

中山大学硕士学位论文

基于GAN的多模态医学影像的合成

Synthesis of Multimodal Medical Images Based on GAN

学位申请人: 瞿毅力

指导老师: 卢宇彤教授

专业名称: 软件工程

答辩委员会(签名):

主席:

委员:

二零二零年五月十六日

论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

学位论文使用授权声明

本人完全了解中山大学有关保留、使用学位论文的规定，即：学校有权保留学位论文并向国家主管部门或其指定机构送交论文的电子版和纸质版；有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆、院系资料室被查阅；有权将学位论文的内容编入有关数据库进行检索；可以采用复印、缩印或其他方法保存学位论文；可以为建立了馆际合作关系的兄弟高校用户提供文献传递服务和交换服务。

保密论文保密期满后，适用本声明。

学位论文作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

基于GAN的多模态医学影像的合成

专业：软件工程

硕士生：瞿毅力

指导老师：卢宇彤教授

摘要

医学影像数据的采集和标注非常困难，尤其是配准的多模态数据。合成的医学影像数据可以很好地缓解此问题，但医学影像包含复杂的生理结构信息，直接合成很容易生成不合理的结构、轮廓。此外，当前医学影像合成的研究还存在模态数量少、训练数据要求配准、不能有效合成病灶、无法从随机噪声无限合成、合成影像可用性未评估等各项未能很好解决的问题。

针对这些问题，本研究设计了一种基于GAN的配准多模态医学影像合成方法，无需配准训练数据，先从随机矩阵合成具有生理结构信息的结构特征图，进而生成一组有病灶标签的多模态配准医学影像。本研究在多个数据集上验证了合成数据中病灶信息的有效程度和合成数据在病灶处理任务中的可用程度。本研究主要工作如下：

1. 本研究提出了一种基于Sobel边缘检测算子的结构特征图的提取与随机生成方法，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。
2. 本研究实现了带标签多模态配准影像的合成。本研究将随机生成的结构特征图随机选取的病灶标签融合，通过生成网络合成多模态医学影像。我们通过实现生成的不同模态影像间的互相转换确保了它们的互相配准，通过对生成的影像的病灶信息的再处理还原出病灶标签确保了生成影像根据输入的标签生成了对应的病灶信息。
3. 本研究对合成数据的可用性进行了客观的验证。本研究使用不同数据量的合成数据和真实数据构建的数据集来训练病灶处理网络，验证了合成数据可以在医学影像智能处理任务作为预训练数据和增强数据来提高模型的能力。

关键词：医学影像、生成对抗网络、图像合成、多模态配准、边缘检测

Synthesis of Multimodal Medical Images Based on GAN

Major: Software Engineering

Name: Yili Qu

Supervisor: Prof. Yutong Lu

Abstract

The acquisition and labeling of medical image data is very difficult, especially for registered multi-modal data. Synthesized medical image data can alleviate this problem well, but medical images have complex physiological structure information, and direct synthesis can easily generate unreasonable physiological structures. In addition, the current research on medical image synthesis also includes a small number of modalities, the need to register multi-modal training data, the inability to synthesize designated lesions and test them, the inability to synthesize from random matrices, the need for additional data to generate physiological structure information, and the evaluation of synthesis quality. Objectives and other issues that are not well resolved.

Therefore, we designed an unsupervised GAN-based registration multi-modal medical image synthesis method. Without the need to register training data, we first synthesized a structural feature map with physiological structure information from a random matrix, and then generated a set of lesion labels. Multimodal registration medical image. We verified the validity of the lesion information in our synthetic data and the availability of the synthetic data in the lesion processing task on multiple data sets. The main work of this article is as follows:

1. We propose a method for extracting and randomly generating structural feature maps based on the Sobel edge detection operator. No additional anatomical structure segmentation labels or label extraction training are required, and structural features can be directly extracted from real images of arbitrary modes To assist GAN learning to generate more reasonable synthetic images.

2. We have realized the synthesis of labeled multi-modal registration images. We

Abstract

randomly fuse the randomly generated lesion feature maps with selected lesion labels and synthesize multi-modal medical images by generating a network. We ensure the mutual registration by implementing the mutual conversion between the different modal images that are generated. By reprocessing the lesion information of the generated images, we restore the lesion labels to ensure that the generated images generate corresponding lesion information based on the input labels. .

3. We have objectively verified the usability of the synthesized data. We used data sets constructed from synthetic data of different data amounts and real data to train the lesion processing network, and verified that the synthetic data can be used as pre-trained data and enhanced data in medical image intelligent processing tasks to improve the model's ability.

Keywords: Medical images, generative adversarial networks, image synthesis, multi-modal registration, edge detection

目录

摘要	I
Abstract	II
第一章 绪论	1
1.1 研究背景和目的	1
1.2 研究思路与技术路线	3
1.3 论文的结构	3
第二章 国内外研究现状	5
第三章 基础方法和数据集	7
3.1 数据集	7
3.2 基于CNN的图像分类	8
3.3 生成对抗网络	8
3.4 变分自编码器	10
3.5 语义分割	11
3.6 目标检测	11
第四章 带病灶标签的配准多模态医学影像的合成	14
4.1 整体架构	14
4.2 结构特征图的提取和生成	14
4.3 多模态影像的合成	21
4.4 多模态影像的配准	23
4.5 病灶信息的添加和合成	24
4.6 合成数据集的构建	25
第五章 合成医学影像的性能评估	28
5.1 评估指标	28
5.2 训练设置	28

目录

5.3 合成方法在脑肿瘤MRI数据集上的对比实验	28
5.4 合成方法在不同数据集上的量化评估	30
5.5 合成病灶的有效性验证实验	33
5.6 合成数据在智能医学影像处理任务中的可用性验证实验	33
5.7 合成数据效果展示	37
第六章 结语	45
参考文献	46
致谢.	54

第一章 绪论

1.1 研究背景和目的

近年来，随着深度学习的提出和发展，深度学习表现出的极强的学习能力为研究者们打开了一扇新世界的大门。如今，深度学习在各个领域得到了广泛的应用，其中，利用深度学习处理医学影像的相关研究更是越来越多，智能医学影像处理也成为了深度学习落地应用最多最有影响力的领域之一。在原理上，深度学习本身依赖于数据驱动，自然，诸多的智能医学影像处理任务需要大量的医学影像数据来进行训练学习。然而，相对于人人都可用手机拍照上传的自然图像来说，医学影像数据的采集和标注要困难得多，尤其是对于配准的多模态医学影像数据。具体来说，医学影像数据集的构建将会面对医学伦理与法律法规、病人隐私保护与数据脱敏、医院多部门的协调、病人及家属的配合、合格的采集设备、有经验的专业医生的标注、充足的病例特别是罕见病病例、常年累月的数据积累等等诸多的问题。这种数据集构建的困难和涉及其他相关利益又迫使一些医疗机构拒绝公开数据。这造成了当前相对于自然图像领域公开数据集的百花齐放、日新月异，医学影像公开数据集则显得寥寥无几、蜗行牛步。

随着生成对抗网络的强大生成能力被逐步挖掘出来，面对这种医学影像数据集稀少、数据量小的现状，通过合成的医学影像数据来进行数据增强以解决样本不足的问题成为了一种可行方案。然而，医学影像的特殊性还在于其包含了复杂的生理结构信息，采用合成自然图像的方式直接从随机噪声合成医学影像极易生成不复合生理逻辑的结构或轮廓。这使得合成方案的设计和合成训练十分具有挑战性。

另一方面，多模态的医学影像包含更多的有医学价值的信息，无论是对医生还是对智能医学影像处理任务，多模态影像都能更好的帮助诊断。但同样的，多模态影像相对于单模态影像采集难度更大，还需要考虑模态之间的配准，因此公开可用的多模态数据集少之又少。然而，在当前医学影像合成的研究中，许多方案只考虑了单模态影像的情况，对于多模态影像的合成，不是简单将多个单模态模型训练多次就能合成可用的多模态影像的，还需要确保合成的多模态影像之间保持相互配准。这是多模态医学影像合成的另一个挑战。

此外，对于医学影像，其最有价值之处在于其中的病灶信息。医学影像中的病灶信息是医生进行诊断的重要依据，也是智能医学影像处理模型推理诊断的重要依据。然而，当前绝大多数的医学影像合成的研究中，未针对病灶信息的合成进行任何针对性研究，致使合成的医学影像尽管整体上与真实影像相似但关键的病灶信息却不能确保有效的合成，即使影像中合成了病灶也无法提供对应的病灶标签，不能对合成的医学影像在医学诊断上的可用性进行评估和检验，更不能单独的对合成病灶的有效性进行评估和检验。这同样是合成医学影像的一个大的挑战。

具体来说，在上述的问题和挑战中，核心的技术背景有如下几项：

- **医学影像及其模态** 医学影像是指为了医疗或医学研究，对人体或人体某部分，以非侵入方式取得的内部组织影像，其中不同的成像方式得到的医学影像我们称之为不同的模态，常见的医学影像模态有核磁共振成像（MRI）、CT成像、PET成像、B超成像、X射线成像等。有的模态在成像时设置不同的参数将得到具有明显视觉差异的不同的子模态，例如CT分低剂量和高剂量、MRI包括T1、T2、T1c等子模态。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确的判断。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。
- **多模态图像的配准** 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是一致的，那么得到的不同模态的影像就是对齐的，我们称之为这些模态之间是配准的。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，可以支撑更多和更复杂的应用场景，满足深度神经网络对训练数据的需求，有助于提供更加高效可靠的智能诊断服务。对于医生来说，获取不同模态的配准影像需要花费更长的时间并且需要患者的耐心配合，或者需要额外的伴随失真的配准计算。
- **病灶的重要意义** 病灶是指人体器官的病变区域，例如肺结节、肿瘤、结石等。医学影像中的病灶信息对医生来说至关重要，它们是医生进行诊断的重要依据。在医学影像中，病灶区域与正常区域不一定具有明显的视觉差异，研究者们尝试使用人工神经网络来学习潜在的差异，以帮助医生进行诊断。对于医学影像

智能处理任务的研究者来说，多模态的医学影像数据集十分稀缺，收集难度非常大，尤其是罕见病数据，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。

- **图像合成的原理和发展** 图像合成为从随机噪声合成图像和从带有指导信息的草图合成目标图像，后者进一步发展为像素到像素的图像转换。图像合成本质上是一个数据分布到另一个数据分布的转换。随着生成对抗网络（GAN）GAN [2]的出现，从随机噪声合成图像发展迅速。图像转换包括风格迁移、人像转换等诸多应用，我们最关注的其中的医学影像的模态转换，即以一种医学影像模态为输入合成另一种模态的医学影像。
- **深度学习在医学影像上的应用** 除了GAN在医学影像的模态转换上的应用，深度学习还广泛应用于多种医学影像处理任务，如图像分割、病灶检测、图像分类等。[3], [4], [5]等研究中详细阐述了该方向的广阔前景。

总的来说，当前的合成医学影像的研究中还存在模态数量少、训练数据要求配准、不能有效合成病灶、无法从随机噪声无限合成、需要额外的数据产生生理结构信息、合成质量评价不客观等各项未能很好解决的问题。本研究的目的正是针对上述这些问题，基于当前的技术和研究成果，进一步探索一套完整的可以同时解决上述多个问题的解决方案。

1.2 研究思路与技术路线

1.3 论文的结构

本文一共有五章，每一章的内容安排如下：

第一章：绪论。介绍本文的研究内容及其背景，对国内外的研究现状进行简单的介绍与总结。

第二章：基础方法和数据集。介绍本文后续章节中的方法和实验将会运用的基础方法和数据集的简介，包括基础的生成对抗网络的介绍、变分自动编码器的介绍、图

像分割任务的介绍、物体检测任务的介绍。

第三章：带病灶标签的配准多模态医学影像的合成。本章是完整的方法描述，先介绍整体架构，再依次介绍结构特征图的提取和生成方法、多模态影像的合成和配准方法、辅助的模态转换训练过程、病灶信息的添加和合成方法、合成数据集的构建过程。

第四章：合成医学影像的评估与展示。本章节介绍对合成的医学影像的评估方法和实验结果，包括通用评估指标、消融实验、通用量化评估结果、合成的病灶的有效性评估、合成数据集在智能医学影像处理任务上的可用性评估。

第五章：结语。总结本文的研究成果，罗列本文的不足之处，分析本文方法落地运用的困难，并对进一步的研究方向进行展望。

第二章 国内外研究现状

- **图像合成的研究现状** 图像到图像转换任务在最近的研究中被公式化为使用编码器和解码器组成的CNN实现对像素到像素的映射 [6], [7], [8], [9], [10], [11]。GAN可采用无监督或自监督训练，也可进行有监督训练，还能使用合成数据进行辅助训练 [12], [13], [14], [15]，因此在数据集样本不足的场景下，GAN更为适用。

在人像合成领域，多域转换的发展最近已经有了进展 [16], [17], [18], [6]。从这些研究中可以看出，采用基本的生成对抗网络进行处理主要有这些缺点：(i) n 个域需要训练 $n(n - 1)$ 数量个转换模型；(ii) 在学习特定的转换模型时，不可利用其他域的数据。对此，诸如CycleGAN、IcGAN [19]、StarGAN [18]、ContrastGAN [17]等可实现图像到图像的多域转换方案或模型被陆续提出，最近的ModularGAN [16]、ComboGAN [20]和XGAN [21]将网络模块化为多个部件开启了另一种思路。

- **医学影像合成的研究现状** 一些研究通过不同模态数据之间的转换 [22], [23], [24], [25]来减少诊断和治疗中给医生和病人带来不必要的代价（例如病人辐射剂量的减少），甚至提高治疗的可行性 [24]，同时不同模态数据之间的转换可以很好的缓解数据样本稀少的难题 [12]。在GAN之前，一些研究使用图字典映射 [24]、稀疏编码 [26], [27]，CNN [28]等探索了医学影像的跨模态转换。在GAN展现了强大的生成能力之后，出现了许多基于GAN的医学影像转换研究 [22], [23], [29], [28], [25]。许多研究使用GAN实现了更高质量的转换结果 [16], [17], [30], [18]。最近一些研究实现了基于未配准的多模态数据的转换 [22], [31]。GAN逐渐被应用到各个部位的器官，诸如脑部MRI到CT图像的转换 [23], [25]、视网膜血管注释到图像的转换 [32]、采用CycleGAN [30]的无监督心脏MRI到CT图像的相互转换与分割 [23]等。当前，在医学影像处理任务中，GAN被广泛应用于医学影像重建 [33], [34]、合成 [12], [32], [14], [15]、转换 [22], [23], [29], [28], [25]、超分辨率 [35], [36]等各类研究。在CycleGAN [30]实现了不成对的图像到图像转换之后，许多基于CycleGAN的研究从各个角度对GAN进行了发展。

但在当前的医学影像合成的研究中，两模态之间的转换合成很多 [22], [23], [24], [27], [29], [28], [25]，对多模态的研究依然稀少 [37], [31], [12]。对于多模态影像的合成，[37]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，[31]进一步实现了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。[38]针对医学图像配准进行了深入研究。[12]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。

[32]研究了基于变分自编码器(VAE) [39], [40]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。[41]在更广泛的数据集上采用Sobel算子实现对医学影像结构信息草图的提取，并通过GAN实现的草图的随机生成，再进一步实现对医学影像的合成。本文基于前述这些研究，对多模态医学影像的合成的各个环节进行更科学的规划和改进，实现了多个数据集上更清晰简明的结构特征图的提取、更稳定的基于VAE的结构特征图的随机合成、更逼真的基于条件GAN的配准多模态医学影像的合成、可控可验证的病灶信息的添加和合成、合成数据的可用性的量化验证。

第三章 基础方法和数据集

3.1 数据集

在我们的实验中使用了六个公开数据集：

- **BRATS2015数据集** [42]公开数据集BRATS2015包含T1 / T2 / T1c / Flair的四个配准模态脑部3D MRI。训练集每个模态有274个大小为 $155 \times 240 \times 240$ 张图及对应的274张肿瘤分割标签图。我们将样本按9:1重新划分训练集和测试集，然后取每个3D MRI的55-105之间的50个切片构建2D数据集。在数据预处理时，我们已对每个图像都进行标准化。
- **Kaggle Chest X-Ray数据集**¹该数据集包括5863个病毒性肺炎、细菌性肺炎和正常肺部的正面2D X射线灰度图，图片尺寸从384*127到2772*2304不等。在数据预处理时，我们已对每个图像都进行标准化，尺寸缩放为512*512。
- **Kaggle Lung CT数据集**²该数据集包含267个胸部至腹部横向截面的2D CT灰度图，尺寸为512*512。在数据预处理时，我们已对每个图像都进行标准化。
- **DRIVE视网膜图像数据集**³该数据集的训练集和测试集各包含20张2D彩色眼底视网膜照片，尺寸均为565*584，训练集还有20张对应的视网膜血管主视图。在数据预处理时，我们已对每个图像都进行标准化，尺寸统一插值为512*512。
- **FIRE视网膜图像数据集**⁴该数据集包含268张2912*2912的2D彩色眼底视网膜照片。在数据预处理时，我们已对每个图像都进行标准化，尺寸缩放为512*512。
- **天池全球数据智能大赛(2019)数据集**⁵该数据集包含肺部3D CT扫描共1837张，训练集1470张，测试集145张。训练集提供的标注信息为：中心坐标+直径（单

¹<https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>

²<https://www.kaggle.com/kmader/finding-lungs-in-ct-data/data/>

³<http://www.isi.uu.nl/Research/Databases/DRIVE/>

⁴<https://projects.ics.forth.gr/cvrl/fire/>

⁵<https://tianchi.aliyun.com/competition/entrance/231724/information>

位为mm) +类别(1-结节, 2-肺密度增高影, 3-肺气肿或肺大泡, 5-索条, 31-动脉硬化或钙化, 32-淋巴结钙化, 33-胸膜增厚)。在数据预处理时, 我们根据标注信息切取有标注的slice及其前后slice组成的3通道图, 再对每个图像都进行标准化, 尺寸缩放为512*512。这样我们得到一个包含17156张512*512*3的CT图, 新的标签为中间通道的原始标签。

3.2 基于CNN的图像分类

图像分类是指采用设计的算法和模型用给定的一组每张图像都被标记了对应的类别的图像集作为训练集进行训练和学习, 然后再对另一组新的测试图像集预测其标签类别的任务。卷积是传统数字图像处理中对图像进行的一种基本操作, 也被称为滤波。卷积神经网络(CNN)是在传统神经网络的基础上应用了卷积操作, 卷积操作与激活函数、批归一化、池化操作、全连接操作等多层复合就得到了一个复杂的神经网络, 再通过神经网络的反向传播机制, CNN就可以自动学习卷积核的参数, 从而能学会处理十分复杂的图像处理任务。深度学习正是从基于CNN的图像分类开始发展。1998年首个CNN模型LeNet提出, 但直到2012年ILSVRC (ImageNet Large Scale Visual Recognition Challenge) 中CNN模型AlexNet [43]以高出10%的正确率力压第二名取得冠军, 卷积神经网络的巨大优势才被人们发现。此后CNN模型进入了飞速发展期, VGGNet [44]、GoogleNet [45]、ResNet [46]等经典网络相继提出, 图像分类任务也随之快速发展。如图所示, 类似于VGGNet的CNN分类模型, 接收图像作为输入, 输入一个类别概率向量, 该向量与输入的类别标签通过损失函数求得损失, 再对损失函数求导, 即可反向传播得到全部模型参数的梯度, 再对模型参数执行梯度更新, 即可使得模型参数得到学习和训练, 通过这样不断地学习, 模型参数最终收敛。使用训练好的VGGNet模型分类时, 通过`argmax()`函数对预测的类别概率向量进行处理即可得到预测的图像类别。

3.3 生成对抗网络

生成式对抗网络(GAN, Generative Adversarial Networks) [47]提出后, 与卷积神经网络结合成为一种无监督学习深度学习模型 [48]。GAN主要包括了两个部分, 即生

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

图 1: VGGNet模型结构设置。随着添加更多的层(添加的层以粗体显示), 配置的深度从左(A)到右(E)逐渐增加。卷积层参数标记格式为“conv< receptive field size>-< number of channels>”。为了简单起见, 没有显示ReLU激活函数。

成器网络 G （Generator）和判别器网络 D （Discriminator），两者不断博弈，进而使 G 学习到数据的分布，如果用到图片生成上，则训练完成后， G 可以从一段随机数中生成逼真的图像。 G 接收一个随机的噪声 z ，通过这个噪声生成图像 x ； D 输入图片 x ，输出该图片为真实图片的概率 $D(x)$ ，如果为1，就代表100%是真实的图片，而输出为0，就代表是完全不真实的图片。训练过程中，生成网络 G 的目标就是尽量生成真实的图片去欺骗判别网络 D 。而 D 的目标就是尽量辨别出 G 生成的假图像和真实的图像。这样， G 和 D 构成了一个动态的“博弈过程”，最终的平衡点即纳什均衡点（Nash equilibrium point）。训练时通过损失函数使二者形成对抗，最终两个网络达到动态均衡的标志是生成器生成的图像接近于真实图像分布，而判别器识别不出真假图像，对于给定图像的预测为真的概率基本接近0.5（处于随机猜测水平）。GAN的一个典型损失函数如下：

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)}[\log(D(x))] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

相比较传统的模型，GAN存在两个不同的网络，而不是单一的网络，并且训练方式采用的是对抗训练方式。GAN中 G 的梯度更新信息来自判别器 D ，而不是直接来自数据样本。GAN具有以下优点：GAN作为一种生成模型，无需复杂的损失函数设计，相比较受限玻尔兹曼机（Restricted Boltzmann machine, RBM）和Generative Stochastic Networks（GSNs）[49]等其他生成模型只用到了反向传播，而不需要复杂的马尔科夫链；GAN相比一般CNN模型，可以产生更加清晰、真实的样本；GAN采用无监督的学习方式训练，可以被广泛用在无监督学习和半监督学习领域；GAN应用场景丰富，比如图片风格迁移、超分辨率、图像补全、去噪。

3.4 变分自编码器

变分自编码器（Variational auto-encoder, VAE）[50]是一类重要的生成模型，它于2013年由Diederik P.Kingma和Max Welling提出，通过2016年Carl Doersch的介绍论文[51]得到飞速发展。VAE与GAN一样，是无监督学习最具前景的方法之一。VAE与GAN两个的目标基本是一致的，即希望构建一个从隐变量 z 生成目标数据 x 的模型，但是实现上有所不同。更准确地讲，两者假设了 z 服从某些常见的分布（比如正态分布或均匀分布），然后希望训练一个模型 $x = G(z)$ ，这个模型能够将原来的概

率分布映射到训练集的概率分布，VAE包含编码器 E 和解码器 G ，编码器将数据分布的高级特征映射到数据的低级表征，低级表征叫作本征向量（latent vector），即 z 。解码器接收数据的低级表征，然后输出同样数据的高级表征，即重建的数据 x_r 。

一个标准VAE中，隐变量 z 服从标准正态分布，即 $z \sim \mathcal{N}(0, 1^2)$ 。编码器 E 将原始数据 x 拟合为一个均值 $\mu(x)$ 和方差 $\sigma(x)$ ，通过重采样技巧构建了一个条件正态分布 $\tilde{z} = \mu(x) + \exp(0.5 \times \sigma(x)) \times z$ ，再用解码器 G 对 \tilde{z} 解码得到 x_r 。VAE使用KL散度来度量两个原始数据的概率分布和重建数据的概率分布之间的差异，以KL散度最小化为优化目标，由此推导得到的损失函数如下所示，详细的推导过程可以参见原论文：

$$\min_{E,D} L(E, D) = \mathbb{E}_x[\|x - D(\tilde{z})\|_2^2], \quad (2)$$

$$\text{s.t. } \tilde{z} = \mu(x) + \exp(0.5 \times \sigma(x)) \times z, \quad (3)$$

$$[\mu(x), \sigma(x)] = E(x), \quad (4)$$

$$z \sim \mathcal{N}(0, 1^2). \quad (5)$$

3.5 语义分割

图像语义分割是数字图像处理领域一个经典任务，指将图像像素按照图像中表达语义含义的不同进行分组（Grouping）或分割（Segmentation）。在深度学习应用到计算机视觉领域之前，人们使用TextonForest和随机森林分类器进行语义分割，2015年全卷积神经网络（FCN）[52]首次将深度学习应用在图像语义分割任务上，通过对全尺寸图片进行端到端的分割取得了十分显著的提升。此后U-net [53]在生物医学图像的语义分割任务上采用编码器-解码器结构并应用skip connections取得了更惊人的分割效果，启发了后来的许多研究。如图所示，U-net接收一张图片输入，输出为分割结果的概率矩阵，通过`argmax()`函数即可得到分割掩膜标签。

3.6 目标检测

图像目标检测同样是数字图像处理领域经典任务之一。目标检测任务是判断图像物体的是否是目标类别并在图中标记出目标类别物体的位置。近几年来，随着深度学习的崛起，基于卷积神经网络的目标检测算法取得了很大的突破。当前比较流

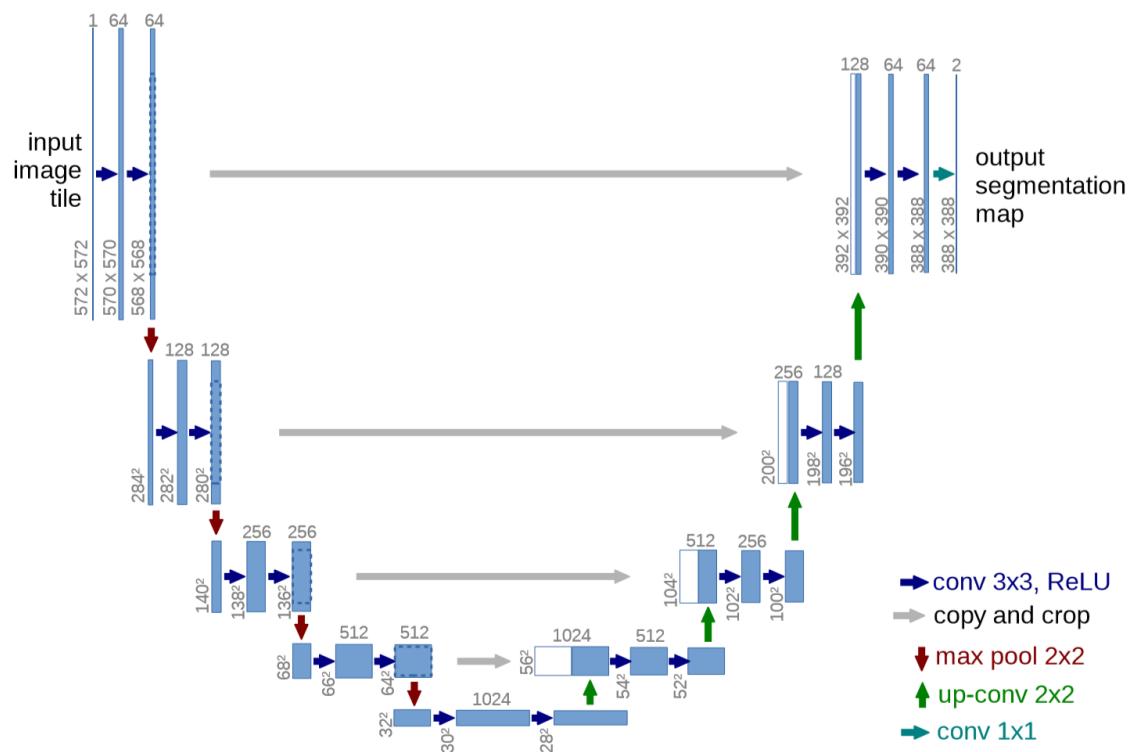


图 2: U-net模型结构。

行的模型可以分为两类：一类是基于Region Proposal的R-CNN [54]类模型（R-CNN, Fast R-CNN [55], Faster R-CNN [56]等），它们是two-stage的，需要先前向推理产生目标候选框，然后再对候选框进行分类和回归；另一类是YOLO [57]、SSD [58]这类one-stage算法，其仅用单个卷积神经网络即可直接预测不同目标的类别与位置。如图所示，SSD基于VGG16模型进行修改，先将输入图片到预训练好的分类网络中

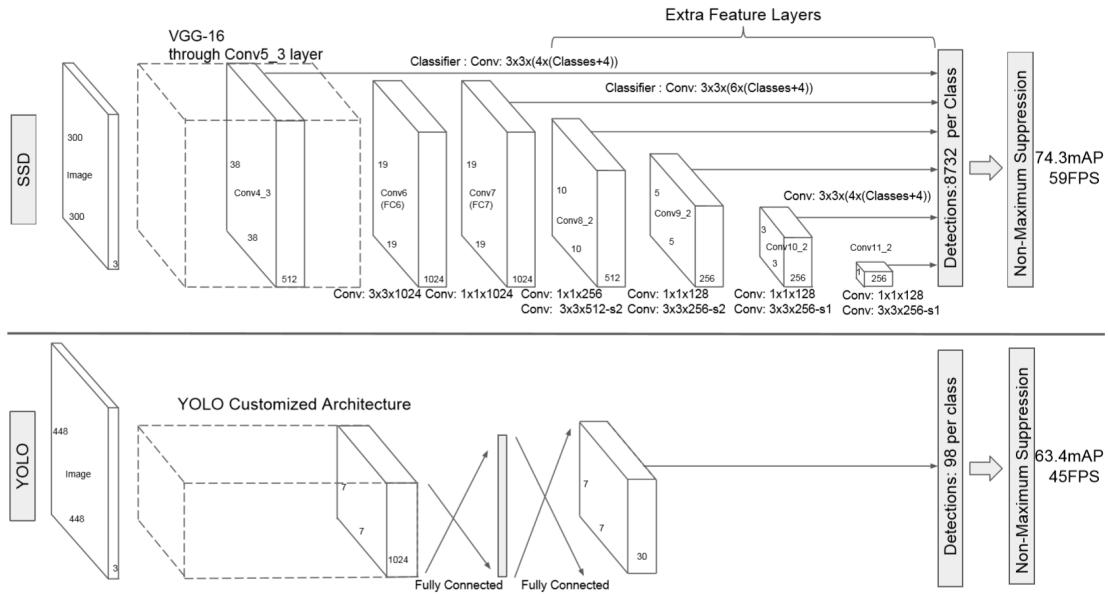


图 3: YOLO和SSD模型结构的对比。

来获得不同大小的特征映射，通过抽取的多层卷积层的feature map构造6个不同尺度大小的bbox，然后分别进行检测和分类，生成多个bbox，再将不同feature map获得的bbox结合起来，经过非极大值抑制（NMS）方法来抑制掉一部分重叠或者不正确的bbox，生成最终的bbox集合，这就是最后输出的目标检测结果。

第四章 带病灶标签的配准多模态医学影像的合成

4.1 整体架构

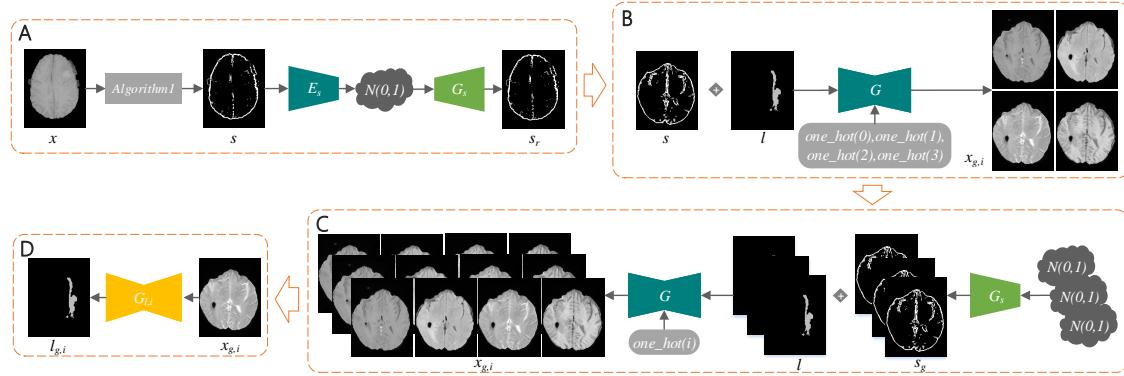


图 4: 整体架构。A 是结构特征图的提取和生成阶段。B 是多模态影像合成阶段。C 是合成数据集构建阶段。D 是合成数据可用性验证阶段。

图 4 所示包括四个主要阶段。在结构特征图提取和生成阶段，我们主要的预期产出为一个结构特征图生成器，该生成器可以从随机正态分布矩阵生成结构特征图。在多模态影像合成阶段，我们以结构特征图为输入训练一个条件生成器，该条件生成器可以根据不同条件合成不同模态的影像。若需要在合成影像中添加指定的病灶信息，则可将病灶标签与结构特征图融合后作为输入，同时通过病灶标签生成器添加病灶生成指导损失。若对合成多模态的影像有高配准度要求，则可以通过模态转换器对合成的多模态影像进行模态转换一致性约束保证像素级配准。在合成数据集构建阶段，我们使用前面两个阶段生成的模型从随机正态分布矩阵合成配准的多模态影像。最后是对合成数据可用性验证阶段，我们对合成数据在智能医学影像处理任务中的可用性进行实验验证，该阶段详见下一章节。

4.2 结构特征图的提取和生成

GAN直接从随机噪声中生成的医学图像很难生成现实的结构信息。我们将提供基本轮廓和结构信息的图像称为结构特征图。例如，视网膜血管分布图可以看作是视网膜图像的结构特征图 [32]。结构特征图可以为合成医学图像提供必要的基本指导。

在合成医学影像时，一些研究从组织分割标签 [12] 获取了基本的结构信息。但是，诸如视网膜血管图和脑组织分割标签之类的一般结构特征，在从原始图像中提取之前需要额外的数据来训练一个提取模型。为此，我们首先设计了一种直接从医学影像直接提取结构特征图的方法，该方法具有操作快速，无需训练，无需附加数据的优点。

4.2.1 结构特征图提取方法

在传统的数字图像处理方法中，Roberts算子、Prewitt算子、Sobel算子等是出色的边缘检测算子。Sobel运算符通常用于处理医学图像。如算法 1 中所示，我们探索了一种从Sobel算子生成的边缘检测图中进一步提取结构特征图的方法。在算法 1 中，我

Algorithm 1 Structural feature map extraction

- 1: Input a real grayscale image x , pixel threshold α and β and γ , gaussian kernel variance σ_1 and σ_2
 - 2: $s_1 = \text{reduce_min}(\text{sobel}(x))$
 - 3: $s_2 = \text{reduce_max}(\text{sobel}(x))$
 - 4: $s_1 = \text{gaussian_blur}(s_1, \sigma_1)$
 - 5: $s_2 = \text{gaussian_blur}(s_2, \sigma_1)$
 - 6: $s_1 = \text{mean}(s_1) - s_1$
 - 7: $s_2 = s_2 - \text{mean}(s_2)$
 - 8: $s_1 = \text{ones} \times (s_1 > \alpha)$
 - 9: $s_2 = \text{ones} \times (s_2 > \alpha)$
 - 10: $s = \text{ones} \times ((s_1 + s_2) > 0)$
 - 11: $s = \text{gaussian_blur}(s, \sigma_2)$
 - 12: $s = \text{ones} \times ((s_1 + s_2) > \beta)$
 - 13: $s = \text{medfilt}(s)$
 - 14: $s = s \times (x > \gamma)$
-

们使用Sobel检测算子 $\text{sobel}()$ 从真实图像中提取水平和垂直边缘检测图。每个边缘检测图执行最大规约 $\text{reduce_min}()$ 和最小规约 $\text{reduce_max}()$ 以获得两个融合水平垂直边缘检测图的结果图，对两张图进行 3×3 的高斯模糊 $\text{gaussian_blur}()$ 可以对线条轮廓加

粗，然后每个融合图计算与通过均值函数 $mean()$ 计算出的平均像素值的差可以去掉大部分背景像素只保留最突出的线条轮廓。根据像素阈值对两个差异图进行二值化，然后对两个二进制图像求和获得完整的轮廓线条，然后完全二值化。再进行一次高斯模糊可以对线条加粗并能使有断点的线条相连成整体，最后和原始图的二值化掩膜相乘可以完全去掉对应于原始图背景区域的噪声，再用 3×3 的中值滤波函数 $medfilt()$ 去掉剩余的孤立的噪点，即可得到我们需要的干净清晰又完整简洁的结构特征图。对于一些有病灶结构信息的结构特征图，我们可以使用病灶分割标签掩膜去除结构特征图中的原始病灶部分的结构信息，或者弃用。本文中，我们对BRATS2015数据集提取的结构特征图使用肿瘤分割标签掩膜去除了结构特征图中的原始肿瘤结构信息，对Kaggle Chest X-Ray数据集中弃用肺炎影像而只选用正常影像生成结构特征图。

[41]利用Sobel边缘检测方法提取初始结构边界，然后利用高斯低通滤波去除孤立噪

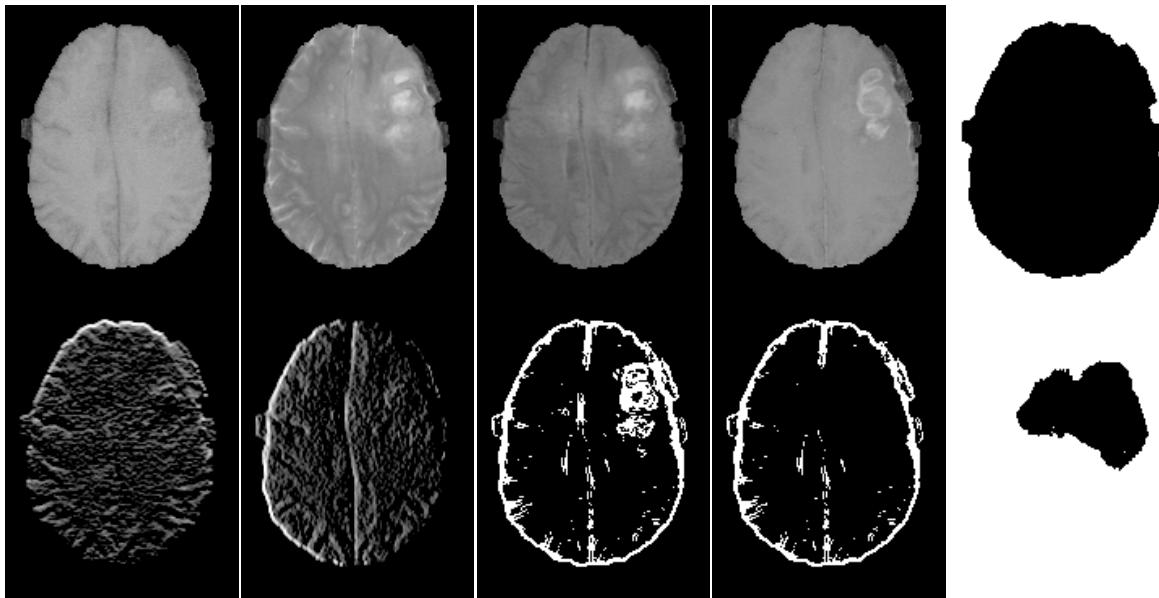


图 5: BRATS2015数据集中医学图像和提取的结构特征图。其中子图依次为T1、T2、T1c、Flair四个模态的MRI、从Flair提取出来的掩膜、从Flair采用Sobel算子提取出来的水平向和垂直向边缘检测结果图、从Flair提取出来的结构特征图、采用肿瘤分割标签掩膜去掉肿瘤病灶结构信息的从Flair提取出来的结构特征图、肿瘤分割标签的二值化掩膜。

声和像素，最后利用一个由开孔过程和闭孔过程组成的形态学操作进一步去除噪声，

填充囊状结构得到了结构草图。和 [41] 的方法相比，我们的方法分别考虑了 Sobel 边缘检测结果的高像素值轮廓和低像素值轮廓，最后再对轮廓组合得到了更完整的轮廓信息。同时两次高斯模糊的应用突出了轮廓线条信息，采用通过与像素均值作差后的二值化即可分离轮廓线条和背景，这与开孔闭孔分离背景的方法相比保留的轮廓复杂程度更低、线条更清晰。最后我们进行的去噪过程能完全去除对应原始图背景区域的部分的噪声和器官内绝大部分噪点而不破坏轮廓线条。从生成结果看，我们的结构特征图更加清晰明了、轮廓更加干净简洁、线条更加符合原图的视觉呈现。

4.2.2 结构特征图生成训练

Algorithm 2 Mask Extraction

- 1: Input a real grayscale image x , pixel threshold α , the expanded pixel value p
 - 2: $m = 1.0 - \text{ones} \times (x > \alpha)$
 - 3: $\text{new_size} = [x.\text{width}() + p, x.\text{length}() + p]$
 - 4: $m = \text{resize}(m, \text{new_size})$
 - 5: $m = \text{crop_padding}(m, p)$
 - 6: $m = \text{medfilt}(m)$
-

生成结构特征图时，[12]仍需要输入真实图像以获取生成的结构特征图，这大大减少了合成数据的多样性。[32]实现了一种基于VAE的方法，用于根据多维正态分布生成视网膜血管分布图。[41]采用GAN从随机噪声合成结构草图。以此为基础，我们设计了一种混合网络，结合了VAE和GAN的特征，来提高合成训练的稳定和鲁棒，实现从随机正态分布矩阵生成具有更好的多样性并且无需其他训练标签的结构特征图。此外，我们训练了一个生成器 G_m ，该生成器从大脑结构特征图中获取了目标器官区域的掩膜，以供以后用于匹配病变标签。生成器与结构特征图生成的训练同步。在 G_m 训练期间，算法 2 提取的掩膜用作标签数据，其中 $\text{resize}()$ 为最近邻插值函数， $\text{crop_padding}()$ 为边距裁切函数。如图 6 所示，具体的训练过程如下：

- 结构特征图 s 使用算法 1 从 x 获得，掩膜 m 通过算法 2 从 x 获得；
- 通过算法 1 从 x 获得结构特征图 s ，通过算法 2 从 x 获得掩膜 m ，从多维正态分布 $\mathcal{N}(0, 1^2)$ 中采样可获得随机噪声 z ；

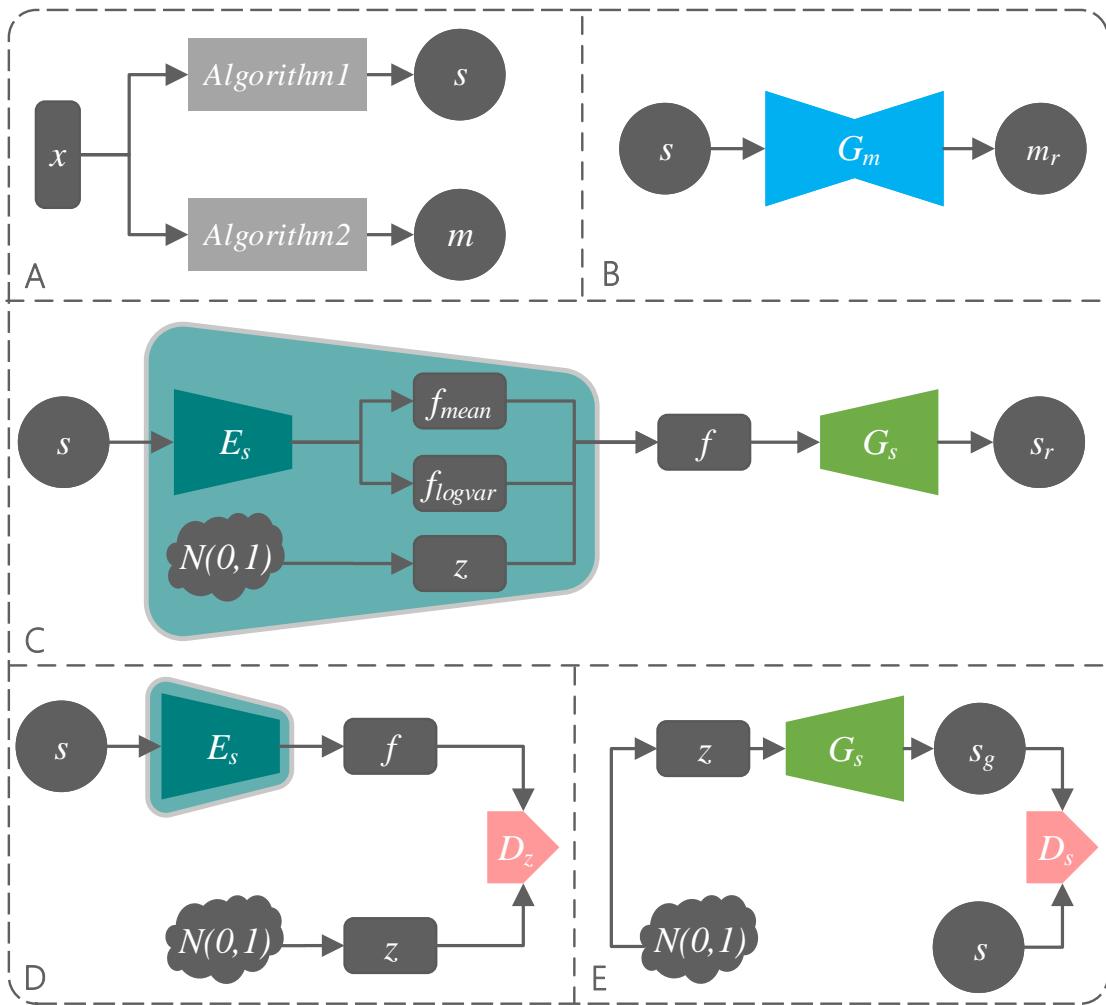


图 6: 结构特征图生成训练。 x 是输入的真实影像, s 是结构特征图。 E_s 是VAE编码器, 输出编码矩阵 f_{mean} 和 f_{logvar} 。 z 是来自多维正态分布 $\mathcal{N}(0, 1^2)$ 的随机噪声采样, 而 f 是近似正态分布矩阵。 G_s 是VAE解码器, s_r 是重构的结构特征图, 而 s_g 是生成的随机结构特征图。 D_s 是结构特征鉴别符。 G_m 是掩膜生成器, m_r 是生成的掩膜。

- 使用 s 和 m 单独训练一个从 s 生成 m 的掩膜生成器 G_m ;
- 使用 VAE 编码器 E_s 对 s 进行编码, 以获得 f_{mean} 和 f_{logvar} , 再与随机生成的噪声 z 一起构造近似正态分布矩阵 $f = f_{mean} + \exp(0.5 \times f_{logvar}) \times z$, 再用 VAE 解码器 G_s 解码 f 以获得重建的结构特征图 s_r ;
- 以 VAE 编码器 E_s 对 s 编码生成的近似正态分布矩阵 f 为负样本, 以随机生成的噪声 z 为正样本, 对特征图分布鉴别器 D_z 和 E_s 进行对抗训练;
- 以 VAE 解码器 G_s 对随机生成的噪声 z 解码生成的随机结构特征图 s_g 为负样本, 以 s 为正样本, 对结构特征图鉴别器 D_s 和 G_s 进行对抗训练;

E_s 、 D_z 和 D_s 均在 VGG11 模型结构的基础上进行了调整适配, G_m 和 G_s 在 U-net 的模型结构的基础上进行了调整适配, 各自的损失函数如下:

- 阶段B: 掩膜生成损失

$$\mathcal{L}_m(G_m) = \mathbb{E}_{m,s}[\|m - m_r\|_2^2], \quad (6)$$

其中 $m_r = G_m(s)$ 。

- 阶段C: 结构特征图重建损失

$$\mathcal{L}_r(E_s, G_s) = \mathbb{E}_{s,f,m}[\|s - s_r\|_2^2 + \|m_r \times s_r\|_2^2], \quad (7)$$

其中 $s_r = G_s(f)$ 。

- 阶段D: 特征图分布对抗训练损失

$$\mathcal{L}_{d1}(D_z) = \mathbb{E}_{s,z}[\|D_z(z) - 1\|_2^2 + \|D_z(f)\|_2^2], \quad (8)$$

$$\mathcal{L}_{g1}(E_s) = \mathbb{E}_z[\|D_z(f) - 1\|_2^2], \quad (9)$$

其中 $f = f_{mean} + \exp(0.5 \times f_{logvar}) \times z$, $[f_{mean}, f_{logvar}] = E_s(s)$.

- 阶段E: 结构特征图对抗训练损失

$$\mathcal{L}_{d2}(D_s) = \mathbb{E}_{s,z}[\|D_s(s) - 1\|_2^2 + \|D_s(s_g)\|_2^2], \quad (10)$$

$$\mathcal{L}_{g2}(G_s) = \mathbb{E}_z[\|D_s(s_g) - 1\|_2^2 + \|m_g \times s_g\|_2^2], \quad (11)$$

其中 $s_g = G_s(z)$, $m_g = G_m(s_g)$ 。

4.2.3 结构特征图与随机噪声的融合

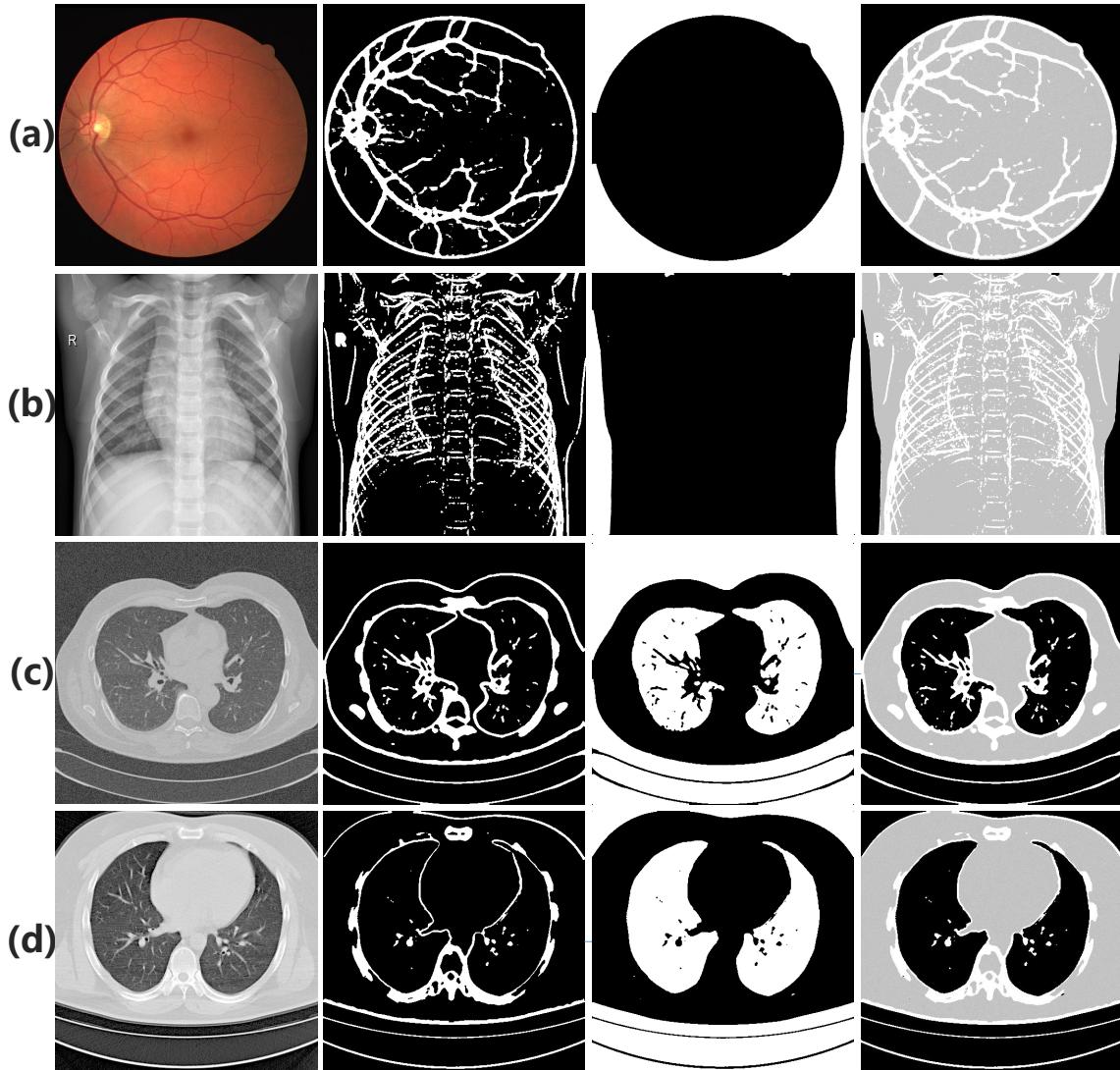


图 7: 各数据集中医学图像和提取的结构特征图。(a) 为DRIVE视网膜图像数据集。(b) 为Kaggle Chest X-Ray数据集。(c) Kaggle Lung CT数据集。(d) 为天池全球数据智能大赛(2019)数据集。以上子图中从左到右依次为原图、结构特征图、掩膜、融合了随机噪声结构特征图。

在使用结构特征图合成医学影像时，由于结构特征图是简单二值图，直接使用结构特征图作为输入会减少输入的多样性和随机性，在使用小数据集时，训练将尤其困难。在普通的GAN的训练中，以为随机噪声矩阵输入，可以带来无限的多样性。因

此，我们设计了如下的计算公式，可以融合结构特征图的结构信息和随机噪声的随机信息：

$$s' = s + z' \times (1 - m) \times (1 - s), \quad (12)$$

其中， z' 是从均匀分布 $\mathcal{U}(\alpha_1, \alpha_2)$ 采样的随机噪声，生成最小值和最大值分别为 α_1 和 α_2 的随机噪声矩阵，默认 α_1, α_2 取值分别为0.5和0.6， m 是与结构特征图 s 配对的二值化掩膜。如图7所示，最终得到的融合结构特征图 s' ，既保留了全部的结构信息，又有丰富的随机信息，同时与预期生成的医学影像更为接近，降低了学习难度。

4.3 多模态影像的合成

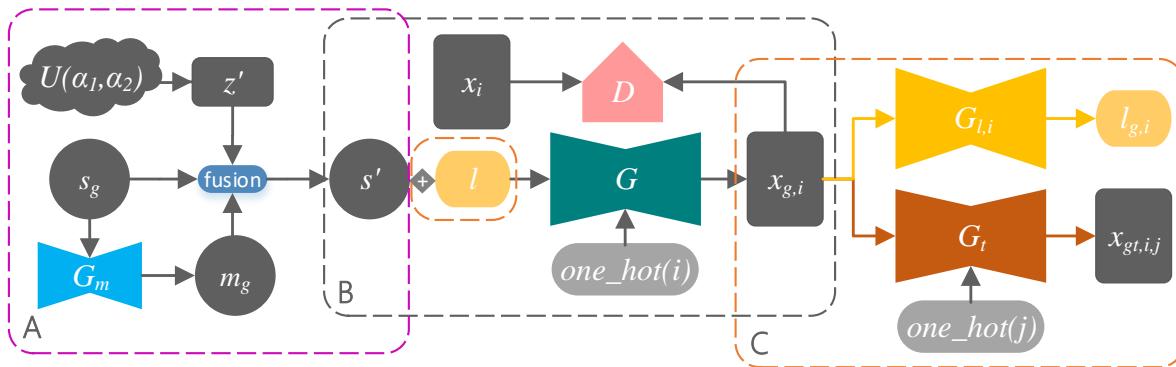


图 8: 多模态影像的合成。紫色框A为结构特征图与随机噪声融合过程，灰色框B为核心过程，黄色框内为可选过程。 s_g 是随机合成的的结构特征图， m_g 是根据 s_g 生成的掩膜，， z' 是从均匀分布 $\mathcal{U}(\alpha_1, \alpha_2)$ 采样的随机噪声， s' 是融合了噪声的结构特征图， l 是随机选择的真实病灶标签。 $one_hot(i)$ 是模态*i*的独热矩阵表示。 G 是条件生成器， D 是鉴别器， x_i 是模态*i*的真实影像。 $G_{l,i}$ 是模态*i*的病灶标签生成器， $l_{g,i}$ 是病灶生成器从 $x_{g,i}$ 还原生成的病灶标签。 G_t 为模态转换器， $x_{gt,i,j}$ 是 $x_{g,i}$ 转换合成的模态*j*的图像。

如图 8所示，首先，我们通过训练好的 G_s 生成结构特征图 s_g ，融合随机噪声后得到 s' ，再根据需要可添加指定的病灶标签 l ，与 s' 在通道向堆叠融合得到最后的融合图。然后，我们使用一个接受独热条件矩阵的条件生成器 G ，对融合图编码后添加不同的条件矩阵 $one_hot(i)$ ，然后再解码生成不同模的合成图像态 $x_{g,i}$ 。我们使用鉴别器 D 提供的对抗性损失和类别指导损失来指导合成图像逼近真实影像的分布。合成的

多模态影像我们可以根据需要通过病灶标签生成器 $G_{l,i}$ 添加病灶生成指导损失、通过模态转换器 G_t 添加模态配准损失。

D 在VGG11模型结构的基础上进行了调整适配， G 在U-net的模型结构的基础上进行了调整适配。在单模态时， D 为单线结构， G 为U-net生成器在首层接收输入， G 与 D 组成一组基本GAN结构。在多模态时， D 为在最后两层双线输出两个结果，即真假鉴别结果矩阵和类别鉴别结果矩阵， G 以U-net生成器为基本结构，在编码缩小尺寸阶段的最后层输出后叠加输入的条件矩阵，再进行后续的解码还原尺寸过程， G 与 D 组成一组ACGAN [62]结构。

多模态影像合成过程的损失项如下，其中 $x_{g,i}$ 是模态*i*的合成图像， $d(x_i)$ 和 $c(x_i)$ 是鉴别器输出 $D(x_i)$ 的真假鉴别结果和模态类别鉴别结果， $d(x_{g,i})$ 、 $c(x_{g,i})$ 为 $D(x_{g,i})$ 的真假鉴别结果和模态类别鉴别结果。

- 多模态合成影像对抗性训练损失

$$\begin{aligned} \mathcal{L}_{d2}(D) = \mathbb{E}_{x,s_g,l} & \left[\sum_{i=0} (\|d(x_i) - 1\|_2^2 + \|d(x_{g,i})\|_2^2 + \right. \\ & \left. \|c(x_i) - i\|_2^2 + \|c(x_{g,i}) - i\|_2^2) \right], \end{aligned} \quad (13)$$

$$\mathcal{L}_g(G) = \mathbb{E}_{s_g,l} \left[\sum_{i=0} (\|d(x_{g,i}) - 1\|_2^2 + \|c(x_{g,i}) - i\|_2^2) \right]. \quad (14)$$

其中 $x_{g,i} = G(concat(s', l), one_hot(i))$ ， $s' = s_g + z' \times (1 - m_g) \times (1 - s_g)$ ， $m_g = G_m(s_g)$ ， z' 是从均匀分布 $\mathcal{U}(\alpha_1, \alpha_2)$ 采样的随机噪声且 $\alpha_1 = 0.5$ 、 $\alpha_2 = 0.6$ ， $concat()$ 为通道堆叠连接函数； $[d(x_i), c(x_i)] = D(x_i)$ ， $[d(x_{g,i}), c(x_{g,i})] = D(x_{g,i})$ 。

- 病灶生成指导损失

$$\mathcal{L}_{les}(G) = \mathbb{E}_{s_g,l} \left[\sum_{i=0} (\|l - l_{g,i}\|_2^2) \right], \quad (15)$$

其中 $l_{g,i} = G_{l,i}(x_{g,i})$ 。

- 模态配准损失

$$\mathcal{L}_{reg}(G) = \mathbb{E}_{s_g,l} \left[\sum_{j=0} \sum_{i=0, i \neq j} (\|x_{g,i} - x_{gt,j,i}\|_2^2) \right], \quad (16)$$

其中 $x_{gt,i,j} = G_t(x_{g,i}, one_hot(j))$ 。

则多模态合成生成器的总损失为：

$$\mathcal{L}(G) = \mathcal{L}_g(G) + \mathcal{L}_{les}(G) + \mathcal{L}_{reg}(G) \quad (17)$$

对于小数据集，我们可以使用从真实影像提取的结构特征图与真实医学影像进行自监督预训练，来降低对抗性训练的难度。预训练过程的损失函数如下：

$$\mathcal{L}_p(G) = \mathbb{E}_{s,l} \left[\sum_{i=0} (\|x_{g,i} - x_i\|_2^2 + \|x_{g,i} \times m_i - x_i \times m_i\|_2^2) \right]. \quad (18)$$

其中 $x_{g,i} = G(concat(s'_i, l_i), one_hot(i))$, $s'_i = s_i + z' \times (1 - m_i) \times (1 - s_i)$, s_i 为采用Algorithm1从真实影像 x_i 提取出的结构特征图, l_i 为真实影像 x_i 的病灶标签, m_i 为采用Algorithm2从真实影像 x_i 提取出的掩膜, z' 是从均匀分布 $\mathcal{U}(\alpha_1, \alpha_2)$ 采样的随机噪声且 $\alpha_1 = 0.5$ 、 $\alpha_2 = 0.6$ 。在SkrGAN [41]的训练中，一方面始终采用从真实数据集中提取的草图进行训练，另一方面SkrGAN将前述自监督损失与对抗性损失加权求和作为总损失，这使得在数据集较小时，训练样本过少，训练过程可视为原始数据集影像的重建，这使得模型的训练将不充分、极易过拟合且缺乏对合成草图的适应能力。当SkrGAN第一步合成的草图出现比原始草图更丰富的多样性时，未经过合成草图训练的SkrGAN图像生成模型将缺乏对合成的草图的进一步合成能力，最终使得模型合成的影像多样性差。在我们的方案中，我们首先采用真实的结构特征图进行真实影像的重建预训练，此时损失函数即为上述的自监督损失，在模型收敛后再使用大量的合成的结构特征图进行对抗性训练合成真实影像。这样我们通过预训练加速训练进程，通过对抗性训练提高模型泛化能力。

4.4 多模态影像的配准

当需要确保合成的不同模态的影像精确配准时，我们可以通过模态转换器对合成图像再进行模态转换一致性约束。模态转换器通过真实的多模态数据预先训练完成，如图 9所示。模态转换器 G_t 同样为一个接受独热条件矩阵的条件生成器，不同条件矩阵指示转换生成不同模态的影像。模态转换器 G_t 与鉴别器 D_t 组成一组循环生成对抗网络（CycleGAN） [30], D_t 在VGG11模型结构的基础上进行了调整适配， G_t 在UNet的模型结构的基础上进行了调整适配，各自的损失函数如下：

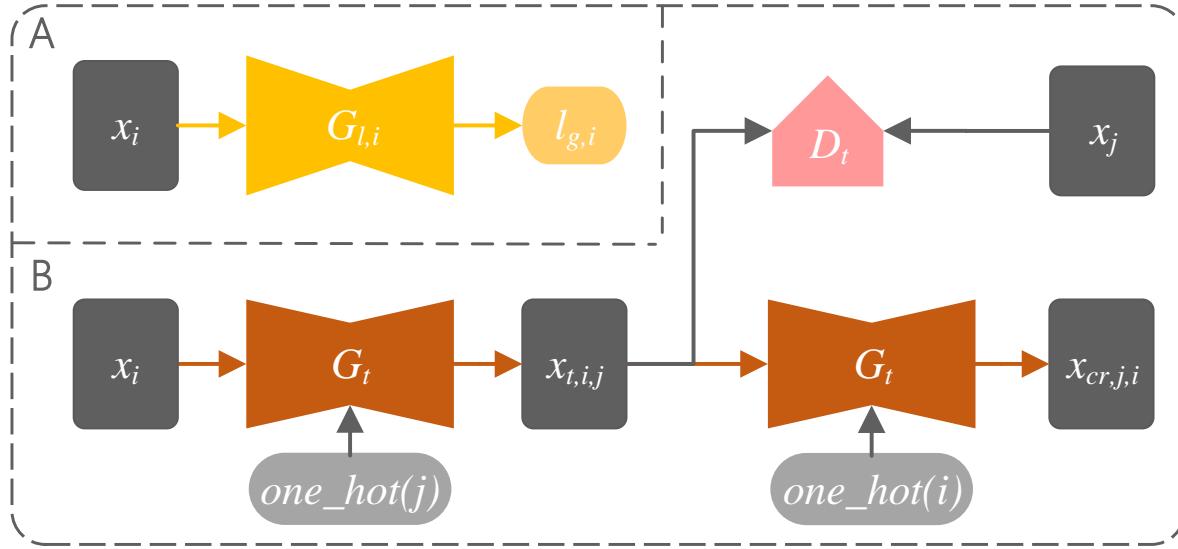


图 9: 模态转换器训练和病灶标签生成器训练

- 模态循环一致性损失

$$\mathcal{L}_{cycle}(G_t) = \mathbb{E}_x [\sum_{j=0} \sum_{i=0, i \neq j} (\|x_i - x_{cr,j,i}\|_2^2)], \quad (19)$$

其中 $x_{cr,j,i} = G_t(x_{t,i,j}, one_hot(i))$, $x_{t,i,j} = G_t(x_i, one_hot(j))$ 。

- 模态转换合成影像对抗性训练损失

$$\begin{aligned} \mathcal{L}_{d3}(D_t) = \mathbb{E}_{x,s,l} [\sum_{i=0} & (\|d(x_i) - 1\|_2^2 + \|d(x_{t,i,j})\|_2^2 + \\ & \|c(x_i) - i\|_2^2 + \|c(x_{t,i,j}) - i\|_2^2)], \end{aligned} \quad (20)$$

$$\mathcal{L}_{g3}(G_t) = \mathbb{E}_{s,l} [\sum_{i=0} (\|d(x_{t,i,j}) - 1\|_2^2 + \|c(x_{t,i,j}) - i\|_2^2)]. \quad (21)$$

其中 $[d(x_i), c(x_i)] = D_t(x_i)$, $[d(x_{t,i,j}), c(x_{t,i,j})] = D_t(x_{t,i,j})$ 。

4.5 病灶信息的添加和合成

当我们需要在合成的多模态影像中添加指定的病灶信息时，我们需要在给生成器输入结构特征图时选取合适的病灶标签与结构特征图堆叠融合后输入。由于随机选择的病灶标签所标示病灶位置可能会出现在结构特征图的目标器官轮廓之外，因此我们

可以使用目标器官的掩膜 m 来过滤病变标签。如果病灶标签所标示病灶位置在 m 的目标器官轮廓内，则可以采用 l ，否则需要重新选择 l 。融合图包含目标部位的基本解剖信息和选定的病灶信息。

如图 9 中所示，为确保合成的多模态图像已根据输入的病灶标签合成了相应的病变内容，我们使用一个病灶标签生成器对每个合成影像的病灶进行提取，还原出输入的病灶标签。病灶标签生成器用真实的影像和标签数据预先训练完成。每个模态均由独立的病灶标签生成器 $G_{l,i}$ 来指导合成，典型的训练过程损失为：

$$\mathcal{L}_{seg}(G_{l,i}) = \mathbb{E}_{l,x}[\|l_i - l_{r,i}\|_2^2], \quad (22)$$

其中 $l_{r,i} = G_{l,i}(x_i)$ 。在实际应用时，损失函数应根据选取的病灶标签生成器和病灶任务的实际需要设计，例如选用U-net作为病灶标签生成器时就可以使用U-net原始的加权交叉熵损失函数 [53]。本文实验中选用的U-net、VGG11、SSD等病灶标签生成器均采用该模型原有的损失函数方案。

4.6 合成数据集的构建

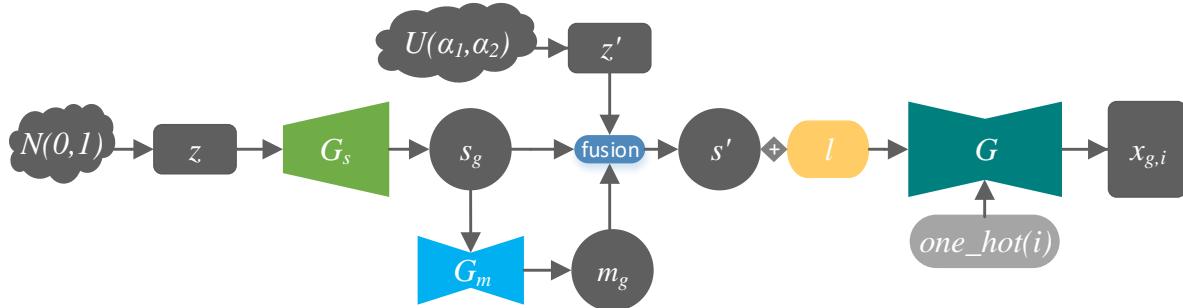


图 10: 构建合成数据集

如图 10 所示，我们可以通过经过训练的结构特征图解码器从随机正态分布矩阵生成任意数量的结构特征图，通过掩膜生成器 G_m 从结构特征图中获取掩膜，再与均匀分布的随机噪声融合可得到融合了噪声的结构特征图。然后，我们可以根据标签的类型对原始标签集进行随机缩放、旋转、平移、翻转或直接生成，以获得随机病变标签集。生成的结构特征图与从随机病变标签集中随机选择的标签融合时，像训练阶段

一样，我们可以通过掩膜生成器 G_m 从结构特征图中获取掩膜，从而选择合适的标签。最后我们将融合图输入生成器，通过添加不同的条件向量即可合成不同模态的多模态影像。

由于可生成的结构特征图数量是无限制的，我们可以根据需要对合成的结构特征图、掩膜和合成影像进行过滤，得到我们需要的由结构特征图，掩膜，病灶标签和多模态合成影像组成的合成数据集。

在基于不同的训练数据和需求时，可以对我们的方案中的一些方法进行选择性应用，在本文中，我们有多个合成任务，具体情况如下：

- **多模态脑肿瘤MRI的合成**我们以BRATS2015数据集作为训练数据，以其肿瘤分割标签为输入的病灶标签，合成T1、T2、T1c、Flair四个模态，使用模态转换器提供模态配准损失，使用一个U-net [53]作为肿瘤分割器提供病灶生成指导损失。此外，在结构特征图生成训练时，我们使用肿瘤分割标签掩膜去除了结构特征图中的原始肿瘤结构信息，在合成多模态MRI训练时，我们通过掩膜对经过数据增强后的肿瘤分割标签过滤后再输入，无预训练过程。
- **眼底视网膜图像的合成**我们以DRIVE视网膜图像数据集和FIRE视网膜图像数据集作为训练数据，无病灶标签和病灶生成指导损失，无模态配准损失，直接合成单模态的彩色眼底视网膜图像，采用在真实影像及其结构特征图上进行的预训练。
- **肺部CT的合成**我们以Kaggle Lung CT数据集作为训练数据，无病灶标签和病灶生成指导损失，无模态配准损失，直接合成单模态的肺部CT，采用在真实影像及其结构特征图上进行的预训练。
- **胸部X-ray的合成**我们以Kaggle Chest X-Ray数据集作为训练数据，将肺炎类别标签扩展为与结构特征图同尺寸的独热矩阵后作为输入的病灶标签，以VGG-11 [44]作为肺炎分类器提供病灶生成指导损失，无模态配准损失，合成具有指定肺炎类别的单模态胸部X-ray，采用在真实影像及其结构特征图上进行的预训练。
- **肺部低剂量CT的合成**我们以天池全球数据智能大赛(2019)数据集作为训练数

据，将目标检测标签扩展为与结构特征图同尺寸的框图后作为输入的病灶标签，以SSD [59]作为病灶检测器提供病灶生成指导损失，无模态配准损失，实现具有根据输入的检测框信息合成的对应病灶的单模态肺部低剂量CT，无预训练过程。

第五章 合成医学影像的性能评估

5.1 评估指标

[41]中同样在一些数据集上实现了较高质量的合成，我们以其中的部分结果作为一个参考基准。这项工作中，我们使用多尺度结构相似性（MS-SSIM），和Freshet Inception距离（FID）来评估合成医学图像的性能。MS-SSIM 是一种广泛使用的指标，用于测量配对图像的相似性，其中 MS-SSIM 越高，性能越好。FID 在像素级别计算真实图像和假图像之间的距离，其中 FID 越低，性能越好。我们使用Dice Score [60]和均方误差（MSE）[61]来评估分割结果，使用敏感度（Sensitivity）、精度（Accuracy）和Area Under the ROC Curve（AUC）来评估血管注释结果，使用精度（Accuracy）来评估我们的分类结果，使用交并比（IOU）来评估检测结果。

5.2 训练设置

如无特殊说明，每个实验的迭代次数等于训练数据集的200个epoch。学习率为 $1e-5$ ，无权重衰减，使用beta为1为0.5的Adam优化器，批处理大小为1。评估结果是2D图像上多模态结果的平均值，每个实验都经过了4次训练保留最佳结果。

5.3 合成方法在脑肿瘤MRI数据集上的对比实验

我们进行了简单的消融对比实验，以验证在多模态合成阶段我们的改进措施的影响。首先我们进行了以相同尺寸的随机噪声代替结构特征图作为MRI生成训练输入的实验来验证添加结构特征图的作用。我们在相同的训练epoch下通过对结构特征图是否融合随机噪声的对比来验证融合随机噪声可以提高模型泛化能力并加快训练收敛。我们进行了无模态配准损失的实验来展示其对边缘配准的矫正作用。我们进行了无病灶生成指导损失的实验来验证其对病灶合成的指导作用。我们通过选择没有掩膜限制的输入标签进行实验来验证掩膜的作用。图11显示了消融对比实验中生成的合成图像的示例。图12为消融实验量化评估结果的柱状图对比情况，不难看出，我们的每一项方法都起到了针对性的提升效力。

表 1: Ablation Experiment Setting on BRATS2015

实验	输入结构特征图	掩膜过滤病灶标签	输入融合随机噪声	模态配准损失 \mathcal{L}_{reg}	病灶生成指导损失 \mathcal{L}_{les}
A	×	×	×	×	×
B	✓	×	×	×	✓
C	✓	✓	×	×	×
D	✓	✓	✓	×	×
E	✓	✓	✓	✓	×
F	✓	✓	✓	✓	✓

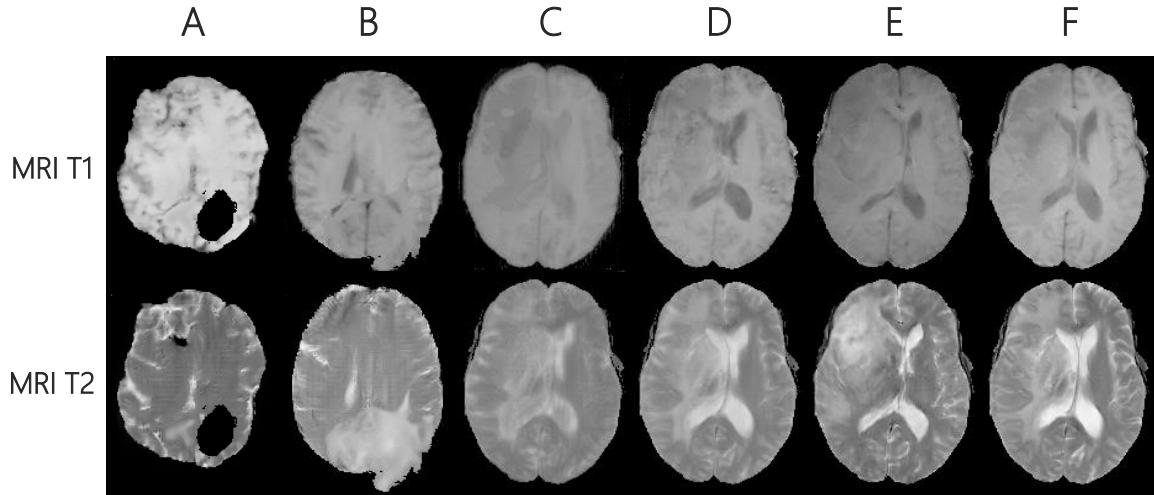


图 11: 消融实验的合成图像。模型A: 用随机噪声替换结构特征图, 由于没有结构特征图在脑轮廓上的约束, 模型A生成的图像符合MRI的特征, 但不符合大脑的结构特征。模型B: 没有掩膜对输入标签过滤, 很容易看出合成的肿瘤超出了大脑的轮廓。模型C: 没有融合随机噪声的结果, 在与其他实验相同的训练epoch后合成质量较差且生成的图像的配准效果不是很好, 也没合成明显的病灶信息。模型D: 没有模态配准损失, 生成的图像的配准效果不是很好, 尤其是边缘细节。模型E: 没有病灶生成指导损失, 可以看出生成的图像中的病变散乱随意, 过度夸张, 与输入的病灶标签不吻合。模型F: 我们完整的模型, 合成了与输入病灶标签吻合的病灶信息, 配准度高, 合成质量高。该组实验中, 实验C-F均采用相同的结构特征图作为输入。

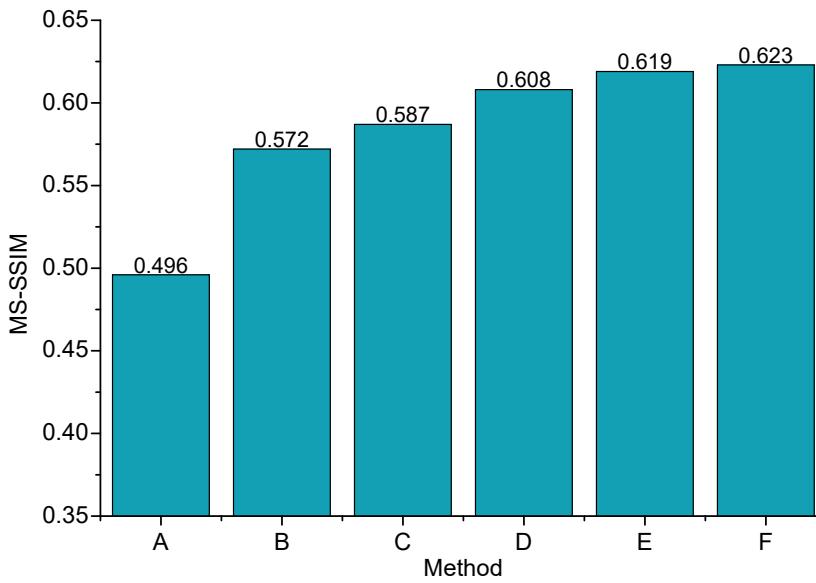


图 12: 脑肿瘤MRI数据集上消融实验量化评估结果的柱状图。

5.4 合成方法在不同数据集上的量化评估

如表 2 和表 3 所示，我们的方法在各个数据集上进行了量化评估，并与当前最先进的方法进行对比。

表 2 中，我们与 [41] 中的数据进行了对比，数据表明，在 Chest X-ray 和 Lung CT 两个数据集上，我们的方法整体合成质量优于表中采用草图为输入的 SkrGAN 和采用噪声为输入的其他方法。

表 3 中，由于 [41] 中两份数据集的不可获取或复现，我们在 Color Fundus 和 BRATS MRI 两个与 [41] 中所用的数据集类似的数据集上，对各个方法进行了复现，复现结果与表 2 中两个数据集上的结果指向的结论是一致的。BRATS MRI 数据集上，我们的方案可直接进行多模态合成，其余方案需对每个模态单独训练后合成，最后的评估结果是多个模态上评估结果的平均值。此外在 BRATS MRI 数据集上，我们添加了肿瘤分割标签作为病灶生成指导标签，评估时，我们根据病灶标签对合成数据集中影像与原数据集中影像的肿瘤切割出来进行评估，并与完整影像的评估进行比对，结果表明，我们的合成肿瘤信息与真实肿瘤具有极高的相似度，并在整体影像的评估中比整体平均情况更好，这说明我们的病灶生成指导方法是有效的，这对于合成影

像的应用也具有重大意义。表 3 中彩色视网膜数据集上，[41] 中所用的数据集的数据量为我们所用的数据集数据量的 22 倍多，因此，我们在我们的方法中采用了前述章节中的自监督预训练。我们以融合了噪声的结构特征图为输入、复现 SkrGAN 时采用我们的算法合成的结构特征图的反二值图（0 与 1 像素值调转）为输入，在采用同样的生成器和鉴别器、无额外的病灶损失和配准损失的情况下，从表中结果来看，我们的方法取得了比 SkrGAN 更好的评估结果，且质量远高于其他方法从随机噪声合成的图像。表 3 中天池肺部 CT 数据集上，我们采用肺炎分类标签作为病灶标签输入，与复现的 SkrGAN 仅是此损失和输入的结构特征图的处理方式不相同，因此两者的评估结果十分接近，得益于我们输入时融合的随机噪声带来的对模型泛化能力的提升，我们的指标略高于复现的 SkrGAN，从全部结果来看，采用结构特征图作为输入的方法远高于采用随机噪声未输入的方法。

综合表 2 和表 3 的结果来看，采用结构特征图为输入的合成图像的质量比采用随机噪声为输入的合成图像的质量要好很多。对结构特征图融合噪声处理比不处理或二值反转处理，模型的泛化能力更强。自监督预训练在小数据集上可以明显提升合成图像质量。相较于 SkrGAN 的草图和其他随机噪声输入，我们采用的结构特征图、自监督预训练、病灶损失等多种举措使得我们的合成医学影像质量更高，与真实图像更接近。

表 2：不同数据集上合成图像质量的量化评估（一）

Dataset	Metric	Ours	SkrGAN [41]	DCGAN [48]	ACGAN [62]	WGAN [63]	PGGAN [64]
Chest	MS-SSIM ↑	0.597	0.506	0.269	0.301	0.401	0.493
X-ray	FID ↓	102.5	114.6	260.3	235.2	300.7	124.2
Lung	MS-SSIM ↑	0.473	0.359	0.199	0.235	0.277	0.328
CT	FID ↓	66.91	79.97	285.0	222.5	349.1	91.89

表3: 不同数据集上合成图像质量的量化评估 (二)

Dataset	Metric	Ours ⁺	Ours	SkrGAN*	GAN#
Color	MS-SSIM ↑	-	0.607	0.584	0.392
Fundus	FID ↓	-	30.13	37.91	227.41
BRATS	MS-SSIM ↑	0.738	0.727	0.713	0.561
MRI	FID ↓	21.15	22.27	28.46	104.53
TC Lung	MS-SSIM ↑	-	0.676	0.667	0.592
CT	FID ↓	-	27.40	29.81	93.65

+ 表示合成肿瘤的评估结果，即根据病灶标签对合成数据集中影像与原数据集中影像的肿瘤切割出来进行评估的结果。

表示基础GAN的评估结果，即在DCGAN [48]的基础上，将生成网络加深为U-net，将鉴别器网络加深为VGG11。

* 表示复现的SkrGAN的评估结果。由于SkrGAN [41]中未给出草图详细算法或开源代码和鉴别器网络结构，我们使用采用我们的算法合成的结构特征图的反二值图（0与1像素值调转）作为输入，以U-net为基础生成模型、以我们采用的鉴别器VGG11为鉴别器合成多模态影像、无病灶损失、无配准损失、无自监督预训练，以此近似为对SkrGAN的复现。

5.5 合成病灶的有效性验证实验

5.5.1 脑肿瘤MRI合成

我们以肿瘤分割标签为病灶标签与结构特征图融合作为输入，合成4个模态的MRI时，每个模态采用一个训练好的肿瘤分割器作为病灶标签生成器，由肿瘤分割器提供分割结果与输入标签的自监督损失来确保病灶信息的合成。

5.5.2 对合成脑肿瘤MRI的肿瘤区域分割检验

如表 4所示，我们在BRATS2015训练数据集上训练了4个模态的肿瘤病变分割器，并在BRATS2015测试数据集上对其进行了测试。然后，我们使用训练有素的分割器对未过滤合成数据进行分割。从分割结果看，采用真实数据训练的分割器能在合成数据上表现良好，这说明合成数据与真实数据具有非常高的相似度，且合成的病灶与真实病灶的相似程度足够让分割器识别出合成病灶与非病灶部分。由此看出，我们的合成病灶是有效的，且能对肿瘤分割器产生实质影响，我们将进一步验证合成数据用于分割器训练的可行性。

表 4: 肿瘤分割器对不同测试数据的分割结果

testing dataset	MSE	Dice Score
real	0.026	0.915
synthetic	0.043	0.838

5.6 合成数据在智能医学影像处理任务中的可用性验证实验

5.6.1 合成脑肿瘤MRI数据用于肿瘤区域分割任务

如表 5所示，我们将不同量的BRATS 2015训练数据与真实BRATS合成数据混合，然后使用混合数据集进行多分割器方法进行分割训练，并在真实BRATS 2015上评估模型的分割能力测试数据集。所有实验均以相同的迭代次数进行了完全训练，该迭代次数等于BRATS2015训练数据集上的100个epoch。同时，我们设置了三种数据混合模

式：随机混合，首先进行实数据训练和首先进行综合数据训练。在实优先实验和综合优先实验中，来自不同来源的数据的训练迭代次数与数量成正比数据的。除表格中的条件外，其他条件相同。如表 5所示，实验NO.3-NO.5的结果表明，合成数据不能完全替代训练中的真实数据。NO.6-NO.8的结果表明，大量合成数据的预训练和少量真实数据的微调性能与完整真实数据的训练相似。在NO.9-NO.11中，不同混合比的结果也完全不同。当两个比率接近时，分割结果与NO.1相似。当合成数据所占比例较高时，合成数据将干扰实际数据的学习，结果低于NO.1。当合成数据所占比例较低时，可以通过合成数据提高模型的泛化能力，结果高于NO.1。在NO.12-NO.17中，我们进一步尝试将不同数量的合成数据添加到真实数据中，这表明添加少量合成数据可以增强学习效果，并且合成数据越多，增强效果越好，但是当综合数据达到一定百分比然后继续增加时，它会达到相反的效果。在NO.18-NO.23中，我们比较了合成数据和通过常规数据增强方法生成的增强数据的增强效果。我们发现两者在增强效果和数据量增加之间的趋势上相似但不相等。总的来说，增强模型在模型对增强数据量的敏感性方面更健壮，但增强效果较高。合成数据的限制远高于增强数据的限制。我们将NO.24-NO.25与NO.15进行了比较，发现合成数据用作预训练数据时性能最佳，而用作补充训练数据时性能较差。当用作增强数据与真实数据混合时，合成数据也可以实现某些增强。

通常，如果存在大量真实数据，则可以将少量合成数据用作增强数据，或者可以将大量合成数据用于预训练，然后对真实数据进行训练。如果真实数据较少，则可以使用大量的综合数据进行预训练，然后对少量真实数据进行微调，其结果可以与完整真实数据的结果相抗衡，因此得出的结论是与 [12]一致。我们不建议将合成数据完全用于训练，也不建议将合成数据用于补充训练。

5.6.2 合成眼底视网膜数据用于视网膜血管注释任务

我们使用DRIVE视网膜图像训练集+对FIRE视网膜图像全数据集作为训练数据合成了大量的视网膜图像。合成时，由于只有单模态数据且结构特征图与标签数据存在冲突，因此无病灶生成指导损失和模态配准损失。我们通过缩放掩膜可以去除结构特征图的外层圆形轮廓，只保留圆内的血管线路，我们以此血管线路图作为采用该结构特征图合成的视网膜图像的血管注释粗标签。我们以U-net为分割模型，使

表 5: 脑肿瘤MRI数据集上的合成数据可用性验证实验.

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

用DRIVE视网膜图像训练集+合成视网膜图像及其粗标签数据集，在DRIVE视网膜图像测试集上，采用随机混合模式训练视网膜血管注释模型，与只使用DRIVE视网膜图像训练集的分割结果对比，来验证合成眼底视网膜数据在视网膜血管注释任务中的可用性。由于我们采用的数据集的数据量仅有SkrGAN的二十分之一，因此尽管在结构特征图粗标签上我们的方法有一些提升，但我们合成的视网膜图像的合成质量和多样性相对来说依然较差。从对任务的提升效果来看，我们的合成图像对分割Accuracy的提升与SkrGAN基本一致，在Sensitivity和AUC指标上略低。但如表6中SkrGAN*所示，当我们采用相同的数据集进行近似复现后，我们的方法在各个指标上都更高。实验结果如表6所示。

表 6: 视网膜合成数据可用性验证实验

train data	test data	Sensitivity	Accuracy	AUC
DRIVE训练集	DRIVE测试集	0.7781	0.9477	0.9705
DRIVE训练集+2000张 SkrGAN合成图像	DRIVE测试集	0.8464	0.9513	0.9762
DRIVE训练集+2000张 SkrGAN*合成图像	DRIVE测试集	0.8297	0.9428	0.9732
DRIVE训练集+2000张 我们的合成图像	DRIVE测试集	0.8416	0.9518	0.9749

* 表示复现的SkrGAN的评估结果。由于SkrGAN [41]中未给出草图详细算法或开源代码和鉴别器网络结构，我们使用采用我们的算法合成的结构特征图的反二值图（0与1像素值调转）作为输入，以U-net为基础生成模型、以我们采用的鉴别器VGG11为鉴别器合成多模态影像、无病灶损失、无配准损失、无自监督预训练，以此近似为对SkrGAN的复现。

5.6.3 合成胸部X光线数据用于肺炎分类任务

我们采用Kaggle Chest X-Ray数据集合成了大量带有病变类别标签的胸部X-Ray数据。合成时，由于只有单模态数据，因此无模态配准损失，输入的病灶标签为扩展成的与结构特征图同尺寸独热矩阵的肺炎类别标签。然后再使用Kaggle Chest X-Ray数据集+合成X-Ray数据采用随机混合模式训练了一个胸部X-Ray肺炎分类模型，我们将

该模型和只采用Kaggle Chest X-Ray数据集训练的模型进行对比。结果如表 ??，可见在该数据集上，我们的合成数据同样可以在分类任务中具有良好的可用性，但相对于分割任务，由于合成时输入的分类标签的指向性较弱，合成图像所具有的类别属性不如分割类任务合成病灶那样集中和突出，因此分类任务使用合成数据的提升效果相对不明显。

表 7: 胸部X光线合成数据可用性验证实验

train data	test data	Accuracy
X-Ray训练集	X-Ray测试集	0.804
X-Ray训练集+2000张 我们的合成图像	X-Ray测试集	0.818

5.6.4 合成肺部CT数据用于肺结节检测任务

我们采用天池全球数据智能大赛(2019)数据集合成了大量带有病灶检测标签的肺部低剂量CT数据。合成时，由于只有单模态数据，因此无模态配准损失，输入的病灶标签为与结构特征图同尺寸的病灶检测标签。我们使用天池全球数据智能大赛(2019)数据集+合成肺部低剂量CT数据采用随机混合模式训练了一个病灶检测模型，将该模型的测试结果与只采用天池全球数据智能大赛(2019)数据集训练的检测模型进行对比。结果如表 8，这表明，在该数据集上，我们的合成数据同样具有良好的可用性。在检测任务中，输入的检测标签较分割标签的指向性弱，但比分类标签集中，从在对应任务中的提升效果也可以看出，合成数据在检测任务中的提升效果介于分类任务和分割任务之间。

5.7 合成数据效果展示

5.7.1 结构特征图多样性效果展示

我们从正态分布随机采样得到如图 13所示的一组结构特征图，图中展示了合成的结构特征图的多样性。

表 8: 肺部CT合成数据可用性验证实验

train data	test data	IOU
Lung CT训练集	Lung CT测试集	0.712
Lung CT训练集+20000张 我们的合成图像	Lung CT测试集	0.727

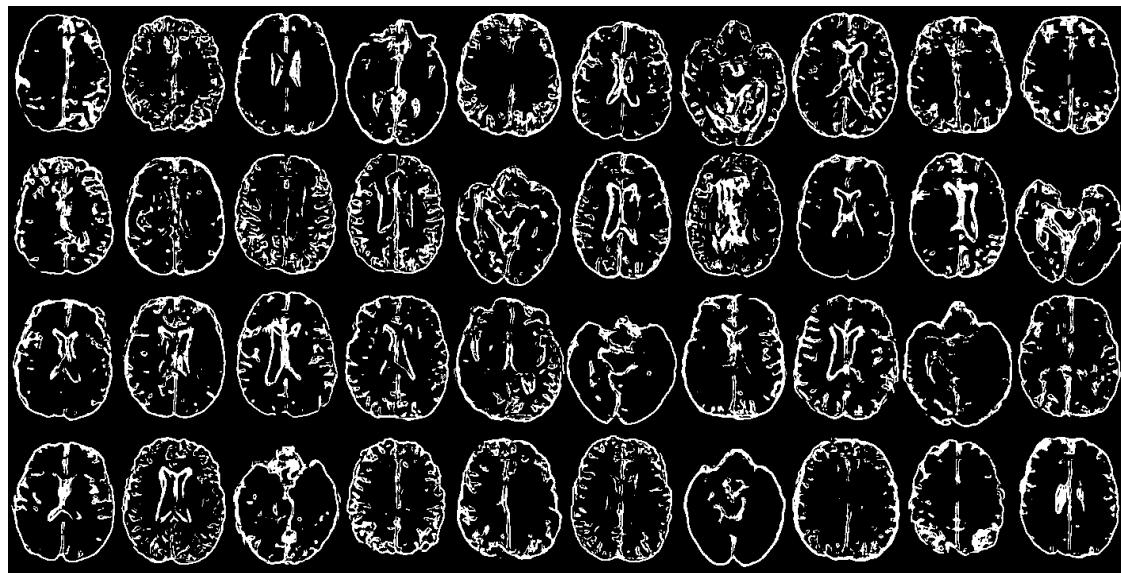


图 13: BRATS2015数据集上随机合成的结构特征图示例。

5.7.2 结构特征图正态分布效果展示



图 14: BRATS2015数据集上从正态分布顺序采样合成的结构特征图示例。

我们对从正态分布进行顺序采样，然后从采样结果解码，获得了如图 14所示的相邻渐变的结构特征图分布。图中任意一条直线上的结构特征图都呈现出渐变效果。

5.7.3 合成医学影像效果展示

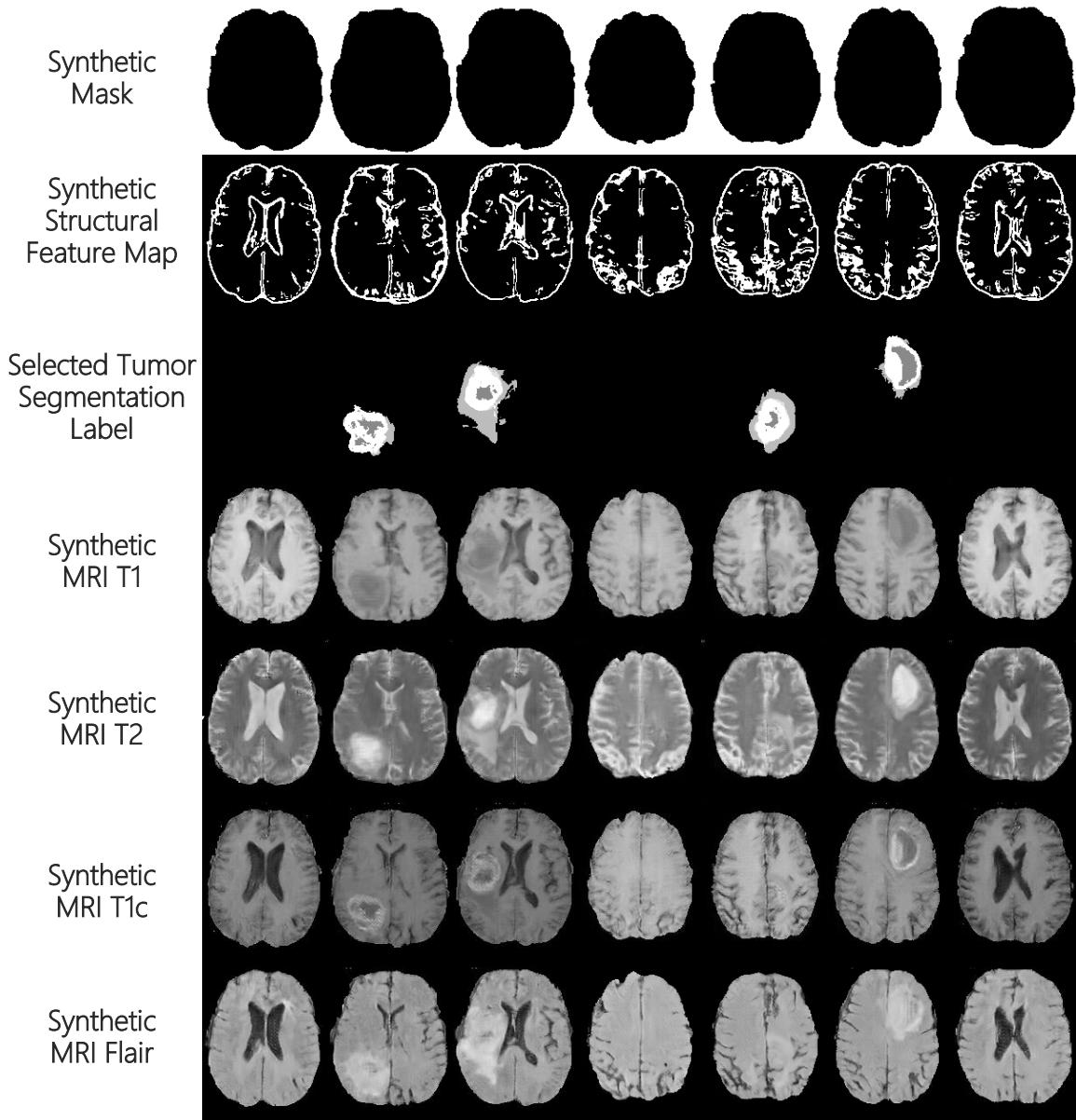


图 15: BRATS2015 上合成的结构特征图和多模态MRI

图 15 中展示了我们在 BRATS2015 数据集上合成的一组掩膜、结构特征图、选择的病灶标签和合成的四个模态的 MRI。

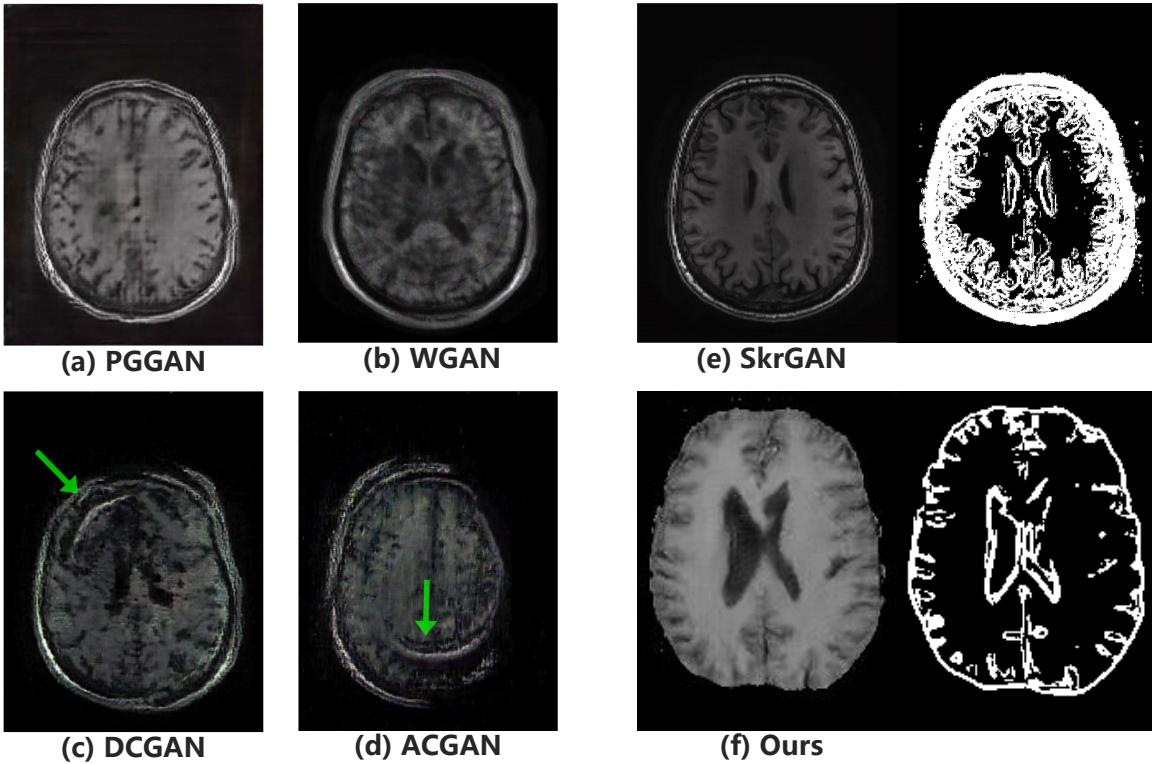


图 16: 脑部MRI合成效果对比。 (a) PGGAN [64] 的合成图像, (b) WGAN [63] 的合成图像, (c) DCGAN [48] 的合成图像, (d) ACGAN [62] 的合成图像, (e) SkrGAN [41] 的合成图像、输入的结构特征图的二值反转图像, (f) 我们的方案的合成图像和输入的结构特征图。 (a)-(e) 为在 ADNI 数据集上满足一些未知条件的筛选出的数据上进行训练的合成效果, (f) 在 BRATS2015 上合成训练。

前述表 3 中展示了我们的方法与其他方法在 BRATS2015 上的量化对比结果。图 16 中展示了 [41] 中其他方法和我们的方法合成的脑 MRI 的视觉效果, 由于 [41] 中采用的 ADNI 数据集上筛选条件的未知, 我们采用在 BRATS2015 上训练的合成脑 MRI 与之在合成视觉效果上简单比较。从图 16 中(e) 和(f) 中不难看出, 我们的结构特征图更加简洁清晰、复杂度低和限制少的同时根据合成指导性。从合成 MRI 来看, 我们的合成的脑 MRI 更加清晰真实、干净整洁, 服从输入的结构特征图的合成指导却有更多样的合成空间。

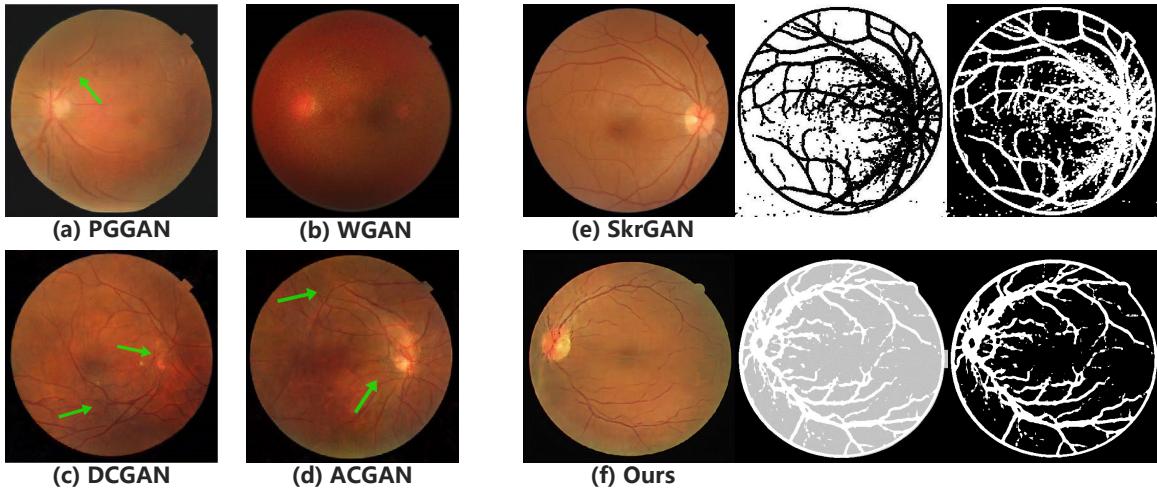


图 17: 视网膜合成效果对比。 (a) PGGAN [64] 的合成图像, (b) WGAN [63] 的合成图像, (c) DCGAN [48] 的合成图像, (d) ACGAN [62] 的合成图像, (e) SkrGAN [41] 的合成图像、输入的结构特征图和二值反转图像, (f) 我们的方案的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。

前述表 3 中展示了我们的方法与其他方法在 DRIVE+FIRE 数据集上的量化对比结果。此处, 图 17 中展示了我们在 DRIVE+FIRE 的 288 张眼底视网膜数据集上合成的结构特征图和视网膜影像, 和 [41] 中其他方法在自建的 6432 张视网膜数据集上的合成的视网膜图像对比, 我们采用更少的训练样本合成的视网膜与 SkrGAN 效果非常接近, 相比其他方法更加逼真, 在血管和神经交汇点等细节上更加合理和真实。其中, (e) 和 (f) 展示了我们的结构特征图和 SkrGAN 的草图的区别, 我们的结构特征图血管线条走向更加自然真实、噪声更少。

图 18 中展示了我们在 X-ray 数据集上合成的结构特征图和 X-ray 影像, 和其他方法合成的结果对比, 我们的合成 X-ray 更加逼真, 肋骨和脊柱等细节更加丰富和符合生理逻辑。其中, (e) 和 (f) 展示了我们的结构特征图和 SkrGAN 的草图的区别, 我们的结构特征图在脊柱等关键生理结构上提取出的信息更完整, 肋骨等核心部位的噪声更少、杂余线条更少、核心的结构信息更突出、线条与线条之间更加独立清晰无干扰, 我们的每条线条都更顺滑均匀, 对应合成的 X-ray 在结构上更加自然真实, 而 SkrGAN 由于其草图脊柱处细节的突兀, 合成 X-ray 在脊柱处也有明显瑕疵。

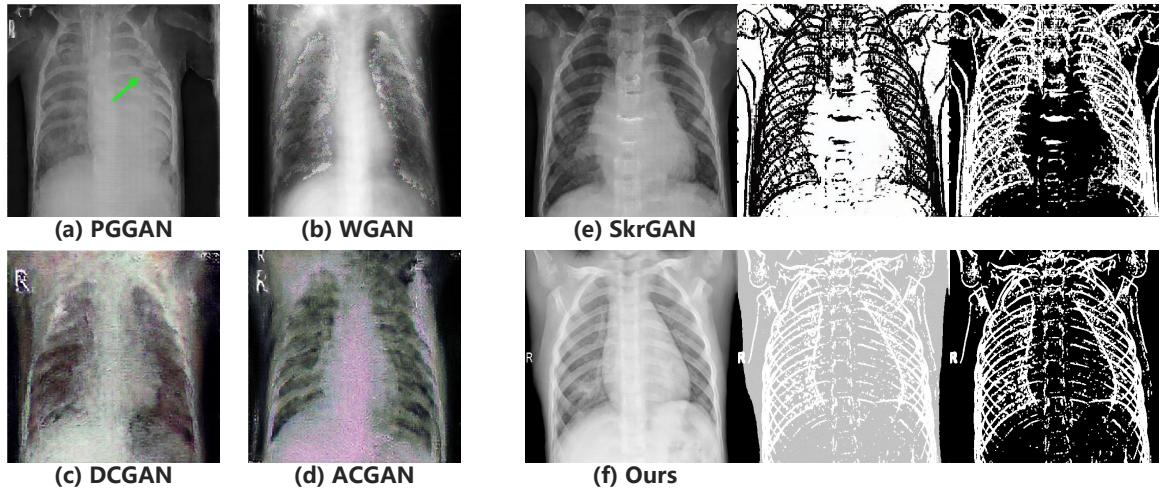


图 18: X-ray合成效果对比。 (a) PGGAN [64] 的合成图像, (b) WGAN [63] 的合成图像, (c) DCGAN [48] 的合成图像, (d) ACGAN [62] 的合成图像, (e) SkrGAN [41] 的合成图像、输入的结构特征图和二值反转图像, (f) 我们的方案的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。

图 19 中(f)展示了我们在肺部 CT 数据集上合成的结构特征图和 CT 影像, 和其他方法合成的结果对比, 我们的合成 CT 更加清晰真实。其中, (e) 和 (f) 展示了我们的结构特征图和 SkrGAN 的草图的区别, 我们提取出的结构特征图复杂度更低, 仅由关键结构的轮廓线条勾画而成, SkrGAN 的草图中包含了大面积的像素区域、黑白交错界限不明且充满噪声杂乱无章, 从合成的 CT 来看, SkrGAN 的合成的肺中缺乏生理结构内容, 而我们的合成的肺中具有符合生理结构的气管血管等结构。图 19 中(g) 展示了我们在天池肺部 CT 数据集上合成的结构特征图和 CT 影像, 效果非常逼真、肺中合成的结构信息丰富详实。

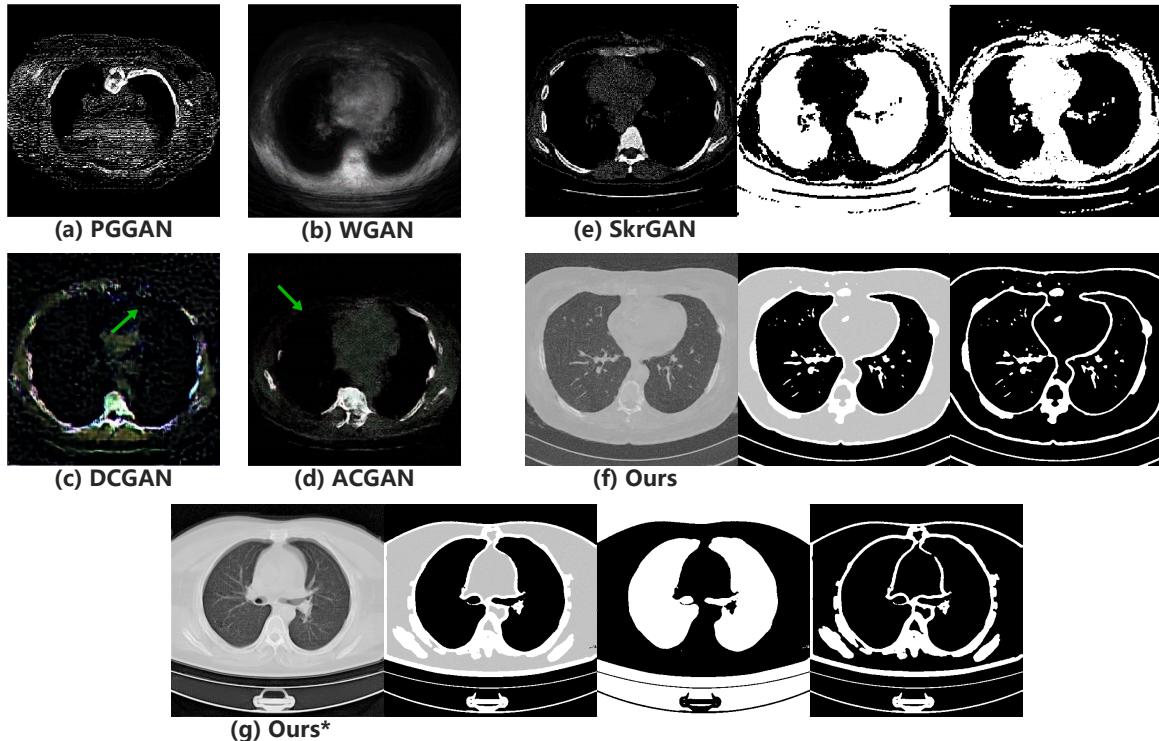


图 19: 肺部CT合成效果对比。 (a) PGGAN [64]的合成图像, (b) WGAN [63]的合成图像, (c) DCGAN [48]的合成图像, (d) ACGAN [62]的合成图像, (e) SkrGAN [41]的合成图像、输入的结构特征图和二值反转图像, (f) 在Kaggle Lung CT数据集上, 我们的方案的合成图像、输入的融合了噪声的结构特征图和原始结构特征图, (g) 在天池肺部CT数据集上, 我们的方案的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。(f)-(g)输出的合成图像像素值在0-1之间, 视觉上与(a)-(e)中非归一化的合成结果有一些差异。

第六章 结语

总的来说，我们提出了一种从正态分布随机噪声分阶段合成可带病灶标签的配准的多模态医学影像的方法，并在多个数据集上进行了充分的实验验证和效果展示。首先，我们提出了一种结构特征图提取方法，无需训练或附加标签数据即可直接从医学图像中提取解剖结构信息，相较于当前最好的草图提取方法，我们提取的结构特征图更加干净简洁、线条清晰、完整合理。在此基础上，我们提出一种结合VAE和GAN的结构特征图生成方法，可以从多维正态分布随机采样生成结构特征图。然后，我们对结构特征图进行噪声融合处理和可控的病灶标签添加处理，再通过无监督训练实现了从结构特征图合成符合生理结构的多模态医学图像，并可以通过添加病灶标签合成对于病灶信息和实现多模态影像的精确配准。最后，我们实现了对合成医学影像用作智能医学图像处理任务的预训练数据或增强数据的可用性验证，多项实验结果表明我们合成的数据可以显着提高模型的泛化能力，尤其是在图像分割类任务中。在本文中，我们的贡献如下：

- 我们提出一种结构特征图提取方法，以直接从医学图像中提取解剖结构信息，而无需训练或附加标签数据。
- 我们结合VAE和GAN的特点提出了一种结构特征图生成方法，以从多维正态分布矩阵生成结构特征图。
- 我们实现了从结构特征图和随机选择的病变标签中合成具有相应病变信息的配准多模态医学图像。
- 我们通过多项合成数据可用性实验，验证合成数据可以用作多种智能医学图像处理任务的预训练数据或增强数据。

参考文献

- [1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *neural information processing systems*, 141(5):1097–1105, 2012.
- [2] Ian J Goodfellow, Jean Pougetabadi, 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] Geert J S Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A W M Van Der Laak, Bram Van Ginneken, and Clara I Sanchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.
- [4] Junegoo Lee, Sanghoon Jun, Youngwon Cho, H S Lee, Guk Bae Kim, Joon Beom Seo, and Namkug Kim. Deep learning in medical imaging: General overview. *Korean Journal of Radiology*, 18(4):570–584, 2017.
- [5] Dinggang Shen, Guorong Wu, and Heung Il Suk. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19(1):221–248, 2017.
- [6] 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.
- [7] Mingyu Liu, Thomas M Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *neural information processing systems*, pages 700–708, 2017.
- [8] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. *international conference on machine learning*, pages 1857–1865, 2017.

- [9] Yizhe Zhu, Mohamed Elhoseiny, Bingchen Liu, and Ahmed M Elgammal. Imagine it for me: Generative adversarial approach for zero-shot learning from noisy texts. *arXiv: Computer Vision and Pattern Recognition*, 2017.
- [10] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. *computer vision and pattern recognition*, pages 3194–3203, 2018.
- [11] Yunye Gong, Srikrishna Karanam, Ziyan Wu, Kuanchuan Peng, Jan Ernst, and Peter C Doerschuk. Learning compositional visual concepts with mutual consistency. *computer vision and pattern recognition*, pages 8659–8668, 2018.
- [12] 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.
- [13] Yuankai Huo, Zhoubing Xu, Shunxing Bao, Albert Assad, Richard G Abramson, and Bennett A Landman. Adversarial synthesis learning enables segmentation without target modality ground truth. *international symposium on biomedical imaging*, pages 1217–1220, 2018.
- [14] Juan Eugenio Iglesias, Ender Konukoglu, Darko Zikic, Ben Glocker, Koen Van Leemput, and Bruce Fischl. Is synthesizing mri contrast useful for inter-modality analysis? *medical image computing and computer assisted intervention*, pages 631–638, 2013.
- [15] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. Learning from simulated and unsupervised images through adversarial training. *computer vision and pattern recognition*, pages 2242–2251, 2017.
- [16] Bo Zhao, Bo Chang, Zequn Jie, and Leonid Sigal. Modular generative adversarial networks. *european conference on computer vision*, pages 157–173, 2018.

- [17] 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.
- [18] 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.
- [19] Guim Perarnau, Joost Van De Weijer, Bogdan Raducanu, and Jose M Alvarez. Invertible conditional gans for image editing. *arXiv: Computer Vision and Pattern Recognition*, 2016.
- [20] Asha Anoosheh, Eirikur Agustsson, Radu Timofte, and Luc Van Gool. Combogan: Unrestrained scalability for image domain translation. *computer vision and pattern recognition*, pages 783–790, 2018.
- [21] Amelie Royer, Konstantinos Bousmalis, Stephan Gouws, Fred Bertsch, Inbar Mosseri, Forrester Cole, and Kevin P Murphy. Xgan: Unsupervised image-to-image translation for many-to-many mappings. *arXiv: Computer Vision and Pattern Recognition*, 2018.
- [22] 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.
- [23] 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.
- [24] 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.

- [25] 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.
- [26] 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.
- [27] Raviteja Vemulapalli, Hien Van Nguyen, and Shaohua Kevin Zhou. Unsupervised cross-modal synthesis of subject-specific scans. *international conference on computer vision*, pages 630–638, 2015.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] Thomas Joyce, Agisilaos Chartsias, and Sotirios A Tsaftaris. Robust multi-modal mr image synthesis. *medical image computing and computer assisted intervention*, pages 347–355, 2017.

- [32] 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.
- [33] Zhiwen Fan, Liyan Sun, Xinghao Ding, Yue Huang, Congbo Cai, and John Paisley. A segmentation-aware deep fusion network for compressed sensing mri. *european conference on computer vision*, pages 55–70, 2018.
- [34] Rushil Anirudh, Hyojin Kim, Jayaraman J Thiagarajan, K Aditya Mohan, Kyle M Champliey, and Timo Bremer. Lose the views: Limited angle ct reconstruction via implicit sinogram completion. *computer vision and pattern recognition*, pages 6343–6352, 2018.
- [35] Chenyu You, Guang Li, Yi Zhang, Xiaoliu Zhang, Hongming Shan, Shenghong Ju, Zhen Zhao, Zhuiyang Zhang, Wenxiang Cong, and Michael W. Vannier. Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble(gan-circle). *arXiv preprint arXiv:1808.04256*, 2018.
- [36] Qing Lyu, Chenyu You, Hongming Shan, and Ge Wang. Super-resolution mri through deep learning. *arXiv preprint arXiv:1810.06776*, 2018.
- [37] 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.
- [38] 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.
- [39] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *international conference on learning representations*, 2014.

- [40] 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.
- [41] Tianyang Zhang, Huazhu Fu, Yitian Zhao, Jun Cheng, Mengjie Guo, Zaiwang Gu, Bing Yang, Yuting Xiao, Shenghua Gao, and Jiang Liu. Skrgan: Sketching-rendering unconditional generative adversarial networks for medical image synthesis. *arXiv: Computer Vision and Pattern Recognition*, 2019.
- [42] 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*, page 33, 2014.
- [43] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of The ACM*, 60(6):84–90, 2017.
- [44] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv: Computer Vision and Pattern Recognition*, 2014.

- [45] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *Computer Vision and Pattern Recognition*, 2015.
- [46] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Computer Vision and Pattern Recognition*, 2016.
- [47] Mathew Salvaris, Danielle Dean, and Wee Hyong Tok. Generative adversarial networks. *arXiv: Machine Learning*, pages 187–208, 2018.
- [48] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv: Learning*, 2015.
- [49] Guillaume Alain, Yoshua Bengio, Li Yao, Jason Yosinski, Eric Thibodeau-laufer, Saizheng Zhang, and Pascal Vincent. Gsns : Generative stochastic networks. *arXiv: Learning*, 2015.
- [50] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv: Machine Learning*, 2013.
- [51] Carl Doersch. Tutorial on variational autoencoders. *arXiv: Machine Learning*, 2016.
- [52] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *arXiv: Computer Vision and Pattern Recognition*, 2014.
- [53] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *medical image computing and computer assisted intervention*, pages 234–241, 2015.
- [54] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. pages 580–587, 2014.
- [55] Ross Girshick. Fast r-cnn. *international conference on computer vision*.

- [56] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017.
- [57] Joseph Redmon, Santosh K Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *arXiv: Computer Vision and Pattern Recognition*, 2015.
- [58] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Chengyang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. *european conference on computer vision*, pages 21–37, 2016.
- [59] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Chengyang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. *european conference on computer vision*, pages 21–37, 2016.
- [60] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [61] 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.
- [62] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. *arXiv: Machine Learning*, 2016.
- [63] Martin Arjovsky, Soumith Chintala, and Leon Bottou. Wasserstein generative adversarial networks. pages 214–223, 2017.
- [64] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv: Neural and Evolutionary Computing*, 2017.

致谢

感谢卢宇彤导师、陈志广老师的悉心指导，感谢国家超级计算广州中心的支持，
感谢郑馥丹师姐、苏琬棋师妹、邓楚富师弟等的无私帮助，感谢家人的理解和陪伴！

瞿毅力

二零二零年四月