Synthesis of Registered Brain Multimodal MRI with Lesions

Yili Qu and **Wanqi Su** and **Chufu Deng** and **Ying Wang**
and **Yutong Lu** and **Zhiguang Chen** and **Nong Xiao** [1]

Abstract. In a large number of data-driven medical image intelligent processing tasks, the collection and acquisition of medical image data is very difficult, especially the registered multimodal medical image data. Synthetic medical image data can alleviate the problem of insufficient data well. In this paper, based on the unsupervised conditional GAN model, we achieve the synthesis of registered multimodal medical images from random normal distribution matrix and the corresponding lesion information can be efficiently generated based on the freely selected lesion label. We conduct 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 generalization ability of the model.

1 INTRODUCTION

MRI is a common medical imaging that has multiple modalities depending on imaging parameters, such as T1 and T2. Different modalities have different reference values for doctors. To make accurate judgments, doctors often need multiple modal images as a comparison. 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 CNN[12] or GAN[6].

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, 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 synthesize 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 image modality translate to other modalities[1, 7, 22, 21]. After GAN has shown its powerful generation ability, there are many researches on medical image translation based on GAN[24,

17, 18, 21, 10]. Many studies use GAN to achieve higher translation quality[25, 13, 26, 3]. Recent studies have implemented the translation on unregistered multimodal data[24, 9]. GAN has been gradually applied to various organs, such as brain MRI to CT translation[17, 10], retinal vascular annotation to image translation[4], cardiac MRI to CT translation and segmentation using unsupervised CycleGAN[26, 17]. For multimodal image synthesis, [2] implements multi-input and multi-output MRI synthesis, but registration is required for input multimodal data. [9] further implements an unregistered multi-input synthesis model, which can perform image synthesis from any subset of its input, but limits the output to unimodal. [15] studies for medical image registration in-depth. [20] uses GAN to synthesize brain tumor images to enhance data and anonymize data, but requires additional training of anatomical structure segmentation network. Besides,[4] implements the random generation of vascular annotation maps based on the idea of variational auto-encoder (VAE)[11], and then generates color retinal images based on the vascular annotation maps. Beyond the field of medical image processing, the research of multi-domain translation is very active[25, 13, 3, 8]. But in the field of medical image synthesis, there are still many researches of translation synthesis on two-modal[24, 17, 1, 22, 18, 21, 10], and rare research on multimodal[2, 9, 20].

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 synthesize from random matrix, and the need of additional training data, etc. Therefore, we design a registered multimodal MRI synthesis scheme based on the Conditional Generative Adversarial Networks (CGAN)[16]. With unsupervised learning method, training data does not need to be registered. Our solution can receive a random normal distribution matrix as input to synthesize a set of registered multimodal MRI with setted lesions. We perform multimodal brain MRI synthesis 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 the code. In general, our main work includes the following three aspects:

- Extraction and generation of structural feature map
  We propose an extraction method for anatomical structure features of medical images. Without additional anatomical segmentation labels or label extraction training, structural feature maps can be extracted directly from arbitrary real images. We also train a structural feature map generator to generate structural feature maps from multidimensional normal distribution. Thus we can gener-

---

ate rich and diverse structural feature maps indefinitely.

- Synthesis of registered multimodal MRI with lesions
We use randomly generated structural feature maps to fuse with random lesion labels and then synthesize the registered multimodal MRI 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 registered data is required except for the lesion labels.

- Objective verification for synthetic data availability
We construct datasets with different amounts 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 medical image intelligent processing tasks to improve generalization ability of the model and improve segmentation precision. So we can present the availability of synthetic data in intelligent lesion processing tasks objectively.

2 METHOD

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

2.1 Architecture

As shown in Fig. 1, our scheme includes four main stages. In the structural feature map extraction and generation stage, we will obtain a structural feature map generator that can generate structural feature maps from random normal distribution matrix. In the multimodal MRI synthesis stage, we train a conditional generator with input of structural feature maps and lesion labels, which can synthesize MRI of different modalities according to different conditionals. In the synthetic datasets construction stage, we use the models produced in the previous stages to synthesize registered multimodal MRI from random normal distribution matrixes. In the synthetic data availability verification stage, we train and test the lesion segmentors on different datasets constructed from different amounts of synthetic data and real data.

2.2 Structural feature map extraction method

Medical images generated directly from random noise by GAN are difficult to generate realistic structural information. We call the image that provides basic contour and structure information as structural feature map. For example, a retinal vascular distribution map can be regarded as a structural feature map of a retinal image[4]. Structural feature maps can provide necessary basic guidance for the synthesis of medical images. When synthesizing brain MRI, some studies obtain basic structural information from the brain segmentation labels[20]. However, general structural features such as retinal vascular maps and brain segmentation labels 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 to further extract structural feature maps from the edge detection maps generated by Sobel operator.

Algorithm 1 Structural feature map extraction

1: Input a real image x and pixel threshold $beta$
2: $f1 = reduce\_min(sobel(x))$
3: $f2 = reduce\_max(sobel(x))$
4: $f1 = mean(f1) - f1$
5: $f2 = f2 - mean(f2)$
6: $f1 = ones \times (f1 > beta)$
7: $f2 = ones \times (f2 > beta)$
8: $f = ones \times ((f1 + f2) > 0.)$

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

2.3 Structural feature map generation training

Algorithm 2 Mask Extraction

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

When generating the structural feature map, [20] still needs to input real image to get the generated structural feature map, which greatly reduces the diversity of synthetic data. [4] 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 MG 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 MG, the mask extracted by 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, the mask $mask$ is obtained from x 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 MG to extract mask $mask_r$ from f;
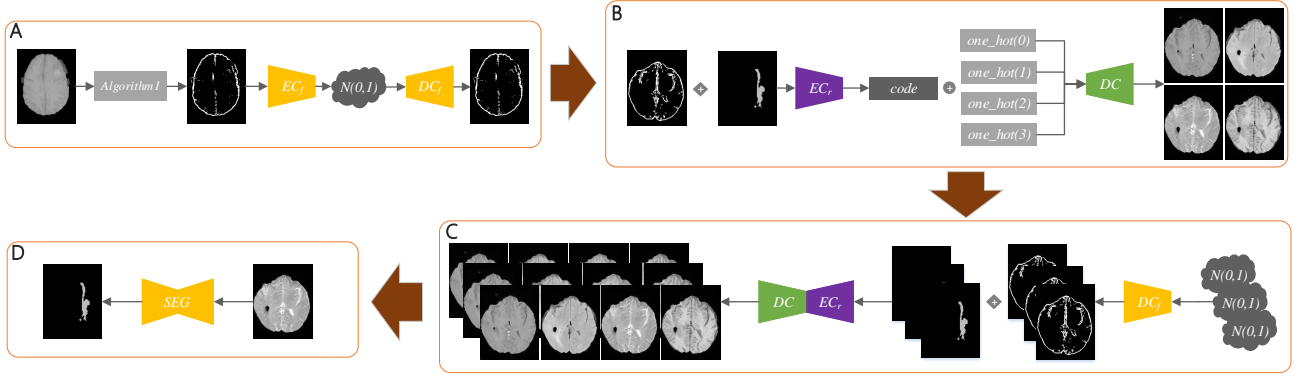- Randomly generate a matrix $code_{f,g}$ that obeys normal distribution $\mathcal{N}(0, 1^2)$;
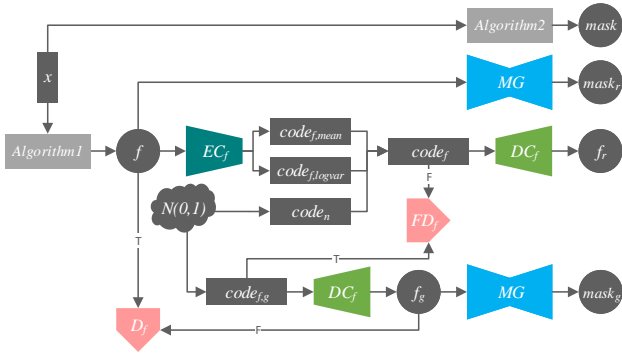
Figure 1. Architecture.

Figure 2. Structural feature map generation training.

- Decode $code_{f,g}$ with VAE decoder DC_f to get the generated random structure feature map f_g;
- Use mask generator MG 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 identifies $code_{f,g}$ and $code_f$, the former is a positive sample and the latter is a negative sample.

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, $mean()$ is the mean function:

- Code matrix distribution discriminator loss

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

- Structural feature map discriminator loss

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

- Structural feature map adversarial loss

$$loss_{g,f} = \|FD_f(code_f) - 1\|_2^2 + \|D_f(f_g) - 1\|_2^2 \quad (3)$$

- Code matrix distribution loss

$$loss_{normal} = \|mean(code_{f,mean})\|_2^2 + \\ \|mean(exp(0.5 \times code_{f,logvar})) - 1\|_2^2 \quad (4)$$

- Structural feature map reconstruction loss

$$loss_{reconstruction,f} = \|f - f_r\|_2^2 + \|f_r \times mask\|_2^2 \quad (5)$$

- Mask reconstruction loss

$$loss_{reconstruction,mask} = \|mask - mask_r\|_2^2 \\ + \|f \times mask_r\|_2^2 + \|f_r \times mask_r\|_2^2 + \|f_g \times mask_g\|_2^2 \quad (6)$$

2.4 Reconstruction and translation training

We perform reconstruction and translation training on real data and synchronize with multimodal MRI synthesis training in subsequent sections. These trianings constrain each component to complete our specified tasks in multimodal MRI synthesis process. In addition, we also use a code matrix distribution discriminator FD 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.
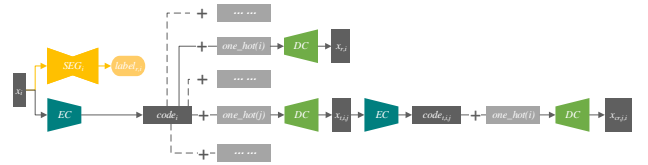

Figure 3. Reconstruction and translation training.

As shown in Fig. 3, when the MRI is reconstructed and translated, encoder EC encodes real MRI x_i of modality i to obtain semantic feature map $code_i$, then it is concatenated to different conditional matrixes and obtain all the modalities through decoder DC. In cycle-reconstruction, we use encoder to re-encode all the obtained translation images, concatenate all re-encoded semantic feature maps with 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 original input modality x_i as the supervised label for 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 discriminator D for $x_{t,i,j}$; $code_{t,i,j}$ represents semantic feature map obtained from $x_{t,i,j}$ by encoder:

- Discriminator loss

$$loss_{d,1} = \sum_{j=0,j\neq i} \sum_{i=0} (\|d_{t,j,i}\|_2^2 + \|c_{t,j,i} - i\|_2^2)$$
$$+ \sum_{i=0} (\|FD(code_i) - 1\|_2^2) \quad (7)$$

- Adversarial and category guidance loss

$$loss_{g,1} = \sum_{j=0,j\neq i} \sum_{i=0} (\|d_{t,j,i} - 1\|_2^2 + \|c_{t,j,i} - i\|_2^2) \quad (8)$$

- MRI reconstruction loss

$$loss_{reconstruction} = \sum_{i=0} (\|x_i - x_{r,i}\|_2^2) \quad (9)$$

- MRI cycle-reconstruction loss

$$loss_{cycle} = \sum_{j=0,j\neq i} \sum_{i=0} (\|x_i - x_{cr,j,i}\|_2^2) +$$
$$\sum_{k=0,k\neq j,k\neq i} \sum_{j=0,j\neq i} \sum_{i=0} (\|x_{cr,j,i} - x_{cr,k,i}\|_2^2) \quad (10)$$

- Semantic consistency loss

$$loss_{code,1} = \sum_{j=0,j\neq i} \sum_{i=0} (\|code_i - code_{t,i,j}\|_2^2) \quad (11)$$

- Lesion segmentation loss

$$loss_{lesion,1} = \sum_{i=0} \|label_i - label_{r,i}\|_2^2 \quad (12)$$

2.5 Fusion of structural feature map and label

After generating structural feature map f_g, we randomly select appropriate lesion segmentation label $label$, and the lesion segmentation label containing n categories is converted into a one-hot matrix $onehot_l$ of n channels. 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 original label segmentation area. At last, we calculate the weighted sum of each channel of $onehot_l$ with f_g, and get a new matrix that fuses information of f_g and $label$.

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

Since the location of randomly selected lesion may appear outside the brain contour of structural feature map, we need to use Algorithm 2 to obtain the brain region mask $mask$ of structural feature map. If the product of $mask$ and 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.
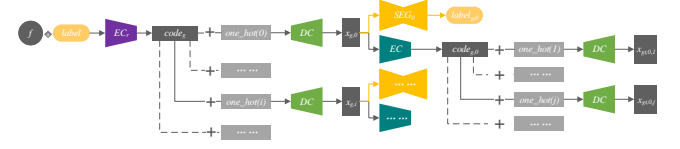
2.6 Multimodal MRI synthesis training

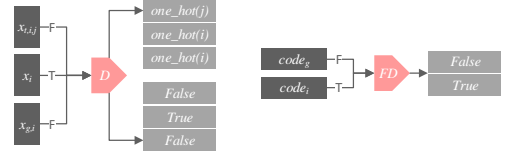**Figure 4.** Synthesis of multimodal MRI.

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

We extract the structural feature map f from real MRI by Algorithm 1, and randomly select the real lesion label $label$ for fusion. The fusion map contains basic anatomical information and lesion information of the target site. The multimodal MRI synthesis process is shown in Fig. 4. First, we use a fusion map encoder EC_r to encode fusion map and obtain the semantic feature map. The semantic feature map is concatenated to different one-hot conditional matrixes as input to MRI decoder, then obtains synthetic image of different modalities. These synthetic images are then performing 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 synthetic 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 synthetic multimodal MRI. In addition, we use the lesion label generation components to segment the tumor lesion segmentation labels from each synthetic MRI to ensure that the synthetic multimodal images have synthesized corresponding lesion content based on the input lesion label. The lesion label generation components are only trained by real data in 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 synthetic 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:

- Discriminator loss

$$loss_{d,2} = \sum_{i=0}(\|d_i - 1\|_2^2 + \|d_{g,i}\|_2^2 + \|c_i - i\|_2^2 + \\ \|c_{g,i} - i\|_2^2) + \|FD(code_g)\|_2^2 \qquad (13)$$

- Adversarial and category guidance loss

$$loss_{g,2} = \sum_{i=0}(\|d_{g,i} - 1\|_2^2 + \|c_{g,i} - i\|_2^2) + \|FD(code_g) - 1\|_2^2 \qquad (14)$$

- Structural feature map consistency loss

$$loss_f = \sum_{i=0}(\|f - f_{g,i}\|_2^2) \qquad (15)$$

- Lesion generation loss

$$loss_{lesion,2} = \sum_{i=0}(\|label - label_{g,i}\|_2^2) \qquad (16)$$

- MRI registration loss

$$loss_{registration} = \sum_{j=0,j\neq i}\sum_{i=0}(\|x_{g,i} - x_{gt,j,i}\|_2^2) \qquad (17)$$

- Semantic consistency loss

$$loss_{code,2} = \sum_{i=0}(\|code_g - code_{g,i}\|_2^2) + \\ \sum_{j=0,j\neq i}\sum_{i=0}(\|code_{g,i} - code_{g,j}\|_2^2) \qquad (18)$$

2.7 Lesion generation guidance method

As shown in Fig. 6, we design the following three lesion label generation components to provide guidance loss for lesion generation in multimodal MRI synthesis training:

1. Single segmentor
   Each MRI modality is segmented by a common complete segmentor SEG.
2. Single lesion encoder + multiple lesion decoders
   Each MRI modality is segmented by a segmentor which is combined by a common lesion encoder EC_l and different lesion decoders $EC_{l,i}$.
3. Multiple segmentors
   Each MRI modality is segmented by a different complete segmentor SEG_i.

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

Besides, we also train the segmentors of these three methods independently. In different experiments, we will train the segmentor in each method 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:

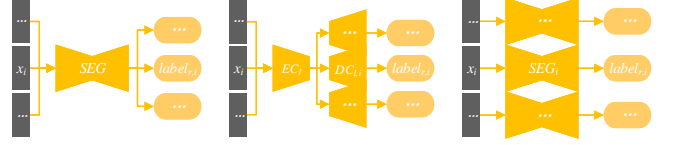$$loss_{segmentation} = \sum_{i=0}\|label_i - label_{r,i}\|_2^2 \qquad (19)$$

Figure 6. Lesion segmentation.

Figure 7. Construction of synthetic datasets.

2.8 Construction of synthetic datasets

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

Due to the influence of the operation that removing tumor information in 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 MG to generate the mask of structural feature map. Then, we perform Gaussian blur[23] on the structural feature map, and use the contour search algorithm and filling algorithm provided by OpenCV[2] 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 traditional algorithm is hollow inside and quite different with the mask generated by generator, so it needs to be regenerated. The algorithm is expressed as Algorithm 3.

Algorithm 3 Structural Feature Map Filtering

1: Input a MAE threshold mae
2: **function** $get\_mask(f)$
3:      $contours = OpenCV.findContours(f)$
4:      $mask = OpenCV.drawContours(f, contours)$
5:      **return** $mask$
6: **do**
7:      $f = DC_f()$
8:      $m = MG(f)$
9:      $m' = get\_mask(f)$
10: **while** $MAE(m', m) <= mae$

In the systhetic multimodal MRI dataset, there are still samples with poor quality of 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

[2] https://opencv.org/

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.

3 EXPERIMENTS

3.1 BRATS2015 dataset

We use the open dataset BRATS2015[14] 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 $155\times240\times240$, and 274 tumor segmentation labels of the same size. We divide the sample into a training set and a test 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 have standardized each image.

3.2 BRATS synthetic dataset

We construct a registered synthetic dataset with tumor labels containing four modalities of T1/T2/T1c/Flair using the method in Section 2.8. The size of the synthetic dataset samples are 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 perform 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 iteration number 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. We use Dice Score [5] and Mean Square Error (MSE)[19] to evaluate the segmentation results. The evaluation results are the average of the 4 modal results on the 2D images, and each experiment is trained four times to retain the best results.

3.5 Experiments of lesion generation methods

As shown in Table 1, we train 4 tumor lesion segmentors using the multiple segmentors method on BRATS2015 training dataset and test them on BRATS2015 test dataset. Then we use the trained segmentors to segment the unfiltered synthetic data from different lesion generation guidance methods.

3.6 Experiments of synthetic data availability

As shown in Table 2, we mix real BRATS2015 training data with BRATS synthetic data in different amounts, then use the mixed dataset for segmentation training using the multiple segmentors method, and evaluate the segmentation ability of the model on real BRATS2015 test dataset. All experiments have been fully trained with the same number of iterations, which equal to the number of 100 epochs on BRATS2015 training dataset. At the same time, we set up three data mixing modes: random mixing, real data training first, and synthetic data training first.In real first experiments and synthetic

first experiments, the training iteration number of data from different sources is proportional to the amount of data. Except the conditions in tabel, other conditions are identical.
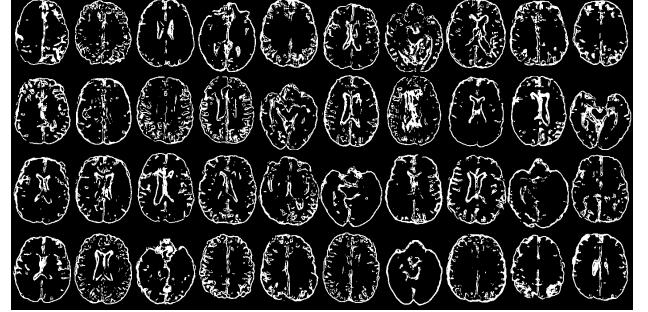
4 RESULTS

Figure 8. Synthetic structural feature maps.

Fig. 8 shows the examples of the structural feature map generated from random normal distribution matrix. Fig. 9 shows the examples of images obtained from all stages.
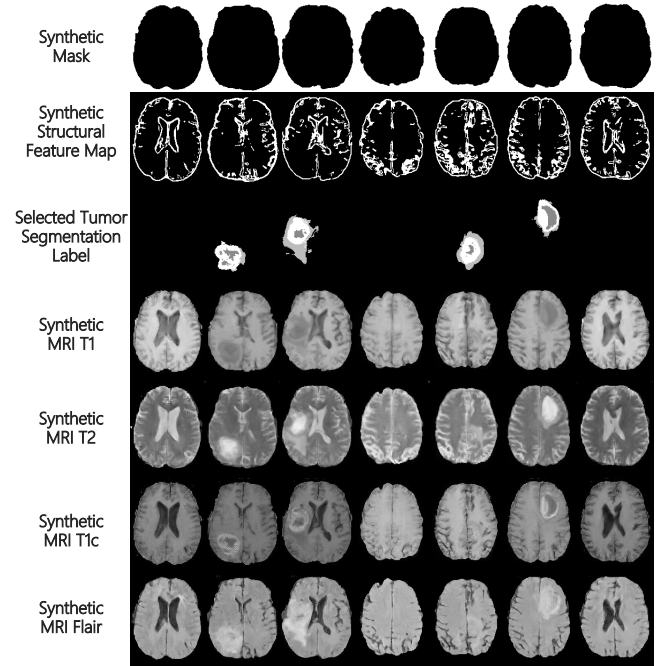

Figure 9. Multimodal MRIs systhesized from structural feature maps and lesion labels.

4.1 Results of lesion generation methods

As shown in Table 1, segmentation test results of the segmentor trained on real data have reached MSE of 0.026 and Dice Score of 0.915. From results, three different lesion label generation methods

Table 1. Lesion generation methods experiments.

segmentation schemes	testing dataset	MSE	Dice Score
$4\,SEG$	real	0.026	0.915
$1\,SEG$	synthetic	0.053	0.741
$1\,EC_l + 4\,DC_l$	synthetic	0.055	0.808
$4\,SEG$	synthetic	**0.043**	**0.838**

have achieved good segmentation results, where the multiple segmentors method is the best.

4.2 Results of Synthetic Data Availability

As shown in Table 2, the results of experiment NO.3-NO.5 show that the synthetic data can not completely replace real data in training. The results of NO.6-NO.8 show that the performances of pre-training with a large amount of synthetic data and fine-tuning on a small amount of real data are similar to those of training with complete real data. In NO.9-NO.11, the results of different mixing ratios are also quite different. When the two ratios are close, the segmentation results are similar to NO.1. When the proportion of synthetic data is high, the synthetic data will interfere with the learning of real data, and the result is lower than that of NO.1. When the proportion of synthetic data is low, the generalization ability of the model can be improved by synthetic data, and the result is higher than NO.1. In NO.12-NO.17, we further try to add synthetic data of different quantities to real data, which show that adding a small amount of synthetic data can enhance the learning, and the more synthetic data, the better the enhancement effect, but when the synthetic data reaches a certain percentage and then continue to increase, it achieves the opposite effect. In NO.18-NO.23, we compare the enhancement effect of synthetic data and enhanced data generated by usual data enhancement methods. We found that the two are similar but not equal in the trend between enhancement effect and the increase in data amount.Overall, the enhanced data is more robust in terms of the sensitivity of model to the amount of enhanced data, but the enhancement effect upper limit of synthetic data is much higher than that of enhanced data. We compared NO.24-NO.25 with NO.15 and found that the performance of synthetic data is the best when it is used as pre-training data, and the performance is poor when it is used as supplementary training data. when using as enhanced data to mixed with real data, the synthetic data can also achieve certain enhancements.

Generally, if there are a large amount of real data, a small amount of synthetic data can be used as enhanced data, or a large amount of synthetic data can be used for pre-training and then training on real data. And if there are less real data, a large number of synthetic data can be used for pre-training, then fine-tuned on a small amount of real data, whose results can compete with the results on complete real data, and this conclusion is consistent with [20]. We do not recommend using synthetic data completely for training, and also do not recommend using synthetic data for supplementary training.

5 CONCLUSION

n short, we realize the synthesis of registered multimodal MRI from random normal distribution matrix through unsupervised training, which can add the lesion information freely. We verify 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 through lesion segmentation experiments. In this paper, our contributions are as follows:

- 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 structure feature map generation method to generate structural feature maps from multidimensional normal distribution matrixes;
- We realize the synthesis of registered multimodal MRI with corresponding lesion information from structural feature map and randomly selected lesion labels;
- We verify that synthetic data can be used as pre-training data or enhanced data for intelligent medical image processing tasks through a number of data availability tests.

ACKNOWLEDGEMENTS

REFERENCES

[1] 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*, 476–484, (2015).

[2] 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).

[3] 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*, 8789–8797, (2018).

[4] Pedro Costa, Adrian Galdran, Maria Ines Meyer, Michael D Abramoff, Meindert Niemeijer, Ana Maria Mendonca, and Aurelio Campilho, 'Towards adversarial retinal image synthesis.', *arXiv: Computer Vision and Pattern Recognition*, (2017).

[5] Lee R Dice, 'Measures of the amount of ecologic association between species', *Ecology*, **26**(3), 297–302, (1945).

[6] Ian J Goodfellow, Jean Pougetabadie, Mehdi Mirza, Bing Xu, David Wardefarley, Sherjil Ozair, Aaron C Courville, and Yoshua Bengio, 'Generative adversarial nets', *neural information processing systems*, 2672–2680, (2014).

[7] 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*, 5787–5796, (2017).

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

[9] Thomas Joyce, Agisilaos Chartsias, and Sotirios A Tsaftaris, 'Robust multi-modal mr image synthesis', *medical image computing and computer assisted intervention*, 347–355, (2017).

[10] 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*, 597–609, (2017).

[11] Diederik P Kingma and Max Welling, 'Auto-encoding variational bayes', *international conference on learning representations*, (2014).

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, 'Imagenet classification with deep convolutional neural networks', *neural information processing systems*, **141**(5), 1097–1105, (2012).

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

Table 2. Synthetic data availability verification experiments.

NO.	real data	synthetic data	enhanced data	mixing modes	MSE	Dice Score
1	×1	0	0	-	0.026	0.915
2	×50%	0	0	-	0.032	0.902
3	0	×1	0	-	0.205	0.708
4	0	×2	0	random mixing	0.206	0.736
5	0	×3	0	random mixing	0.205	0.754
6	×10%	×1	0	synthetic first	0.031	0.908
7	×10%	×2	0	synthetic first	0.028	0.907
8	×10%	×3	0	synthetic first	0.030	0.907
9	×20%	×80%	0	random mixing	0.041	0.850
10	×50%	×50%	0	random mixing	0.031	0.904
11	×80%	×20%	0	random mixing	0.024	0.935
12	×1	×20%	0	random mixing	0.025	0.921
13	×1	×50%	0	random mixing	**0.023**	**0.939**
14	×1	×80%	0	random mixing	0.026	0.916
15	×1	×1	0	random mixing	0.027	0.913
16	×1	×2	0	random mixing	0.033	0.901
17	×1	×3	0	random mixing	0.034	0.897
18	×1	0	×20%	random mixing	0.027	0.911
19	×1	0	×50%	random mixing	0.025	0.927
20	×1	0	×80%	random mixing	0.026	0.920
21	×1	0	×1	random mixing	0.026	0.915
22	×1	0	×2	random mixing	0.032	0.898
23	×1	0	×3	random mixing	0.036	0.885
24	×1	×1	0	real first	0.195	0.795
25	×1	×1	0	synthetic first	**0.021**	**0.940**

[14] Bjoern Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lanczi, Elisabeth Gerstner, Marc-Andre Weber, Tal Arbel, Brian Avants, Nicholas Ayache, Patricia Buendia, Louis Collins, Nicolas Cordier, Jason Corso, Antonio Criminisi, Tilak Das, Hervé Delingette, Cagatay Demiralp, Christopher Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan Iftekharuddin, Raj Jena, Nigel John, Ender Konukoglu, Danial Lashkari, Jose Antonio Mariz, Raphael Meier, Sergio Pereira, Doina Precup, S. J. Price, Tammy Riklin-Raviv, Syed Reza, Michael Ryan, Lawrence Schwartz, Hoo-Chang Shin, Jamie Shotton, Carlos Silva, Nuno Sousa, Nagesh Subbanna, Gabor Szekely, Thomas Taylor, Owen Thomas, Nicholas Tustison, Gozde Unal, Flor Vasseur, Max Wintermark, Dong Hye Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa, Mauricio Reyes, and Koen Van Leemput, 'The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)', *IEEE Transactions on Medical Imaging*, 33, (2014).

[15] Shun Miao, Sebastien Piat, Peter Walter Fischer, Ahmet Tuysuzoglu, Philip Mewes, Tommaso Mansi, and Rui Liao, 'Dilated fcn for multi-agent 2d/3d medical image registration', *national conference on artificial intelligence*, 4694–4701, (2018).

[16] Mehdi Mirza and Simon Osindero, 'Conditional generative adversarial nets', *Computer Science*, 2672–2680, (2014).

[17] 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*, 417–425, (2017).

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

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

[20] 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*, 1–11, (2018).

[21] 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*, 677–684, (2015).

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

[23] Alle Meije Wink and Jos B T M Roerdink, 'Denoising functional mr images: a comparison of wavelet denoising and gaussian smoothing', *IEEE Transactions on Medical Imaging*, **23**(3), 374–387, (2004).

[24] 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*, 9242–9251, (2018).

[25] Bo Zhao, Bo Chang, Zequn Jie, and Leonid Sigal, 'Modular generative adversarial networks', *european conference on computer vision*, 157–173, (2018).

[26] 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*, 2242–2251, (2017).