

基于条件生成对抗网络的配准多模态脑MRI合成 Registered Multimodal MRI with Lesion Label

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

在基于大量数据驱动的医学影像智能处理任务中，医学影像数据的收集和采集是非常困难的，尤其是配准的多模态医学影像数据。合成的医学影像数据可以很好地缓解数据不足的问题。我们基于无监督的条件生成对抗网络模型实现了在已有单模态影像数据时可扩展出多模态影像数据并保留原模态的病灶信息，进一步实现了完全从随机噪声生成配准的多模态医学影像并可自由添加病灶信息。我们验证了合成的医学影像可以在医学影像智能处理任务中与真实数据混合使用来提高模型的泛化能力。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 well alleviate the problem of insufficient data. In this paper, based on the unsupervised CGAN model, we extend multimodal image data from existing unimodal image data and retain lesion information of the original modal, furthermore we achieve the generation of registered multimodal medical images from random noise and the freedom to add lesion information. We verify the synthetic medical images can be mixed with real data in medical image intelligent processing tasks to improve the generalization capabilities of the model.

1 介绍Introduction

核磁共振成像（MRI）是一种常见的医学影像，根据成像参数的不同可以有多种模态，例如T1、T2、T1c等。不同的模态对医生具有不同的参考价

值，医生往往需要多个模态的影像互相对照才能做出准备的判断。在医学影像的智能处理任务的训练和学习中，我们往往也期望获得更多模态的影像，例如采用卷积神经网络（CNN）[1]或生成对抗网络（GAN）[2]进行的医学图像处理任务。

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 the 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)[1] or Generative Adversarial Networks (GAN)[2].

当同一个病人的同一个部位通过不同的成像技术得到不同的模态时，如果成像位置和视角一致的则被认为这些模态是配准的。相较于单模态数据，配准的多模态影像数据能提供更多的信息，可以支撑更多更复杂的应用场景，满足深度神经网络的训练数据的需求，有助于提供更加高效可靠的智能诊断服务。对于医生来说，获取不同模态的影像需要花费更长的时间并且需要患者的耐心配合，对于医学影像的智能处理任务的研究者来说，多模态的MRI数据集十分稀缺，收集难度非常大，尤其是罕见病，而配准的数据则更加稀少，这使得很多的训练任务无法实现。因此，通过应用图像合成

技术扩展数据集,从已有的单模态图像转换为配准的多模态图像、从随机噪声生成配准的多模态医学影像,有着广泛的用途和深远的意义。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 diseases, 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, generate registered multimodal images from random noise, has a wide range of uses and far-reaching significance.

在GAN之前,一些研究使用图字典映射[3]、稀疏编码[4],[5],CNN[6]等探索了医学影像的跨模态转换。此后许多研究使用GAN能产生更高质量的转换结果[7],[8],[9],[10]。得益于GAN的强大能力,目前,采用GAN实现跨模态医学影像转换成为主流[11],[12],[13],[6],[14]。一般的转换基于成对的数据,最近也有研究从不成对的跨域数据中学习[11]。最近的研究有将像素到像素的GAN应用于脑部MRI到CT图像的转换[12],[14]、视网膜血管注释到图像的转换[15]、基于CycleGAN[9]的心脏MRI到CT图像的相互转换与分割[12]等。对于多模态的合成,[16]实现多输入多输出的MRI合成,但对输入的多模态数据要求配准。基于此,[17]改进实现存在缺失或未配准的多输入合成模型,能够从其输入的任何子集执行MRI图像合成,但限制了输出为单一模态,且模型不可扩展。[18]针对医学图

像配准进行了深入研究。[19]应用GAN合成脑肿瘤图像实现数据增强和数据匿名化,但需要额外训练解剖结构分割网络,且要求数据集带有病灶分割标签,模型泛化能力弱。[15]研究了基于变分自编码器(VAE)[20,21]的思想实现血管注释图的随机生成,进而合成彩色视网膜图像。在当前的这些医学影像合成的研究中,大多仅探索了两个不同模态之间的转换[11],[12],[3],[5],[13],[6],[14],对多模态的研究还很稀少[16],[17],[19],而在医学影像处理领域之外,多域转换的发展最近已经有了进展[7],[8],[10],[22]。

Some studies explored cross-modal medical images translation prior to GAN by using graph dictionary mapping[3], sparse coding[4],[5], and CNN[6]. Since then, many studies used GAN to generate higher quality translation results[7],[8],[9],[10]. Owe to the powerful capabilities of GAN, it has become the mainstream to achieve multimodal medical image translation[11],[12],[13],[6],[14]. The general translation is based on paired data, some studies have also learned from unpaired cross-modal data[11]. Recent studies have realized brain MRI to CT image translation with pixel-to-pixel GAN[12],[14], retinal vascular annotation to image translation[15], CycleGAN-based[9] cardiac MRI to CT image translation and segmentation[12]. For multimodal synthesis, [16] implements MRI synthesis of multiple inputs multiple outputs, but requires registration for input multimodal data. Based on this, [17] improves and implements a missing or unregistered multi-input synthesis model that can perform MRI image synthesis from any subset of its inputs, but limits the output to a single modality and the model is not scalable. [18] has conducted in-depth research on medical image registration. [19] applies GAN to synthesize brain tumor images for data enhancement and data anonymization, but additional training of anatomical segmentation networks is required, and the dataset is required to have lesion segmentation labels, so the model generalization ability is weak. [15] studies the random generation of vascular annotation maps based on the idea

of Variational Auto-Encoder (VAE)[20, 21], and then synthesizes color retinal images. In these current studies of medical image synthesis, most are two-modal translation[11],[12],[3],[5],[13],[6],[14], and the study of multimodal translation is very rare[16],[17],[19]. Outside the field of medical image processing, the development of many-to-many translation has recently made progress[7],[8],[10],[22].

目前针对医学影像合成的研究存在模态数量难以扩展、需要配准训练数据、依赖于复杂的大型网络、无法添加或保留病灶、需要从真实数据扩展而无法从随机矩阵开始生成、需要额外的训练数据等各项问题,且大多数研究对合成数据的评价依赖于经验医师的人工视觉效果评估,没有进行客观的量化检验。因此,我们设计了一种基于条件生成对抗网络(CGAN)[23]的多模态配准图像生成的方法,采用无监督学习方法,训练数据无需配准,可以接收一个符合我们设计规范的随机输入,进而生成一组有标签的多模态配准图像。具体来说,我们的贡献体现在以下三个方面: 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 need to be extended from real data and cannot be generated 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 multi-modal registration image generation scheme based on the Conditional Generative Adversarial Networks (CGAN)[23]. With unsupervised learning method, training data do not need to be registered. Our solution can receive a random input that meets our design specifications to generate a set of labeled multimodal registration images. Specifically, our contribution is in the following three areas:

- **结构特征图的提取与随机生成** 我们针对医学影像

提出了一种解剖结构特征的提取方法,无需额外的解剖结构分割标签或标签提取训练,可直接从任意模态的真实影像提取得到结构特征,用以辅助GAN学习生成更合理的合成影像。我们训练了一个结构特征图生成器实现了从多维正态分布生成结构特征图。我们的提取方法可以直接获取真实影像的解剖结构特征,在提升合成影像质量同时不带来额外参数、计算开销小,而随机生成方法可以无限地生成丰富多样的结构特征图。

- **带标签多模态配准影像的合成** 我们使用随机生成的结构特征图,融合随机选择的病灶信息标签,通过生成器合成配准的多模态MRI。我们探讨了多种病灶生成指导方法,并通过病灶生成指导方法实现了在多模态MRI的合成过程中根据输入病灶标签有效地生成对应的病灶信息。在训练时我们无需配准的数据,除病灶信息标签外无需额外的数据标签,而合成的数据是配准的,且输入的随机选择病灶信息标签即为合成数据的病灶信息标签。我们的方法能够便捷快速地构建带标签的配准多模态MRI数据集。
- **合成数据可用性的客观验证方法** 我们使用不同比重的合成数据和真实数据构建的数据集来训练病灶分割网络,验证了合成数据可以在医学影像智能处理任务与真实数据混合使用来提高模型的泛化能力。对比传统合成影像质量的主观评价方法,我们的方法能客观真实地验证合成数据的可用性。

2 方法

我们在脑MRI数据的合成任务上进行了我们的方法的展示,其中的病灶为肿瘤,病灶标签为肿瘤分割标签,病灶处理任务为病灶分割。我们的方法对合成部位、病灶处理任务和模态数量等不做限制,通过下述的展示,我们能很容易的将该方法推广应用到其他的类似任务。

2.1 整体架构

如图1所示,我们的方案包括结构特征图提取和生成、多模态MRI生成、构建合成数据集、合成数据

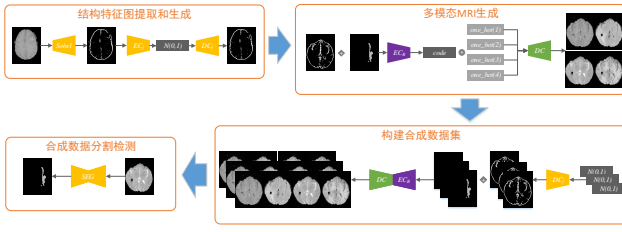


图 1: 整体架构图.

分割检测四个主要阶段。

结构特征图提取和生成阶段我们将获得一个结构特征图生成器，能从随机的正态分布矩阵生成结构特征图。该阶段我们训练的模型模包括一个结构特征图编码器、一个结构特征图解码器和一个结构特征图鉴别器。

多模态MRI生成阶段我们产出一个条件生成器，其以结构特征图为输入，能根据不同的独热条件向量生成不同模态的MRI，并且可在结构特征图上添加病灶标签使得生成的MRI具有对应的病灶信息。该阶段我们训练的模型模块包括一个结构特征与病灶标签融合图编码器、一组病灶标签分割组件、一个MRI编码器、一个MRI解码器、一个MRI鉴别器和一个MRI编码鉴别器。

在构建合成数据集阶段，我们使用前两个阶段产出的模型先从随机正态分布矩阵生成足量的结构特征图再随机添加病灶标签，最后生成配准的多模态MRI，构建出一个合成数据集。

在合成数据分割检测阶段，我们根据真实的数据为每个模态的MRI单独训练一个病灶分割网络，并在真实数据集中进行分割能力测试。此外，我们使用由不同比重的合成数据和真实数据构建的数据集来对病灶信息检测阶段中的病灶分割网络进行训练，训练充分后再在真实测试数据集上进行分割能力测试，对比各项测试结果，以验证合成数据在模型训练中的可用性。该阶段我们训练的模型包括每个模态的MRI的基于不同数据比重的病灶分割器。

2.2 结构特征图提取方法

直接从随机噪声通过生成对抗训练生成的医学影像通常训练困难且难以生成真实的结构信息。我们将医学影像中提供基本轮廓结构信息的图像称

为其结构特征图，例如视网膜血管分布图可视为视网膜图像的结构特征图[15]。结构特征图可以为医学影像的合成提供必要的指导信息，例如合成脑部MRI图像时一些研究从脑分割标签图获取基本的结构信息[19]。然而，视网膜血管分布图和脑分割标签图等常用的结构特征图都需要额外的数据和训练才能实现从原图提取出结构特征图。为此，我们首先设计了一种直接从脑MRI提取结构特征图的方法，该方法具有运算快、无需训练、无需额外数据等优点。

在传统的数字图像处理方法中，Roberts算子[24]、Prewitt算子[25]、Sobel算子[26]等是十分优秀的边缘检测算子，其中Sobel算子常用于脑部医学图像的处理，其卷积核参数和计算公式如图所示。我们探索出了从Sobel算子生成的边缘检测图中进一步提取结构特征的方法，如算法 1所示。

Algorithm 1 Structural Feature Extraction

- 1: 输入一张真实图像 x , β 为元素阈值
 - 2: $f1 = \text{reduce_min}(\text{sobel}(x))$
 - 3: $f2 = \text{reduce_max}(\text{sobel}(x))$
 - 4: $f1 = \text{mean}(f1) - f1$
 - 5: $f2 = f2 - \text{mean}(f2)$
 - 6: $f1 = \text{ones} * (f1 > \beta)$
 - 7: $f2 = \text{ones} * (f2 > \beta)$
 - 8: $f = f1 + f2$
 - 9: $f = \text{ones} * (f > 0.)$
-

简单来说，我们对一张真实图像用Sobel算子提取得到其横向和纵向的边缘检测图，对两张边缘检测图进行最大值规约和最小值规约得到两张新的边缘检测融合图，然后两张边缘检测融合图分别与各自的平均像素值求差，再对两张差值图根据设定像素阈值进行二值化处理，两张二值图求和后再进行完全的二值化，最后得到的就是我们需要的结构特征图。

2.3 随机结构特征图的生成训练

在生成结构特征图时，[19]仍然需要真实的模态图作为输入来得到生成的结构特征图，这大大降低

- **结构特征对抗性损失** 这部分损失包括：使得结构特征图编码服从正态分布的对抗性损失，以及使得随机正态分布矩阵解码出结构特征图更逼真的对抗性损失

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

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

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

- **结构特征编码的监督损失** 这项损失使得结构特征图中间编码结果服从正态分布，其中， $mean()$ 函数为均值函数。

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

- **结构特征图及掩模的自监督损失** 结构特征图和掩模两次重建融合后与原始结构特征图和掩模的两两自监督一致性损失

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

- **Mask生成器损失** 从 f 生成对应 $mask$ 的Mask生成器损失

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

2.4 辅助的模态重建和模态转换训练

在辅助的模态重建和模态转换训练中，我们用到了套模态编码器、模态解码器和鉴别器，以及一组病灶标签生成组件，我们在真实的MRI上对这三个部件进行了下面同步的模态重建和模态转换训练，以约束三个部件在多模态MRI生成的训练中完成我们指定的学习任务。此外我们还通过一个特征鉴别器来对两个训练过程进行一致性指导。

如图所示，模态转换和重建时，编码器将模态 i 的真实MRI x_i 编码得到语义特征图 $code_{x,i}$ ，然后我们将其与不同的条件向量连接，通过解码器解码出全部的模态。循环重建时，我们对所得到的转换

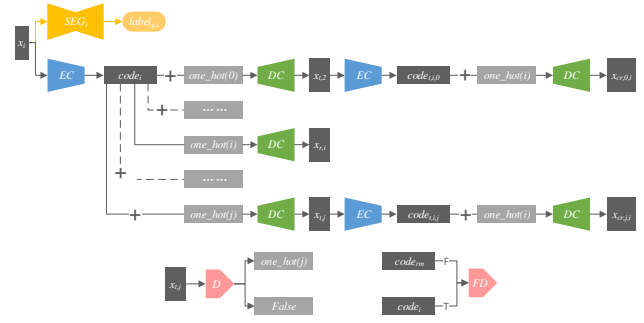


图 3: 辅助的模态重建和模态转换训练。

图采用编码器全部进行再编码，将全部再编码得到的语义特征图均与模态 i 的条件向量进行连接，最后再用解码器全部解码得到循环重建的 $x_{r,i}$ 。模态重建的循环重建都是自监督训练。在上述过程中，我们以原始输入模态 x_i 对应的病灶标签 l_i 作为病灶还原训练的监督标签，对 x_i 用病灶标签生成组件得到 $l_{r,i}$ 。

我们的鉴别器模块独立更新，其他模块通过一个优化器更新训练，损失项包括鉴别器提供的对抗性损失、模态重建自监督损失、模态循环重建自监督损失、模态循环重建一致性损失、语义一致性损失、病灶监督损失。详细损失如下，其中 $x_{r,i}$ 表示模态 i 重建得到的MRI， $x_{j,t,i}$ 指由模态 j 转换生成的模态 i 的MRI， $d_{x,j,t,i}$ 和 $c_{x,j,t,i}$ 分别为鉴别器对 $x_{j,t,i}$ 的真假鉴别和类别鉴别结果， $x_{cr,j,i}$ 表示模态 i 转换为模态 j 再转换回模态 i 的MRI； $code_i$ 表示编码 x_i 编码器编码后得到的语义特征图， $code_{i,t,j}$ 表示 x_i 转换生成的模态 j 的MRI再经过编码器编码后得到的语义特征图； l_i 表示 x_i 的真实病灶标签， $l_{g,i}$ 表示 x_i 经过病灶标签生成组件生成的病灶标签：

- 鉴别器损失

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

- 对抗性损失和类别指导损失

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

- 模态重建自监督损失

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

- 模态循环重建自监督损失

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

- 模态循环重建一致性损失

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

- 语义一致性损失

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

- 病灶监督损失

$$loss_{sv,l} = \sum_{i=0} \|l_i - l_{g,i}\|_2^2$$

2.5 结构特征图与分割标签图的融合

我们已经可以得到从标准正态分布随机生成的结构特征图 f_g ，然后我们随机选择分割标签图 l ，将包含 n 个类别的分割标签图转为 n 个通道的独热矩阵 $onehot_l$ ，每个通道对应一个分割类别，每个通道内的像素值为0或1，与对应类别分割位置相同的区域像素值为1其余部分为0，这样各个1像素区域与分割标签图中的各个分割区域是配准的。接下来，我们将 $onehot_l$ 的每个通道与 f_g 按位求取加权，得到一个新的融合了 f_g 和 l 信息的矩阵。

如果结构特征图 f' 是从随机MRI m 中提取的，那么提取出的结构特征有可能包含肿瘤结构信息，会干扰与随机标签 l 中的肿瘤信息，影响融合后生成的图像，所以 f' 需要在与随机标签 l 融合前消除肿瘤信息，得到无肿瘤信息的结构特征图 f ，使生成图像的肿瘤信息只来源于标签 l 。我们通过 m 的分割标签 l_m 生成无边界扩充的分割掩膜 $mask_{l,m}$ ，则 $f = mask_{l,m} \times f'$ 。

由于随机选择的病灶的位置上可能出现在结构特征图的脑部轮廓之外，因此我们选取标签

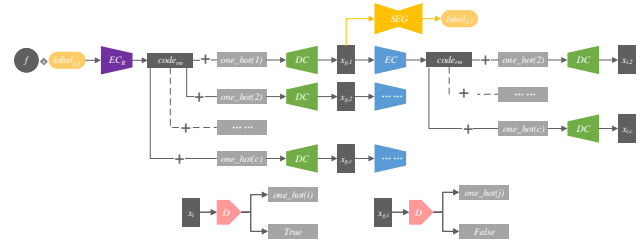


图 4: 多模态MRI生成.

图时需要获取结构特征图的脑部区域掩膜 $mask$ ，若 $mask$ 与选取的 l 求积后为0，则说明肿瘤标签像素在 $mask$ 的脑轮廓内部，可以采用，否则需要重新选取 l 。

2.6 多模态MRI生成

结构特征图与分割标签图的信息融合图包含了目标部位的基本解剖结构信息和病灶信息，从该图生成多模态图比直接从随机噪声生成多模态图更易训练，生成的模态图更加合理和逼真。多模态MRI生成过程如图所示，首先，我们使用一个单独的编码器对信息融合图进行编码得到语义特征图，语义特征图与不同的条件向量堆叠后通过一个模态解码器解码，得到不同模态的合成图。我们通过一个模态鉴别器提供的对抗性损失和类别指导损失来使得生成的各个模态的合成图逼近于真实的MRI。各个模态的合成图再通过一个模态编码器得到模态语义特征图，然后这些语义特征图与不同的条件向量堆叠后通过模态解码器解码就实现了模态的转换，我们通过损失约束所有语义特征图的一致性，以及转换图的一致性，以此保证了生成的多模态影像互相配准。此外，我们使用一组病灶标签生成组件从各合成MRI中分割还原出肿瘤分割标签，确保生成的多模态影像根据输入病灶标签生成了对应的病灶内容。

我们的鉴别器模块独立更新，除只用于病灶标签生成组件中的模块外其他模块通过一个优化器更新训练，只用于病灶标签生成组件中的模块仅在辅助的转换和重建训练中采用真实数据训练更新，合成训练过程仅用于为生成模块提供语义指导损失。多模态MRI生成过程的具体损失函数如

下，其中 $d_{x,i}$ 和 $c_{x,i}$ 为鉴别器 $D(x_i)$ 的真假鉴别输出和类别鉴别输出， $d_{x,g,i}, c_{x,g,i}$ 为 $D(x_{i,g})$ 的输出； $x_{g,i}$ 为模态 i 的合成图， $x_{g,j,t,i}$ 为模态 j 的合成图转换生成的模态 i 的转换图； $code_g$ 为输入信息编码后的语义特征图， $code_{x,g,i}$ 为 x_i 编码后的语义特征图； l 为输入的标签图， $l_{x,g,i}$ 为 $code_{x,g,i}$ 解码得到的重建标签图； f 为输入的结构特征图， $f_{x,g,i}$ 为从 $x_{g,i}$ 提取得到的结构特征图：

- 使得随机正态分布矩阵解码出结构特征图更逼真的对抗性损失

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

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

- 模态类别指导损失

$$loss_{D,class} = \sum_{i=0} (\|c_{x,i} - i\|_2^2 + \|c_{x,g,i} - i\|_2^2)$$

$$loss_{G,class} = \sum_{i=0} (\|c_{x,g,i} - i\|_2^2)$$

- 输入的结构特征图的重建自监督损失

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

- 与输入的结构特征图融合后输入的肿瘤分割标签图的重建自监督损失

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

- 生成的模态图进行转换得到的转换图与生成图的自监督损失

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

- 自监督语义一致性损失

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



图 5: 构建合成数据集.

2.7 病灶标签生成组件设计方案

我们设计了如下五组病灶标签生成组件。

- **单分割器方案** 每个模态由一个共同的完整的分割器从合成的MRI还原得到各自的病灶标签。
- **单病灶编码器+多病灶解码器方案** 每个模态由一个采用共同的病灶编码器与每个模态的独立的病灶解码器组合得到的分割器，从合成的MRI还原得到各自的病灶标签。
- **多分割器方案** 每个模态由一个独立的完整的分割器从合成的MRI还原得到各自的病灶标签。

上述三种方案的损失函数与前文所述损失是一致的，各个模块均只使用辅助的模态重建和模态转换训练中的病灶监督损失进行训练，多模态MRI生成中的与输入的结构特征图融合后输入的肿瘤分割标签图的重建自监督损失只用于MRI生成组件的训练。

2.8 构建合成数据集

如图所示，我们通过训练好的结构特征图解码器即可从随机生成的正太分布矩阵生成任意数量的结构特征图。然后，我们对原始的标签集进行了随机的缩放、旋转、平移、翻转等改变得到随机病灶标签集。我们再将随机生成的结构特征图和从随机病灶标签集中随机选择的病灶信息标签融合，同样地，我们借助生成的与结构特征图匹配的掩膜，可以筛选得到合适的随机病灶标签。最后，我们通过多模态生成模型即可从融合信息图中生成配准的多模态影像，选取的病灶信息标签就是生成的多模态影像的病灶标签。由此，我们可以从随机矩阵构建带有病灶标签的多模态配准影像数据集。

由于训练时去除肿瘤部分的结构的影响，合成的结构特征图中，存在一些质量较差的脑轮廓未闭合结构特征图，我们对此设计了一个结构特征图过滤算法。首先，我们使用生成器生成一张结构特征图和其对应的掩膜，我们先对结构特征图进行高

斯模糊[27]，再采用OpenCV[]提供的轮廓查找算法和填充算法获取高斯模糊图所有的闭合轮廓并进行填充，这样我们得到一个采用通常的算法的得到的掩膜，最后我们计算两张掩膜的均差(MAE)。若MAE低于我们设定的阈值则说明该结构特征图主要的脑部轮廓较为完整，该特征图可以使用；否则说明该结构特征图主要的脑部轮廓有残缺，采用通常的算法得到的掩膜内部是空心的，与生成器生成的掩膜差异较大，因此，需要重新生成。算法表示如下：

Algorithm 3 Structural feature map filtering

```

1: function GetMaskFromF(img)
2:   contours = OpenCV.findContours(img)
3:   img = OpenCV.drawContours(img, contours)
4:   return img
5: end function
6:
7: mae = 0.05
8: do
9:   f, m = Generator()
10:  m' = GetMaskFromF(f)
11: while MAE(m', m) <= mae

```

从筛选出来的结构特征图和匹配的病灶标签得到的多模态MRI中，同样存在病灶信息生成情况较差的样本。此时，我们通过预先训练好的病灶分割网络对我们的合成MRI数据进行分割，然后将分割结果与输入的病灶标签进行骰子评分评估，只保留评分高于设定阈值（默认0.95）的样本。

经过多重筛选，我们得到最终的由随机结构特征图、配对掩膜、随机病灶分割标签、多模态MRI组成的合成数据集。我们要求该数据集使用在真实数据上训练得到的分割网络进行病灶分割时，得到的分割标签与数据集中的标签进行骰子分数评估，能取得0.99以上的分数。

2.9 病灶分割

我们的病灶分割训练过程如图所示，为了验证合成数据中是否生成了与随机输入的病灶标签一致



图 6: 病灶分割.

的病灶信息，我们用真实的数据为每个模态单独训练一个病灶分割网络，并在真实数据集中进行分割能力测试。然后使用训练好的分割器对合成数据集中的影像进行分割，将分割结果与输入的病灶标签比对评估，以此检验合成的影像中包含了预期的病灶信息。

此外，我们还采用由不同比重的合成数据和真实数据构建的数据集，来训练病灶分割网络，训练充分后再在真实测试数据集上进行分割能力测试，对比各项测试结果，以验证合成数据对分割训练的提升效果。

分割训练的损失函数如下，其中 $l_{r,i}$ 表示 x_i 经过病灶分割器生成的病灶标签：

- 病灶监督损失

$$loss_l = \sum_{i=0} \|l_i - l_{r,i}\|_2^2$$

3 实验

3.1 训练设置

每项实验训练50个epoch；基础学习率分别为1e-4，无权重衰减；采用优亚当优化器，beta1取0.5；在输入层进行0.1的Dropout；Batch size为1；在生成器中使用均值滤波器的参数进行参数初始化，具体来说，对于一个卷积核尺寸为 $[k, k, f]$ 的卷积层，我们采用random normal initializer设定该层卷积核的 $k \times k \times f$ 个参数均设置为 $1/(k \times k \times f)$ ，偏置量为0；在鉴别器中我们使用0均值和标准差为0.2的random normal initializer，偏置量为0。我们采用骰子分数[28]和均方差（MSE）[29]进行分割结果的评估，评估结果为2D图像的评估结果的均值，每项实验训练四次保留最佳结果。

3.2 BRATS2015数据集

我们采用了公开的BRATS2015[30]数据集进行实验，该数据集包含已配准的T1、T2、T1c、Flair四个模态，训练集每个模态有274张3D MRI，大小为 $155 \times 240 \times 240$ ，同时配有274张相同尺寸的肿瘤分割标签。我们将样本按9:1划分训练集和测试集，取每张3D MRI第55-105间的50个slice构建2D的数据集。在数据预处理阶段，我们将每张图进行了标准化。

3.3 BRATS合成数据集

我们采用 2.8中的方法构建了一个配准的包含T1、T2、T1c、Flair四个模态的具有肿瘤标签的合成数据集。合成数据集样本的尺寸与BRATS2015数据集一致，但样本的多少可以根据实验需要进行任意数量的合成。

3.4 BRATS增强数据集

我们对原始的BRATS2015数据集进行了随机的缩放、旋转、平移、翻转等改变，得到增强数据。增强数据集样本的尺寸与BRATS2015数据集一致，但样本的多少可以根据实验需要进行任意数量的生产。

3.5 病灶生成组件各方案对比实验

我们使用在处理后的BRATS2015数据集的训练集对病灶分割网络进行了100个epoch的充分训练，然后我们在BRATS2015数据集的测试集和采用不同病灶生成组件方案的未经分割器筛选的的BRATS合成数据集上分别进行分割测试，除了测试数据来源的不同外，测试数据的样本量等其他条件完全相同。

3.6 合成数据可用性验证实验

如表所示，我们将真实的BRATS2015训练数据与BRATS合成数据进行了不同数量的混合，再用构建的混合数据集进行分割训练，最后再在真实的BRATS2015测试数据上进行模型的分割能力评估，所有实验都进行训练100个epoch的充分训练，除了训练数据源的不同外其他条件完全相同。同时，我们还进行了单独的合成数据的训练、真实数据与通

表 1: 病灶检测实验结果.

synthetic method	test data type	MSE	Dice Score
-	real	0.014	0.961
1SEG	synthetic	0.053	0.741
1ECL+4DCL	synthetic	0.055	0.808
4SEG	synthetic	0.043	0.838

常的数据增强数据的混合训练作为对比。我们设定了随机混合、先真后假、先假后真三种数据混合方式。

4 结果

4.1 实验量化结果

病灶检测实验结果 如表 1所示，我们在BRATS2015训练数据经过100个epoch的充分训练后，在真实的测试数据集上，分割测试结果达到了0.014的MSE和0.961的SSIM[31]。之后我们使用这个表现优秀的分割网络对我们未经过滤的合成数据进行分割测试，在与真实测试数据集相同数据量的合成数据集上，不同病灶标签生成组件设计方案都取得了较好的分割结果，其中一个模态一个分割器的方案取得了最好的结果，Dice Score也达到了0.838。

合成数据可用性验证实验结果 如表 2所示，我们在不同的成分和数量的数据集上都经过训练后在真实测试集上评估，得到了表中的结果。实验1在BRATS2015训练数据经过100个epoch的充分训练后，在真实的测试数据集上，分割测试结果达到了0.033的MSE和0.954的SSIM。从表中实验2-4单独使用合成数据训练的结果看，与真实数据的训练结果有一定差距，说明我们的合成数据尚不能完全替代真实数据充当训练集。实验7-9结果表明使用大量合成数据进行预训练，再在少量真实数据上微调能达到和实验1全使用真实数据训练十分接近的结果。实验12-14一定比例的真实数据与合成数据随机混合，能达到和真实数据训练差不多的效果，比例相近时效果与实验1相差不大，合成数据比例偏高时，结果比实验1略低，合成数据比例偏低时，能提供模型的泛化能力，取得高于实验1的结果。实验15-

表 2: 合成数据可用性验证实验结果.

序号	真实量	合成量	增强量	混合方式	MSE	Dice Score
1	15070	0	0	无混合	0.033	0.954
31	15070×0.5	0	0	无混合	0.026	0.916
2	0	15070	0	无混合		
3	0	15070×2	0	随机混合		
4	0	15070×3	0	随机混合		
7	15070×0.1	15070	0	先假后真		
8	15070×0.1	15070×2	0	先假后真		
9	15070×0.1	15070×3	0	先假后真		
12	15070×0.5	15070×0.5	0	随机混合		
13	15070×0.2	15070×0.8	0	随机混合		
14	15070×0.8	15070×0.2	0	随机混合		
32	15070	15070×0.2	0	随机混合	0.026	0.921
33	15070	15070×0.5	0	随机混合	0.023	0.939
34	15070	15070×0.8	0	随机混合		
15	15070	15070	0	随机混合		
16	15070	15070	0	先真后假		
17	15070	15070	0	先假后真		
18	15070	15070×2	0	随机混合		
19	15070	15070×3	0	随机混合		
35	15070	0	15070×0.2	随机混合		
36	15070	0	15070×0.5	随机混合		
37	15070	0	15070×0.8	随机混合		
22	15070	0	15070	随机混合		
23	15070	0	15070×2	随机混合		
24	15070	0	15070×3	随机混合		

17等比例混合真实数据与合成数据, 随机混合时合成数据用作数据增强, 效果比真实数据训练效果好, 说明合成数据起到数据增强的作用, 能有效提升训练效果; 先训练真实数据再用合成数据进行补充训练, 结果较差, 与完全使用合成数据训练效果差不多; 先使用合成数据进行与训练, 再使用真实数据进行训练, 能取得非常好的结果。实验15, 32-34, 18-19按多种比例混合真实数据与合成数据, 开始时(实验33-34, 15)随着合成数据的增加能提升训练效果, 但随着合成数据比例继续增大, 训练效果会下降。最后实验35-37, 22-24是使用传统数据增强方法进行数据增强, 实验结果表明合成数据比传统数据增强的训练效果更好。但在真实数据充当训练集的基础上, 可以使用合成数据集作为增强数据来提高模型泛化能力, 数据最好的组合使用方式是随机混合使用。我们也注意到, 合成数据也并不是越多越好的, 当合成数据所占比重过大时, 为降低真实数据的影响力。在我们的实验中, 采用【】倍的合成数据与真实数据进行随机混合取得了最好的提升效果, 相比完全的真实数据, 分割测试结果提升了【%】的MSE和【%】的SSIM, 远远大于通常的数据



图 7: 合成的结构特征图机器掩膜.



图 8: 合成的多模态MRI.

增强的提升幅度。与[19]的结论一致, 我们的合成数据也可用于预训练, 然后再使用少量真实数据微调, 最后能得到和完全真实数据训练相竞争的结果。

4.2 合成数据可视化

上述各项实验的一些生成结果如图 7 8 9 10所示, 从图中我们能更直观的取得结论。

5 结论与未来工作

我们基于条件生成对抗网络, 通过无监督的训练, 实现了能在已有单模态数据时可扩展出多模态影像数据并保留原模态的病灶信息, 同时可从随机噪声生成配准的多模态医学影像并可自由添加病灶信息。我们通过病灶分割实验验证了合成的医学影像可以用于医学影像智能处理任务的预训练或者补充训练来提高模型的泛化能力。具体来说, 我们的贡献包括以下几点:

- 结构特征图提取方法。我们从医学影像直接提取解剖结构信息, 无须训练, 无需额外数据;
- 随机结构特征图生成方法。我们实现了从多维正态分布采样生成结构特征图, 使结构特征图能便捷地大量生成并使最后合成的影像具有良好的多样性;
- 模态转换方法。我们实现了多模态脑MRI之间的互转, 使单模态数据可转换生成多模态影像数据并保留原模态的病灶信息;



图 9: 使用真实训练数据的分割模型在不同测试集上的分割示例.



图 10: 使用不同训练数据的分割模型在真实测试集上的分割示例。

- 基于结构特征图生成多模态图方法。我们从真实影像提取的结构特征图及随机选择的标签生成带病灶标签的配准的多模态影像；
- 构建带标签的多模态配准数据集方法。我们组合使用随机结构特征图生成方法和基于结构特征图生成多模态图方法，实现了从多维正态分布随机生成结构特征图，结合随机选择的标签生成带标签的配准的多模态MRI数据；
- 合成数据可用性测试方法。我们将合成数据与真实数据数据集进行不同比重的混合，通过病灶分割网络验证了合成数据在分割训练中的提升效果。

未来，我们将进一步在CT、PET等不同模态和其他部位中对我们的方法进行改进，并将进一步简化训练过程和提升合成质量。

6 Acknowledgments

感谢NSCCGZ提供的计算支持。

参考文献

- [1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. 2012.
- [2] 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*, pages 2672–2680, 2014.
- [3] 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.
- [4] 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.
- [5] Raviteja Vemulapalli, Hien Van Nguyen, and Shao-hua Kevin Zhou. Unsupervised cross-modal synthesis of subject-specific scans. *international conference on computer vision*, pages 630–638, 2015.
- [6] 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.
- [7] Bo Zhao, Bo Chang, Zequn Jie, and Leonid Sigal. Modular generative adversarial networks. *european conference on computer vision*, pages 157–173, 2018.
- [8] 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.
- [9] 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.
- [10] 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.

- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] Agisilaos Chatsias, 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.
- [17] Thomas Joyce, Agisilaos Chatsias, and Sotirios A Tsaftaris. Robust multi-modal mr image synthesis. pages 347–355, 2017.
- [18] 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*, pages 4694–4701, 2018.
- [19] 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.
- [20] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *international conference on learning representations*, 2014.
- [21] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. *international conference on learning representations*, pages 1278–1286, 2014.
- [22] 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.
- [23] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *Computer Science*, pages 2672–2680, 2014.
- [24] Roberts Lawrence G. Machine perception of three-dimensional solids. *Optical and Electro-Optical Information Processing*, 1965.
- [25] J.M.S. Prewitt. Object enhancement and extraction. *Picture processing and Psychopictorics*, 1970.
- [26] Irwin Sobel. Camera models and machine perception. *Ph.D.dissertation*, 1970.

- [27] 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.
- [28] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [29] 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.
- [30] Bjoern Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lenczi, 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*, page 33, 2014.
- [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.