapters to constrain each component to complete our specified tasks in the multimodal MRI generation process. In addition, we also use a code matrix distribution discriminator to guide the consistency of code matrixes distribution between the two training processes. All discriminator components are updated independently, the discriminator training process is shown in Fig. 5

As shown in Fig. 3, when the MRI is reconstructed and translated, the encoder encodes the real MRI x_i of modality i to obtain the semantic feature map $code_i$, then we connect it to different conditional matrixes and decode all the

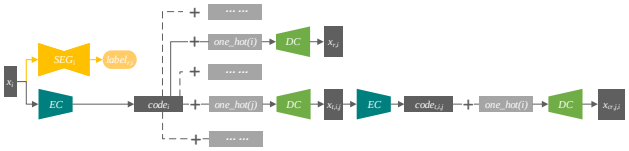

Figure 3: Auxiliary modality reconstruction and modality translation training.

modalities through the decoder. In cycle-reconstruction, we use encoder to re-encode all the obtained translation images, connect all the re-encoded semantic feature maps with the conditional vector of modality i, and finally decode them by decoder to get cycle-reconstruction image $x_{rc,j,i}$. In the above process, we use the lesion label l_i of the original input modality x_i as supervised label for the lesion generation training, and use lesion label generation components for x_i to get $l_{r,i}$.

The detailed losses are as follows, where $x_{t,j,i}$ refers to MRI x_i of modality i translated by modality j; $d_{t,j,i}$ and $c_{t,j,i}$ are the True/False discrimination and category discrimination results of the discriminator for $x_{t,i,j}$; $code_{t,i,j}$ represents the semantic feature map obtained from $x_{t,i,j}$ by encoder:

- **Reconstruction and translation discriminator loss**

$$loss_{D,assist} = \sum_{j=0,j\neq i} \sum_{i=0} (\|d_{t,j,i}\|_2^2 + \|c_{t,j,i} - i\|_2^2)$$

- **Adversarial and category guidance loss**

$$loss_{G,assist} = \sum_{j=0,j\neq i} \sum_{i=0} (\|d_{t,j,i} - 1\|_2^2 + \|c_{t,j,i} - i\|_2^2)$$

- **MRI reconstruction loss**

$$loss_{sv} = \sum_{i=0} (\|x_i - x_{r,i}\|_2^2)$$

- **MRI cycle-reconstruction loss**

$$loss_{cycle} = \sum_{j=0,j\neq i} \sum_{i=0} (\|x_i - x_{cr,j,i}\|_2^2)$$

- **MRI cycle-reconstruction consistency loss**

$$loss_{cycle,consistency} = \sum_{k=0,k\neq j,k\neq i} \sum_{j=0,j\neq i} \sum_{i=0} (\|x_{cr,j,i} - x_{cr,k,i}\|_2^2)$$

- **Semantic consistency loss**

$$loss_{code,consistency} = \sum_{j=0,j\neq i} \sum_{i=0} (\|code_i - code_{t,i,j}\|_2^2)$$

- **Lesion generation loss**

$$loss_{sv,l} = \sum_{i=0} \|label_i - label_{r,i}\|_2^2$$

2.5 Structural feature map and lesion label fusion

First we generate structural feature map f_g, then randomly select the appropriate lesion segmentation label $label$, and then the lesion segmentation label containing n categories is converted into a one-hot matrix $onehot_l$ of n channels,

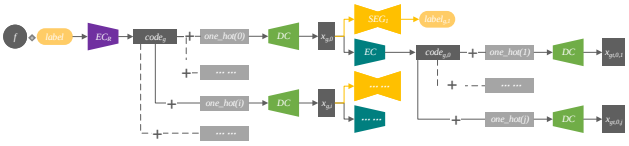

Figure 4: Generation of multimodal MRI.

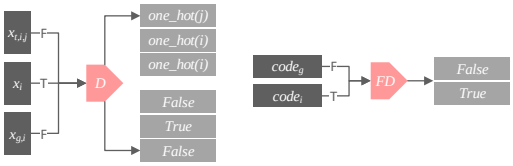

Figure 5: Discriminator training during reconstruction and translation training on real MRI and multimodal MRI generation training.

and each channel corresponds to a segmentation category, the pixel value in each channel is 0 or 1. Each 1 pixel area is registered with each segmentation area in the original segmentation label. At last, we calculate the weighted sum of each channel of $onehot_l$ with f_g, and get a new matrix that fuses the information of f_g and $label$.

If the structural feature map f' is extracted from the random MRI x, then the extracted structural features may contain tumor structure information, which may interfere with the tumor information at random label l and affect the fusion generation MRI, so f' needs to eliminate the tumor information before fusing with the random label $label$, and get the structural feature map f without tumor information, so that the tumor information of the generated image is only derived from the label $label$. We generate a mask without boundary expansion $mask_{l,x}$ for segmentation label $label_x$ of x by the Algorithm 2, then we have $f = mask_{l,x} \times f'$.

Since the location of the randomly selected lesion may appear outside the brain contour of structural feature map, we need to use the Algorithm 2 to obtain the brain region mask $mask$ of the structural feature map. If the product of $mask$ and the selected $label$ is 0, then the tumor label pixel is inside the brain contour of $mask$, which can be adopted, otherwise the $label$ needs to be re-selected.

2.6 Multimodal MRI generation training

We extract the structural feature map f from the real MRI by the Algorithm 1, and randomly select the real lesion label $label$ to fusion. The fusion map contains basic anatomical information and lesion information of the target site. The multimodal MRI generation process is shown in Fig. 4. First, we use a fusion map encoder to encode fusion map to obtain the semantic feature map. The semantic feature map is stacked with different one-hot conditional matrixes and decodes by MRI decoder to obtains synthesis images of different modalities. These synthesis images are then perform modal translation process between each other by MRI encoder and MRI decoder. We use adversarial loss and category guidance loss provided by MRI discriminator to constrain synthesis image to approximate real MRI, and constrain the consistency of all semantic feature maps and translation images by loss, thus ensuring the mutual registration of generated multimodal MRI. In addition, we used the lesion label generation components to segment the tumor lesion segmentation labels from each synthetic MRI to ensure that the generated multimodal images have generated corresponding lesion content based on the input lesion label. The lesion label generation components are only trained by real data in synchronous reconstruction and translation training.

The specific loss items are as follows, where d_i and c_i are the True/False discrimination and category discrimination of the discriminator $D(x_i)$, $d_{g,i}$,$c_{g,i}$ is the output of $D(x_{g,i})$; $x_{g,i}$ is the synthetic image of modality i, $x_{gt,j,i}$ is the translation image of the modality i translated by synthetic image of modality j; $code_g$ is the semantic feature map obtained by fusion map encoderEC_R, $code_{g,i}$ is the semantic feature map encoded from x_i; l is the input label, $l_{g,i}$ is the label obtained by lesion label generation component from $x_{g,i}$; f is the input structural feature map, $f_{g,i}$ is the structural feature map extracted from $x_{g,i}$ by Algorithm 1:

- **Generation Discriminator loss**

$$loss_D = \sum_{i=0}(\|d_i - 1\|_2^2 + \|d_{g,i}\|_2^2 + \|c_i - i\|_2^2 + \|c_{g,i} - i\|_2^2)$$

- **Generation adversarial and category guidance loss**

$$loss_G = \sum_{i=0}(\|d_{g,i} - 1\|_2^2 + \|c_{g,i} - i\|_2^2)$$

- **Structural feature map consistency loss**

$$loss_{sv,f} = \sum_{i=0}(\|f - f_{g,i}\|_2^2)$$

- **Lesion label generation loss**

$$loss_{sv,l} = \sum_{i=0}(\|label - label_{g,i}\|_2^2)$$

- **MRI registration loss**

$$loss_{trans} = \sum_{j=0,j\neq i}\sum_{i=0}(\|x_{g,i} - x_{gt,j,i}\|_2^2)$$

- **Semantic consistency loss**

$$loss_{trans,code} = \sum_{i=0}(\|code_g - code_{g,i}\|_2^2) + \sum_{j=0,j\neq i}\sum_{i=0}(\|code_{g,i} - code_{g,j}\|_2^2)$$

2.7 Lesion label generation guidance method

We design the following three lesion label generation components to provide guidance loss for lesion generation in multimodal MRI generation training:

- **Single segmentor** Each MRI modality is segmented by a common complete segmentor.
- **Single lesion encoder + multiple lesion decoders** Each MRI modality is segmented by a segmentors which is combined by a common lesion encoder and different lesion decoders.
- **multiple segmentors** Each MRI modality is segmented by a different complete segmentor.

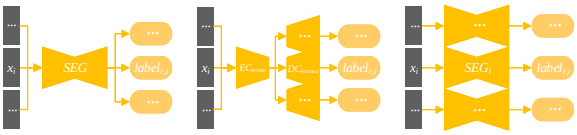

Figure 6: lesion segmentation.

Figure 7: construction of synthetic datasets.

The loss item of the above three schemes in reconstruction and translation training is consistent with the loss item described above. And in the generation training, they only provide the same lesion label generation loss for MRI generation components, no learning.

Besides, we will also independently train the segmentor of these three schemes. In different experiments, we will train the segmentor in each scheme with different amounts of synthetic data or real data. The loss function of the training is as follows, where $label_i$ is the supervised label and $label_{r,i}$ is the segmentation result:

$$loss_{l,independent} = \sum_{i=0} \|label_i - label_{r,i}\|_2^2$$

2.8 Constructing synthetic datasets

As shown in Fig. 7, we can generate any number of structural feature maps from randomly generated normal distribution matrix through the trained structural feature map decoder. Then, we randomly scaled, rotated, translated, flipped, etc. the original set of labels to get a random lesion label set. We then fuse the generated structural feature map with the randomly selected label from the random lesion label set. Like training, we can select suitable labels by obtaining mask from structural feature map through the mask generator $MASK$.

Due to the influence of the operation that removing tumor information during training, there are some structural feature maps with poor quality that brain contour is not closed, so we design a structural feature map filtering algorithm for this. First, we use the generator $MASK$ to generate the mask of structural feature map. Then, we perform Gaussian blur(Wink and Roerdink 2004) on the structural feature map, and use the contour search algorithm and filling algorithm provided by OpenCV[TODO] to obtain all the closed contours of the Gaussian blur image and fill them. So we get another mask by the traditional algorithm. Finally, we calculate the Mean Absolute Error (MAE) of the two masks. If the MAE is lower than the threshold we set, that means the main brain contour of the structural feature map is relatively complete, and the feature map can be used; otherwise, it means that the mask obtained by the traditional algorithm is hollow inside and quite different with the mask generated by the generator, so it needs to be regenerated. The algorithm is expressed as Algorithm 3.

In the generated multimodal MRIs, there still are samples with poor lesion generation. At this point, we segment our synthetic MRI data through the lesion segmentor pre-trained on real data.We filter according to the segmentation evaluation score(default threshold is 0.95 on dice score). After multiple filtering, we obtain the final synthetic dataset consisting of structural feature maps, masks, lesion segmentation labels, and multimodal MRI.

Algorithm 3 Structural feature map filtering

Input a MAE threshold mae
function $get\_mask(f)$
    $contours = OpenCV.findContours(f)$
    $mask = OpenCV.drawContours(f, contours)$
    return $mask$
end function

do
    $f = DC_f()$
    $m = MASK(f)$
    $m' = get\_mask(f)$
while $MAE(m', m) <= mae$

3 Experiments

3.1 BRATS2015 dataset

We use the open dataset BRATS2015 for experiments, which has four registered modalities of T1/T2/T1c/Flair. The training dataset contains 274 3D MRIs per modality, with the size of 155×240×240, and 274 tumor segmentation labels of the same size. We divide the sample into a training set and a testing set by 9:1, and construct a 2D data set from 50 slices of each 55-105 of the 3D MRI. In data preprocessing, we standardized each image.

3.2 BRATS synthetic dataset

We constructed a registered synthetic dataset with tumor labels containing four modalities of T1, T2, T1c, and Flair using the method in 2.8. The size of the synthetic dataset sample is consistent with the BRATS2015 dataset, but the number of samples can be any number according to the needs of the experiment.

3.3 BRATS Enhanced dataset

We performed random scaling, rotation, translation, flipping, etc. on the original BRATS2015 dataset to obtain enhanced data. The size of the enhanced dataset sample is consistent with the BRATS2015 dataset, but the number of samples can be any number according to the needs of the experiment.

3.4 Training settings

The number of iterations of each experiment is equal to 100 epochs of the BRATS2015 training dataset. The learning rate is 1e-4 without weight decay. we use Adam optimizer with beta1 of 0.5 and 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

Table 1: Lesion segmentation result

synthetic method	test data type	MSE	Dice Score
-	real	0.026	0.915
1SEG	synthetic	0.053	0.741
1ECL+4DCL	synthetic	0.055	0.808
4SEG	synthetic	0.043	0.838

parameter. Specifically, for a convolutional layer with a convolution kernel size of $[k, k, f]$, we initialize the $k \times k \times f$ convolution kernel parameters of this layer to $1/(k \times k \times f)$, and the bias is 0. In discriminator, we use a random normal initializer with mean of 0 and standard deviation of 0.2, and the bias is 0. We used Dice Score (Dice 1945) and Mean Square Error (MSE)(Prasad and Rao 1990) to evaluate the segmentation results. The evaluation results are the average of the evaluation results of the 2D images, and each experiment is trained four times to retain the best results.

3.5 Contrast experiment of lesion generation method

We used the training set of the processed BRATS2015 dataset to fully train the lesion segmentor for the same iteration step, and then we selected the test set of the BRATS2015 dataset and the unfiltered synthetic dataset by segmentor from different lesion generation component methods. Except for the difference in the source of the test data, other conditions such as the sample size of the test data are identical. Among them, the segmentation method we adopt on the real data is the multiple segmentors method in the section 2.7, that is, each modality trains a separate segmentor.

3.6 verification experiment of Synthetic data availability

As shown in the table, we mixed real BRATS2015 training data with BRATS synthetic data in different amounts, then used the mixed data set for segmentation training, and finally evaluated the segmentation ability of the model on real BRATS2015 test data. All experiments were fully trained with the same number of iterations. Except for the difference in the source of the training data, other conditions such as the sample size of the test data are identical. At the same time, as a comparison, we also conduct a separate training of synthetic data, mixed training of real data and normal enhancement data. We set up three data mixing modes: random mixing, real first, and synthetic first.

4 Results

4.1 Quantitative result

Lesion segmentation result As shown in the table 1, after the BRATS2015 training data is fully trained by the same iteration step, the segmentation test results reach the MSE of 0.026 and the Dice Score of 0.915 on the real test data set. Then we use this excellent segmentor to segment our unfiltered synthetic data. On the synthetic dataset with the same amount of data as the real test dataset, three different lesion label generation component methods have achieved good segmentation results. As a result, the method in which each modality trains a separate segmentor achieves the best results, and the Dice Score also reaches 0.838.

Table 2: Verification experiment results for synthetic data availability

Num	real data	synthetic data	Enhanced data	mixing modes	MSE	Dice Score
1	15070	0	0	-	0.026	0.915
31	15070× 0.5	0	0	-	0.032	0.902
2	0	15070	0	-	0.205	0.708
3	0	15070× 2	0	random mixing	0.206	0.736
4	0	15070× 3	0	random mixing	0.205	0.754
7	15070× 0.1	15070	0	synthetic first	0.031	0.908
8	15070× 0.1	15070× 2	0	synthetic first	0.028	0.907
9	15070× 0.1	15070× 3	0	synthetic first	0.030	0.907
13	15070× 0.2	15070× 0.8	0	random mixing	0.041	0.850
12	15070× 0.5	15070× 0.5	0	random mixing	0.031	0.904
14	15070× 0.8	15070× 0.2	0	random mixing	0.024	0.935
32	15070	15070× 0.2	0	random mixing	0.025	0.921
33	15070	15070× 0.5	0	random mixing	0.023	0.939
34	15070	15070× 0.8	0	random mixing	0.026	0.916
15	15070	15070	0	random mixing	0.027	0.913
18	15070	15070× 2	0	random mixing	0.033	0.901
19	15070	15070× 3	0	random mixing	0.034	0.897
35	15070	0	15070× 0.2	random mixing	0.027	0.911
36	15070	0	15070× 0.5	random mixing	0.025	0.927
37	15070	0	15070× 0.8	random mixing	0.026	0.920
22	15070	0	15070	random mixing	0.026	0.915
23	15070	0	15070× 2	random mixing	0.032	0.898
24	15070	0	15070× 3	random mixing	0.036	0.885
16	15070	15070	0	real first	0.195	0.795
17	15070	15070	0	synthetic first	0.021	0.940

Verification experiment results for synthetic data availability TO DOAs shown in table 2, we have been trained on data sets with different components and quantities and then evaluated on real test sets to get the results in the table.In experiment 1, after the full training of 100 epochs of BRATS2015 training data, the segmentation test results reached 0.026 MSE and 0.915 Dice Score on the real test data set.According to the training results of experiment 2-4 using synthetic data alone in the table, there is still a certain gap between the training results of the synthetic data and the real data, which indicates that the synthetic data cannot completely replace the real data as the training set.The results of experiment 7-9 show that pre-training with a large amount of synthetic data and fine-tuning on a small amount of real data can achieve a result very close to that of training with real data in experiment 1. This indicates that the synthesized data is very suitable as a pre-training data set.Experiment 12-14 a percentage of the real data and synthetic data random mixing proportion of different result difference is very big, when both scale similar segmentation result with experiment 1 were similar, synthesis of high data rate, the result is lower than the experiment 1, synthetic data rate is low, can provide the generalization ability of the model, obtained is higher than the result of experiment 1.Experiment a further attempt to 32-34, 18th - 19 in the total amount of real data synthesis data of adding different amount of data, we found that fewer synthetic data can improve the generalization ability of the model, have the effect of data enhancement, synthetic data enhancement effect, the better, the more but when synthetic data to increase again after a certain proportion will have the opposite effect, the synthetic data, the more the influence of the real data.Experiment 35 to 37 and 22 to 24, we use the enhancement data generated by a data enhancement method to enhancement effect compared with

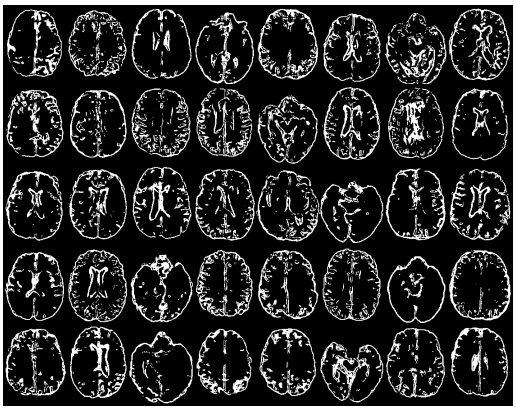

Figure 8: Synthetic structural feature map.

the synthetic data, we found the data volume increased and enhanced effect on the change tendency of the two is consistent, but the difference in specific enhancement effect as the change of enhanced data volume change curve there are some differences. Overall, in terms of the sensitivity of the model to the amount of enhanced data, the enhanced data is more robust, but the upper limit of the enhancement effect that can be achieved by the composite data is much higher than that of the former strong data.In experiment 16-17, we compared experiment 15 and found that the synthesized data with the same amount of real data had the best performance as pre-training data before real data training, the mixed training effect as enhanced data and real data was in the middle, and the training performance after real data set as supplementary training data was very poor.

TO DOAsIn general, we find that when the amount of real data is large, a small amount of composite data can be used as mixed enhancement data, or a large amount of composite data can be used for pre-training before training on real data sets. When real data is scarce, large amounts of synthetic data can be used for pre-training, and then fine-tuned on a small amount of real data to obtain results that compete with the training of completely real data. This is consistent with the conclusion of (Shin et al. 2018). We do not recommend full use of synthetic data for training, and contrary to the conclusion of (Shin et al. 2018), we do not recommend supplementary training using synthetic data either.

4.2 Synthetic images

TO DOAsFigure 8 shows an example of the structural feature graph we generated from the random normal distribution matrix. In figure 9, we show several examples of structural feature maps generated from random normal distribution matrices, corresponding masks generated from structural feature maps, randomly selected lesion segmentation labels, and multi-modal MRI generated from structural feature maps and lesion segmentation labels.

5 Conclusion and future work

Based on the conditional generation antagonism network, we realized the generation of registered multimodal MRI

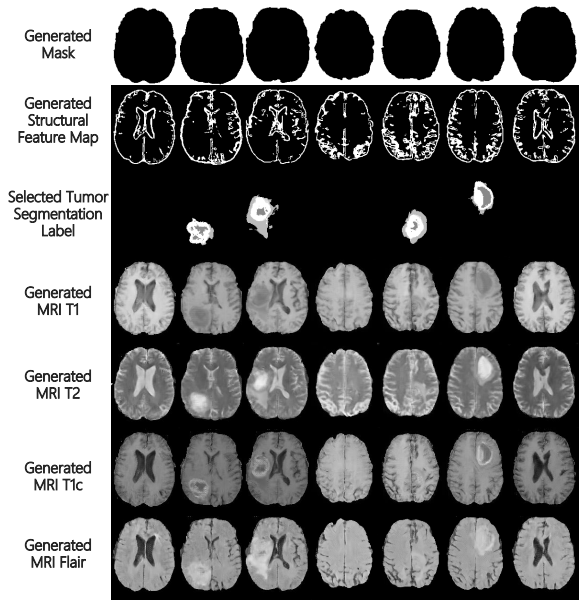

Figure 9: Multimodal MRI generated from the random structural feature map and lesion label.

from the random normal distribution matrix through unsupervised training, and could add the lesion information freely. We verified through lesion segmentation experiments that synthetic MRI can be used as pre-training data or enhanced data for intelligent medical image processing tasks and can significantly improve the generalization ability of the model. In this paper, our contributions include the following:

- We propose a structural feature map extraction method to extract anatomical structure information directly from medical images without training or additional label data;
- We propose a random structure feature map generation method to generate structural feature maps from multi-dimensional normal distribution sampling;
- We realized the generation of registered multimodal MRI with corresponding lesion information from structural feature map and randomly selected lesion labels;
- We verified that synthetic data can be used as pre-training data or enhanced data for intelligent medical image processing tasks through a number of data usability tests on our synthetic dataset.

In the future, we will further improve our method in CT, PET and other modes, in heart, lung and other parts, in detection, classification and other lesion processing tasks. We are committed to further simplifying the training process and synthesizing higher quality medical images.

6 Acknowledgments

Thanks for the computing support provided by NSCCGZ.

References

Burgos, N.; Cardoso, M. J.; Guerreiro, F.; Veiga, C.; Modat, M.; Mcclelland, J.; Knopf, A.; Punwani, S.; Atkinson, D.; Arridge, S. R.; et al. 2015. Robust ct synthesis for radiotherapy planning: Application to the head and neck region. *medical image computing and computer assisted intervention* 476–484.

Chartsias, A.; Joyce, T.; Giuffrida, M. V.; and Tsaftaris, S. A. 2018. Multimodal mr synthesis via modality-invariant latent representation. *IEEE Transactions on Medical Imaging* 37(3):803–814.

Choi, Y.; Choi, M.; Kim, M.; Ha, J.; Kim, S.; and Choo, J. 2018. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. *computer vision and pattern recognition* 8789–8797.

Costa, P.; Galdran, A.; Meyer, M. I.; Abramoff, M. D.; Niemeijer, M.; Mendonca, A. M.; and Campilho, A. 2017. Towards adversarial retinal image synthesis. *arXiv: Computer Vision and Pattern Recognition*.

Dice, L. R. 1945. Measures of the amount of ecologic association between species. *Ecology* 26(3):297–302.

Goodfellow, I. J.; Pougetabadie, J.; Mirza, M.; Xu, B.; Wardefarley, D.; Ozair, S.; Courville, A. C.; and Bengio, Y. 2014. Generative adversarial nets. *neural information processing systems* 2672–2680.

Huang, Y.; Shao, L.; and Frangi, A. F. 2017. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. *computer vision and pattern recognition* 5787–5796.

Isola, P.; Zhu, J.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. *computer vision and pattern recognition* 5967–5976.

Joyce, T.; Chartsias, A.; and Tsaftaris, S. A. 2017. Robust multi-modal mr image synthesis. *medical image computing and computer assisted intervention* 347–355.

Kamnitsas, K.; Baumgartner, C. F.; Ledig, C.; Newcombe, V.; Simpson, J. P.; Kane, A. D.; Menon, D. K.; Nori, A. V.; Criminisi, A.; Rueckert, D.; et al. 2017. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. *information processing in medical imaging* 597–609.

Kingma, D. P., and Welling, M. 2014. Auto-encoding variational bayes. *international conference on learning representations*.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. *neural information processing systems* 141(5):1097–1105.

Liang, X.; Zhang, H.; Lin, L.; and Xing, E. P. 2018. Generative semantic manipulation with mask-contrasting gan. *european conference on computer vision* 574–590.

Miao, S.; Piat, S.; Fischer, P. W.; Tuysuzoglu, A.; Mewes, P.; Mansi, T.; and Liao, R. 2018. Dilated fcn for multi-agent 2d/3d medical image registration. *national conference on artificial intelligence* 4694–4701.

Mirza, M., and Osindero, S. 2014. Conditional generative adversarial nets. *Computer Science* 2672–2680.

Nie, D.; Trullo, R.; Lian, J.; Petitjean, C.; Ruan, S.; Wang, Q.; and Shen, D. 2017. Medical image synthesis with context-aware generative adversarial networks. *medical image computing and computer assisted intervention* 417–425.

Osokin, A.; Chessel, A.; Salas, R. E. C.; and Vaggi, F. 2017. Gans for biological image synthesis. *international conference on computer vision* 2252–2261.

Prasad, N. G. N., and Rao, J. N. K. 1990. The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association* 85(409):163–171.

Shin, H.; Tenenholtz, N. A.; Rogers, J. K.; Schwarz, C. G.; Senjem, M. L.; Gunter, J. L.; Andriole, K. P.; and Michalski, M. 2018. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. *arXiv: Computer Vision and Pattern Recognition* 1–11.

Van Nguyen, H.; Zhou, K. S.; and Vemulapalli, R. 2015. Cross-domain synthesis of medical images using efficient location-sensitive deep network. *medical image computing and computer assisted intervention* 677–684.

Vemulapalli, R.; Van Nguyen, H.; and Zhou, S. K. 2015. Unsupervised cross-modal synthesis of subject-specific scans. *international conference on computer vision* 630–638.

Wink, A. M., and Roerdink, J. B. T. M. 2004. Denoising functional mr images: a comparison of wavelet denoising and gaussian smoothing. *IEEE Transactions on Medical Imaging* 23(3):374–387.

Zhang, Z.; Yang, L.; and Zheng, Y. 2018. Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network. *computer vision and pattern recognition* 9242–9251.

Zhao, B.; Chang, B.; Jie, Z.; and Sigal, L. 2018. Modular generative adversarial networks. *european conference on computer vision* 157–173.

Zhu, J.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *international conference on computer vision* 2242–2251.