

中山大学硕士学位论文

基于生成对抗网络的多模态医学影像合成研究

Research on Multimodal Medical Images Synthesis
Based on GAN

学位申请人: 瞿毅力

指导老师: 王青副教授 (卢宇彤教授)

专业名称: 软件工程

答辩委员会 (签名):

主席:

委员:

二零二零年五月三十一日

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

基于生成对抗网络的多模态医学影像合成研究

专业：软件工程

硕士生：瞿毅力

指导老师：王青 副教授（卢宇彤 教授）

摘要

医学影像数据的采集和标注一直是医学影像处理任务面临的挑战，尤其是基于配准多模态数据的应用。利用图像合成技术可以有效缓解这一问题。但医学影像包含复杂的生理结构信息，现有方法直接合成医学影像会生成不合理的结构、轮廓和不可控的病灶；合成多种模态的医学影像时如何确保模态之间的配准面临严峻的挑战；如何控制医学影像中最关键的病灶信息的合成也是一大难题；另外，合成影像和合成病灶需要一种更全面的方式来评估其性能。

针对这些问题，本研究提出了一种基于生成对抗网络的多模态医学影像合成方案，可从随机噪声合成具有指定病灶的配准多模态医学影像。多个数据集上的实验验证了合成影像的可用性、病灶的有效性和模态间的配准度。本研究的主要工作包括：

1. 本研究提出了一种基于Sobel算子的结构特征图提取和生成方法，无需额外的结构信息标签或标签提取训练，可直接从真实影像提取出结构特征图并用于变分自编码器的训练，最后得到的模型可从正态分布随机采样合成任意数量的结构特征图。

2. 本研究提出了一种可控制病灶合成的多模态配准医学影像合成方案，可将合成的结构特征图与选定的病灶标签融合后再合成多模态医学影像，通过模态转换器提供合成多模态之间的转换一致性约束以确保合成多模态之间的配准，通过病灶处理器实现对重建合成影像病灶标签的约束从而确保合成影像根据输入标签生成对应病灶。

3. 本研究在采用一般相似性评估对合成和真实数据直接评估的基础上，进一步使用合成影像训练医学影像处理模型，通过评估训练好的模型间接评估合成影像的性能。结果表明合成影像可在多种任务中通过用于预训练和数据增强来提升模型性能。

关键词：医学影像，图像合成，多模态配准，病灶，生成对抗网络

Research on Multimodal Medical Images Synthesis Based on GAN

Major: Software Engineering

Name: Yili Qu

Supervisor: Associate Prof. Qing Wang (Prof. Yutong Lu)

Abstract

The collection and annotation of medical images data have always been a challenge in many data-driven medical images processing tasks, especially for those based on registered multimodal medical image data. This problem can be largely alleviated by utilizing the image synthesis technology. However, medical images contain complex physiological structure information, and the directly-synthesized medical images generated by current methods usually have unreasonable structures or contours and uncontrollable lesions. The registration of different modalities when synthesizing multimodal medical images remains difficult. How to control the synthesis of the most critical lesion information in medical images is also a major problem. In addition, a more comprehensive way is needed to evaluate the performance of synthetic images and lesions.

To solve these problems, this study proposed a multimodal medical image synthesis method based on Generative Adversarial Networks, which can synthesize registration multimodal medical images with specified lesions. Experiments on multiple datasets have verified the availability of the synthesized images, the effectiveness of the synthesized lesions and the registration of the synthesized multimodal. The main contributions are as follows:

1. This study proposed a structural map extraction and synthesis method based on the Sobel operator, which does not require additional structural information labels or label extraction training. The method can extract structural map directly from the real images and then feed it to Variational Auto-Encoder for training. Finally, the trained model can achieve any number of structural maps synthesis from the random normal distribution matrixes.

Abstract

2.This study proposed a multimodal registration medical images synthesis scheme that can control lesion synthesis. The randomly generated structural maps are used to fuse with the selected lesion labels and then synthesize multimodal medical images. The modal translator constrains the consistency of the translations between the multimodal synthesis and ensures the registration between the multimodal synthesis. The lesion processor constrains the restoration of the lesion label in the synthetic image to ensure that the synthetic image generates a corresponding lesion according to the input label.

3.This study directly evaluated the similarity of the synthesized data and the real data by the general evaluation method in the first step. Next, the synthesized images were further used to train the intelligent medical image processing model, and the performance of the synthesized images were indirectly evaluated by evaluating the trained model. The results showed that the synthesized image can be used for pre-training and data augmentation in a variety of tasks to improve model performance.

Keywords: Medical Images, Image Synthesis, Multimodal Registration, Lesions, Generative Adversarial Networks

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 研究内容和技术路线	4
1.3 本文的创新点	7
1.4 本文的结构	8
第二章 国内外研究现状	9
2.1 图像合成的研究现状	9
2.2 医学影像合成的研究现状	10
2.3 小结	12
第三章 智能医学影像处理的研究基础	13
3.1 基于卷积神经网络的图像分类	13
3.2 基于生成对抗网络的图像合成	15
3.3 基于变分自编码器的图像合成	16
3.4 基于卷积神经网络的语义分割	17
3.5 基于卷积神经网络的目标检测	18
第四章 医学影像结构特征图的提取和生成方法	21
4.1 结构特征图提取方法	22
4.2 结构特征图的病灶结构移除	25
4.3 结构特征图生成训练	26
4.4 结构特征图的过滤	30
4.5 结构特征图与随机噪声的融合	31

第五章 基于结构特征图的多模态医学影像合成方法	33
5.1 整体架构	33
5.2 多模态影像的合成	34
5.3 无监督多模态影像的配准	38
5.4 病灶信息的添加和合成	39
5.5 合成数据集的构建	42
第六章 合成医学影像的性能评估	45
6.1 数据集	45
6.2 实验设置	47
6.3 多模态医学影像合成方法的消融实验	50
6.4 合成医学影像数据与真实医学影像数据的相似性评估	52
6.5 合成医学影像的病灶有效性评估	54
6.6 合成多模态医学影像的配准有效性评估	55
6.7 合成医学影像在智能医学影像病灶处理任务中的可用性评估	57
6.8 合成影像的视觉评估	62
第七章 结语	71
参考文献	73
附录一 攻读硕士学位期间取得的科研成果	87
致谢	89

第一章 绪论

1.1 研究背景和目的

近年来，随着深度学习（Deep Learning）的提出和发展，深度学习表现出的极强的学习能力吸引了许多的研究者投入其中。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（Medical Images）的相关研究也越来越多，智能医学影像处理成为了深度学习商业化应用最为成熟和影响最为广泛的领域之一。在原理上，深度学习本身依赖于数据驱动，这使得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获取的自然图像来说，医学影像数据的采集和标注（Annotate）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的风险，需要考虑病人隐私保护与数据脱敏，需要医院多个部门的协调和配备合格的采集设备，需要病人及家属的配合，需要常年累月的数据积累和有经验的专业医生的标注，有时还需要有充足的病例特别是罕见病病例。在构建数据集需要极大代价且牵涉多方利益的情况下，许多机构很难具备公开自己的数据的条件。这使得相对于当前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（GAN, Generative Adversarial Networks）的强大生成能力在一系列研究中逐渐展现。许多基于深度学习的计算机视觉研究使用生成对抗网络获得大量生成样本来进行数据增强（Data Augmentation）以获得模型性能的提升 [1–4]，或在小数据集上通过补充生成样本来提高应用深度学习方法的可行性 [5]。同样地，面对上述医学影像数据采集和标注困难、公开数据集稀少且数据量小的现状，利用合成的医学影像数据来缓解样本不足的问题成为了一种可行方案。然而，医学影像具有一定的特殊性，其中通常包含有复杂精细的生理结构信息。这意味着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符合生理逻辑的结构或轮廓。这是合成医学影像首先要面对的挑战。

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模型，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，简单地将多个单模态合成模型训练多次得到的合成多模态影像，很难实现模态之间的相互配准。这是多模态医学影像合成的另一个挑战。

再者，对于医学影像，其最有价值的地方在于其中包含的病灶（Lesion）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很难确保有效地合成，即使影像中合成了病灶也难以提供对应的病灶标签，难以对合成的医学影像在医疗诊断上的可用性进行评估和检验，更很难单独地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研究中未能很好解决的一个问题。

最后，当前采用一般相似性评估方法对合成医学影像数据和真实数据的相似性评估或采用主观打分方式对合成医学影像质量的评估难以验证合成数据的可用性、合成病灶的有效性和合成多模态的配准情况。除了采用一般相似性量化评估方法外，更全面的评估还需要一种更直观有效地验证合成影像各项性能的评估方法。

在上述的问题和挑战中，涉及到的技术背景的核心概念包括深度学习、医学影像及其模态、多模态医学影像的配准、病灶与标注、图像合成等，下面是对这几项概念的详细解释和介绍：

- **深度学习** 深度学习是机器学习（Machine Learning）领域中一个新发展起来的研究方向，它展显现出的能力使得机器学习更加接近人工智能（Artificial Intelligence）这个终极目标。在实现上，深度学习一般指基于深层人工神经网络模型、采用随机梯度下降等最优化方法对给定样本数据的内在规律和表示层次进行训练学习和推理的过程。深度学习中常见的神经网络模型包括卷积神经网络（CNN, Convolutional Neural Networks）模型 [6]、循环神经网络（RNN, Recurrent Neural Network）模型 [7]、生成对抗网络（GAN）模型 [8]等。深度学习可以处理图像、文本、语音等多种形式的信息，在搜索推荐、机器翻译、

语音合成、人像识别、图像识别等诸多领域应用广泛。深度学习同样广泛应用于多种医学影像处理任务，如医学影像模态转换、医学图像分割、病灶检测、医学图像分类等。深度学习在医学影像处理的相关研究中具有十分广阔前景 [9–11]。

- **医学影像及其模态** 医学影像是指对个体通过特定的成像方式取得的用于医疗诊断或医学研究的影像，其中不同的成像方式得到的医学影像被称为不同的模态（Modality），常见的医学影像模态有核磁共振成像（MRI, Magnetic Resonance Imaging）、计算机断层扫描（CT, Computed Tomography）成像、正电子发射型计算机断层（PET, Positron Emission Computed Tomography）成像、超声波成像（B超）、X射线（X-ray）成像等。有的模态在成像时设置不同的参数将得到具有视觉差异的不同的子模态，例如CT根据辐射剂量的不同分为低剂量CT（Low-Dose CT）和高剂量CT（Normal -Dose CT）、MRI根据成像时长和造影剂及射频方向等不同分为T1、T2、T1c、Flair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。在采用CNN或GAN等深度学习方法进行的智能医学影像处理模型的训练和学习中，研究者们也期望获得更多模态的影像来提升模型的诊断能力。
- **多模态医学影像的配准** 同一个病人的同一个部位通过不同的成像技术或参数得到的一组不同的模态被称为多模态（Multimodal），如果成像位置和视角是一致的，那么得到的多模态影像就是对齐的，这些模态之间的这种对齐关系称之为配准（Registration）。相较于未配准的多模态医学影像，配准后的多模态医学影像，对于医生来说可以更直观地对比关键部位在不同模态中的状态来加以诊断，对于智能医学影像处理任务来说则可以满足一些处理复杂任务的深度学习模型对多模态训练数据的要求以提供更加高效可靠的智能诊断服务。在采集时，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，同时还需要额外的伴随失真的配准计算过程。
- **病灶与标注** 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，病灶组织与正常组织、病灶的不同种类、病灶的严重程度等相互

之间不一定具有很明显的视觉差异，往往需要极具经验的专业医生进行判断。在智能医学影像处理任务中，病灶信息同样是模型进行学习和推理的重要信息，因此，医学影像数据集需要具有丰富经验的专业医生对医学影像中的病灶进行标注得到可用于指导智能医学影像处理模型训练的标签（Label），这是十分巨大的工作量。由于职业的特殊性，医生们的本职工作常常十分繁重，这样大量的标注任务往往需要医生长期牺牲自己的空余时间来逐步完成。

- **图像合成** 图像合成（Image Synthesis）或图像生成是指从一个随机噪声矩阵或一个具有一定指导信息的矩阵合成一张预期的完整目标图像的过程，其中，从带有指导信息的矩阵合成图像进一步发展为像素到像素映射的图像转换。图像转换是图像合成的一种。从数学本质上，图像合成可以视为一个数据分布到另一个数据分布的变换。随着生成对抗网络（GAN）[8]的提出，从随机噪声合成图像的相关研究发展迅速。随后包括风格迁移 [12–15]、人像转换 [16–19] 等诸多图像转换应用被提出。其中，医学影像的合成指以医学影像为预期输出目标的图像合成，医学影像的模态转换指以一种医学影像模态为输入合成另一种医学影像模态的像素到像素的图像转换过程。

总的来说，当前合成医学影像的研究中还存在合成模态数量偏少、合成多模态难以配准、合成病灶难以控制、合成数量存在限制、合成生理结构不够合理、合成质量评估不够全面等各项未能很好解决的问题。本研究的目的正是针对上述的这些问题，基于当前的图像合成技术和相关研究成果，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，实现可控制病灶合成并带有病灶标签的配准的多模态医学影像的合成。

1.2 研究内容和技术路线

针对医学影像包含的复杂的生理结构信息难以合成的问题，本研究采用渐进式合成思路，分成两个阶段实现医学影像的合成，第一个阶段是从随机噪声合成目标部位的包含结构信息的简单线条轮廓二值图，本文中将这样的图称为结构特征图（Structural Map），然后以结构特征图为输入进一步合成多模态医学影像。

针对多模态医学影像的合成和配准问题，本研究先采用一个基于ACGAN [20]结构的条件生成器进行多模态影像的合成，然后采用一个基于CycleGAN [14]结构的条件转换器进行合成的多模态影像之间的互相转换，通过模态之间像素到像素的转换一致性损失实现对合成多模态影像的配准约束。

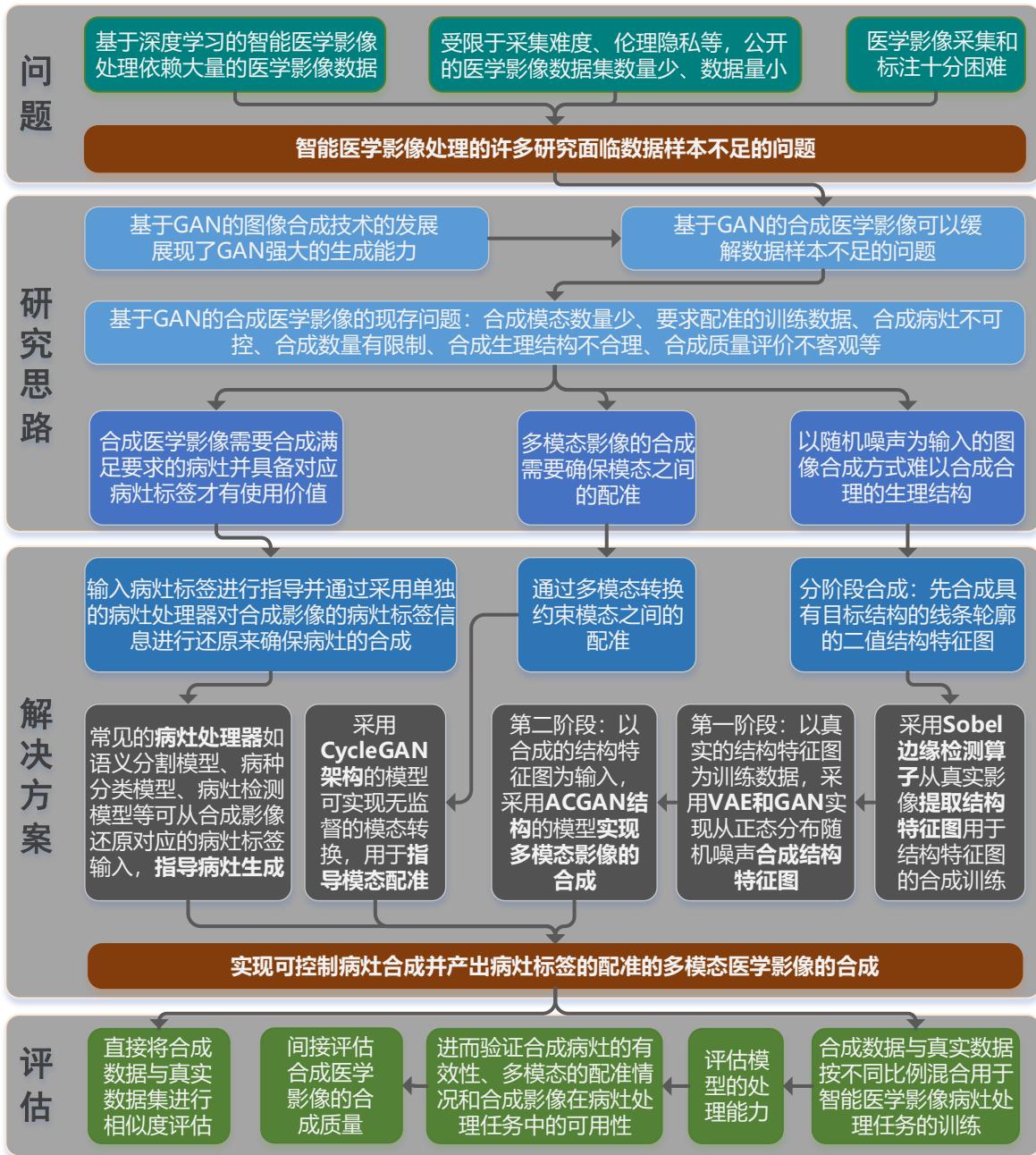


图 1.1: 研究内容。

针对合成病灶的问题，本研究首先在输入结构特征图时额外输入一张病灶标签图用以提供合成病灶的指导信息，然后采用一个病灶处理器，即一个独立的智能医学影像处理网络，来对合成的医学影像进行诊断推理提取出合成影像中的病灶信息，通过输入的病灶信息与提取出的病灶信息的自监督一致性损失约束生成器根据输入的标签合成指定的病灶内容。

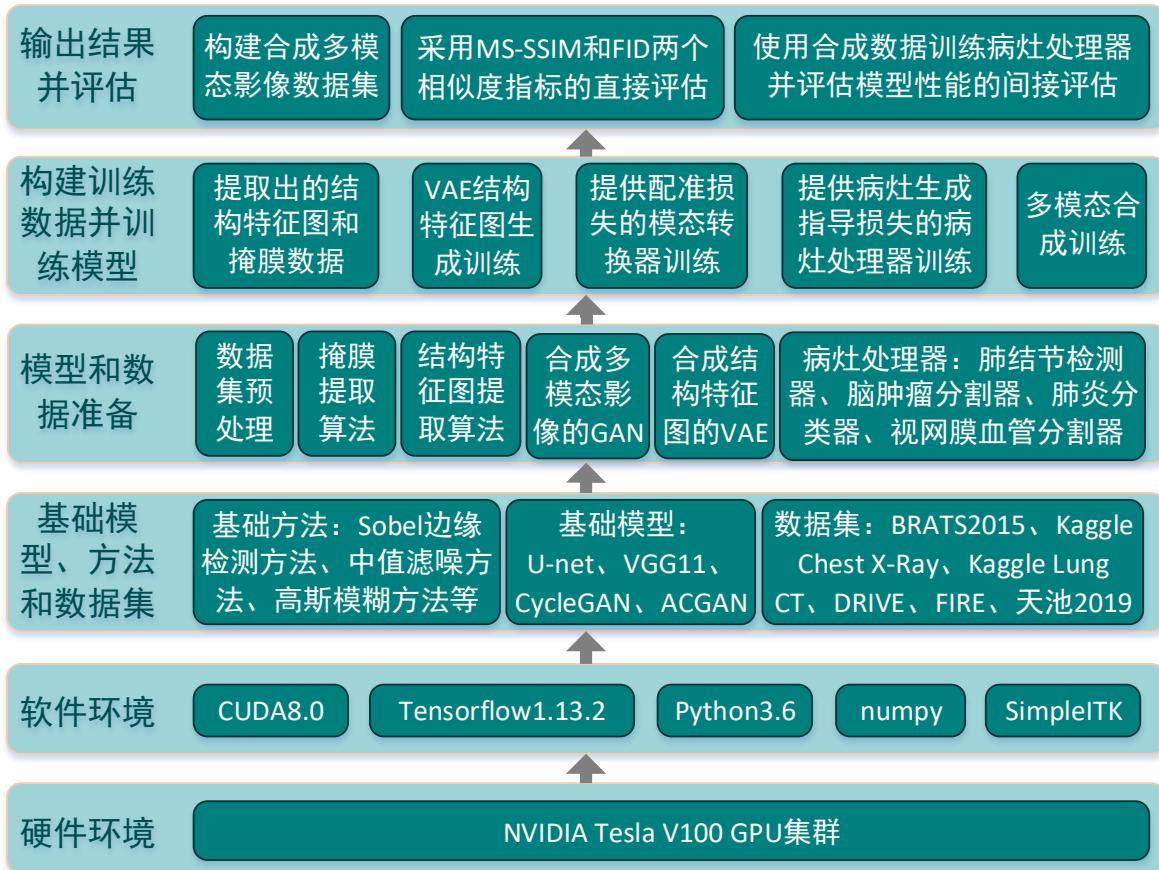


图 1.2: 技术路线。

对合成的多模态医学影像，本研究除了采用一般相似性评估方法将其与真实影像进行直接相似性评估外，还将其与真实影像按照不同比例进行混合用于训练智能医学影像处理模型，然后根据训练好的模型的评估结果来验证合成病灶的有效性和合成影像在智能医学影像处理任务中的可用性，并以此作为一种对合成医学影像质量的评估方式。

本研究详细的研究内容如图1.1所示，图中简要直观地展示了上述的研究思路和

研究过程。本研究的技术路线图如图1.2所示，图中直观地展示了在研究过程中采用的技术、模型及简易的调用路线。

1.3 本文的创新点

针对本研究面临的各项挑战，本文提出了一种基于GAN的多模态医学影像合成方案，可从随机矩阵合成具有指定病灶的配准多模态医学影像，并在多个数据集上验证了合成方法在多种不同成像模态、成像部位和病灶种类的医学影像合成中鲁棒性、合成医学影像的病灶有效性、合成多模态影像的配准有效性和合成影像在多种病灶处理任务中的可用性。合成的医学影像可用于多种智能医学影像处理任务的预训练和数据增强并能在任务中显著提升模型的泛化能力，以此起到缓解智能医学影像处理任务中数据不足问题的作用。本文的创新点包括：

- 本文提出了一种基于Sobel算子（Sobel Operator）[21]和变分自编码器的结构特征图提取和合成方法。提取结构特征图时，无需额外的结构信息标签和标签提取训练，可直接从真实影像提取出结构特征图。相较于当前最好的结构信息提取方法，本文方法提取的结构特征图在更具核心结构描绘能力的同时轮廓更加完整合理、线条更加清晰简明。生成结构特征图时，本文方法可直接从随机正态分布矩阵合成任意数量的、更具有多样性的结构特征图。
- 本文提出了一种可控制病灶合成的配准多模态医学影像合成方法，输入结构特征图和选取的病灶标签，通过执行病灶处理器还原合成影像中病灶标签的操作来约束合成影像生成器根据输入的病灶标签生成对应的病灶内容，通过模态转换器提供的合成多模态影像之间的转换一致性损失约束来确保合成多模态之间的配准。
- 本文在采用一般相似性评估方法对合成影像数据集与真实影像数据集进行直接评估的基础上，进一步提出了一种合成医学影像可用性、合成病灶有效性和合成多模态配准情况的评估方法，通过将合成影像用于智能医学影像处理任务的训练，再评估训练出的模型，间接地评估了合成影像的各项性能。

1.4 本文的结构

本文的正文有七个章节，每章的内容安排如下：

第一章：绪论。介绍本文研究的问题背景、研究目的、研究思路、技术路线和本文的创新点。

第二章：国内外研究现状。介绍与本研究相关的国内外的研究现状，并进行简单地分析与总结

第三章：智能医学影像的研究基础。介绍本文后续章节在方法和实验中将会运用的智能医学影像研究中使用的基础方法，包括基于卷积神经网络的图像分类、基于生成对抗网络的图像合成、基于变分自编码器的图像合成、基于卷积神经网络的语义分割和基于卷积神经网络的目标检测。

第四章：结构特征图的提取和生成方法。本章是本文第一个关键阶段方法的完整描述，首先对结构特征图的提取过程进行介绍，再对结构特征图的生成训练进行了介绍，穿插了对结构特征图的前处理和后处理方法的介绍。

第五章：配准多模态医学影像与病灶的合成方法。本章是本文第二个关键阶段方法的完整描述，首先对本文方法整体架构进行了介绍，再依次对多模态影像的合成、多模态影像的配准、病灶信息的添加和合成方法、合成数据集的构建过程进行介绍。

第六章：合成医学影像的性能评估。本章介绍对合成的医学影像的评估方法和实验结果，包括数据集介绍、评估指标介绍、消融实验、合成医学影像整体质量的直接量化评估、合成医学影像的病灶有效性评估、合成多模态影像的配准有效性评估、合成医学影像在智能医学影像病灶处理任务中的可用性评估。

第七章：结语。包括对本文研究成果的总结，对本文不足之处的分析，和对进一步研究方向的展望。

第二章 国内外研究现状

上一章介绍了本研究的问题背景、研究意义和研究思路，本章继续对与本研究相关的国内外研究现状进行细致分析。

在深度学习的发展过程中，国内外发表了许多与本研究相关的研究成果。首先，许多在一般自然图像或人像上进行的图像合成的相关研究使得图像合成技术取得了充分的发展，尤其是基于GAN的图像合成技术，在许多研究中展现出了强大的合成能力和巨大的潜能。其次，在与本研究更相关的医学影像合成领域，广阔的应用前景吸引着研究者们参与研究，并在近两年也出现了一些优秀的研究成果。本研究将基于大量优秀研究为研究基础，从国内外的研究现状详细分析探寻当前仍存在的问题，深入研究和充分实验，再提出新的解决方案。

2.1 图像合成的研究现状

前述章节对图像合成的概念进行了介绍和解释，本节对其研究现状进行更详细地介绍。在最近的图像合成研究中，CNN和GAN成为了应用最广泛的技术。语义分割模型全卷积神经网络（FCN）[22]是采用CNN输出整张图像的最早的研究之一，之后在语义分割领域，优秀的CNN模型不断涌现，从最初经典的 SegNet [23]、U-Net [24]、DeepLab系列 [25–28]、FCDenseNet [29]、Mask R-CNN [30]、PSPNet [31]、RefineNet [32]等模型再到最近性能卓越的 DenseASPP [33]、ResNet DUC [34]、DFANet [35]、DANet [36]、Auto Deeplab [37]、APCNet [38]、CANet [39] 等模型，这些模型在CNN模型结构上一步步发展，许多通用的模型结构设计方法和训练方法被提出和不断成熟。在以语义分割为主的图像合成任务中，CNN图像生成模型也在发展过程中在结构上形成了由一个编码器和一个解码器连接起来的编解码结构模式。采用CNN合成图像的大多数研究是有监督训练，且与其他多数CNN任务一样，任务中CNN起到的作用是提取输入的某种特征或对输入的特征进行转换而不是对输入添加更加丰富的细节，例如图像转换任务 [12, 13, 40, 41]或前述的一系列语义分割任务。CNN对训练标签的要求限制了CNN在图像合成领域的应用场景。

相反，GAN自然是采用无监督训练的方法，同时也可添加有监督损失进行有监督

训练 [42–44]。因此，最近越来越多的研究采用GAN来进行图像合成任务。由于编解码结构在CNN上的有效性，在采用GAN的图像合成的发展中，图像合成任务也被模式化为使用编码器和解码器组成的生成器进行像素到像素的映射 [19, 45–49]，其中编码器实现对输入进行特征提取和融合，解码器实现对特征进行填充还原和添加细节。GAN在图像生成任务中的惊人表现和相关数据集的发展逐步催生了图像合成领域中一些研究方向的火热，例如人像合成 [16–19]、风格迁移 [14, 15]和草图补全 [19, 50]等。在结构设计和训练方法上，DCGAN [51]通过采用更深层和复杂的CNN作为生成器和鉴别器使得合成质量在原始GAN的基础上有了很大的提升，展现了GAN的巨大潜力。随后ACGAN [20]在GAN最重要的变体CGAN [52]的基础上提出了多类别图像合成方法，为GAN的应用展现了一个更加广阔前景。之后CycleGAN [14]的出现，启发了后续许多基于GAN的多域图像转换的研究和应用。在CycleGAN的基础上，IcGAN [53]、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 [54]和XGAN [55]将网络模块化为多个部件又开启了另一种思路。

在研究者们的热情参与和贡献中，无论是基于CNN还是GAN的图像合成都得到了充分的发展，尤其是语义分割、人像合成等领域涌现出了大量的优秀图像合成应用。在这些研究的基础上实现高质量的医学影像的合成将具备充足的可行性，这一方面开拓了图像合成方法的应用领域，同时也能进一步的发展图像合成方法和技术。

2.2 医学影像合成的研究现状

在深度学习提出之前，一些研究使用图字典映射 [56]、稀疏编码 [57, 58]等方法进行了医学影像合成的最初尝试。随后在CNN的发展浪潮中，也有基于CNN的医学影像合成的研究出现 [40, 41]。

随着语义分割、人像合成等领域中基于GAN的图像合成的发展，将语义分割、人像合成等领域中优秀的图像合成方法，应用于高质量的医学影像的合成成为了医学影像合成研究的趋势。GAN逐步被广泛应用于医学影像的分割 [59]，重建 [60, 61]、合成 [42, 44, 62, 63]、转换 [41, 59, 64–69]和超分辨率 [70, 71]等各类研究。

其中，医学影像的模态转换是当前医学影像合成的研究中发展最成熟的方向之

一。一些医学影像模态转换的研究旨在减少医生和病人在诊断和治疗中产生的代价，例如无需配准数据的心脏MRI与CT之间的互相转换 [64]，脑部MRI转换合成CT图像 [65]，低剂量肺部CT合成高剂量CT [67–69]等一些通过无辐射或低辐射影像合成高辐射影像以减少病人辐射剂量的研究，并声称他们合成的医学影像可以直接提供给医生进行诊断，也有的研究甚至表示转换结果可以提高治疗的可行性 [56]。这些医学影像模态转换合成的研究中，许多研究仅尝试了两个模态之间的转换 [41,56,58,59,64–66]，对于多个模态的互相转换合成的研究则相对较少 [42,72,73]。在多模态医学影像合成的研究中，有研究 [72]实现了多输入多输出的MRI合成，但对输入的多模态数据要求配准，而在一些医学影像的转换配准的研究 [40]中CNN又详细展示了在模态配准方向的潜力，于是相关研究 [73]通过模型的自动学习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更高质量的MRI合成，但又限制了输出为单一模态。

与上述模态转换研究的目的不同，最近的另一些研究则试图使用合成医学影像来缓解医学影像数据样本稀少的难题，如脑MRI的合成研究 [42]、视网膜的合成研究 [62]、涵盖多种身体部位的单模态医学影像合成研究 [74]。

其中，脑MRI合成研究 [42]应用GAN合成脑肿瘤图像实现了通过合成影像的方式进行数据增强和数据匿名化，但其输入为从真实数据中提取的脑组织分割图，不仅需要额外的数据标签和额外训练一个脑组织分割网络，输入的分割图的数量也受限于真实数据集的大小，这使得该方法的合成受多种因素的限制。该研究还首次通过在输入时添加肿瘤分割标签来指导病灶的合成，然而其合成过程没有针对病灶合成的损失约束，仅采用了鉴别器提供的整体的对抗性损失进行指导，这使得病灶的合成极具不确定性，同时该研究也并未对其合成的肿瘤病灶的有效性进行深入验证和分析。

一项针对视网膜合成的研究 [62]通过变分自编码器(VAE) [75,76]实现了血管分割图的随机生成，进而以合成的血管分割图合成了彩色视网膜图像，可以实现任意数量的合成，但该研究仍需训练一个血管分割网络从真实图像中提取血管分割图来训练血管分割图的VAE生成网络。

而最近提出的SkrGAN [74]与本研究思路一定程度不谋而合，该研究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（Sketch）的提取，并通过GAN先实现草图的随机生成，再进一步实现了对医学影像的合成。SkrGAN在训

练中，仅进行了单模态影像的合成尝试，一方面由于始终采用从真实数据集中提取的草图进行训练，并以输入草图对应的真实影像为监督标签，使得其模型的训练极易过拟合且缺乏对合成草图的适应能力，最终导致模型合成的医学影像缺乏多样性；另一方面提取的草图对原图生理结构信息的勾勒效果较差，使得合成的影像在结构细节上存在残缺或模糊的情况。与前述绝大部分的研究相同的是，SkrGAN合成的医学影像没有考虑病灶信息的合成，也没有对应的病灶标签的产出，这使得绝大多数的合成数据难以在智能医学影像处理任务中直接使用。

2.3 小结

在上述对相关研究进行的分析中不难发现，当前合成医学影像尽管取得了长足的进展，具备了很好的研究基础，但依然有许多问题亟待解决，依然有继续深入研究的巨大价值和空间。总的来说，在前述各项研究的基础之上，本研究需要对多模态医学影像的合成方案的各个环节进行更科学的规划和改进，以实现多个数据集上更清晰简明的结构特征图的提取、更稳定且无限制的结构特征图的随机合成、更逼真的配准多模态医学影像的合成、可控可验证的病灶合成和更全面的合成数据的可用性评估。

第三章 智能医学影像处理的研究基础

智能医学影像分类、智能医学影像病灶分割、智能医学影像病灶检测等智能医学影像处理研究是一般图像分类、语义分割、目标检测等计算机视觉任务在医学影像上的应用扩展，均为基于卷积神经网络、生成对抗网络等深度学习方法进行的图像处理任务。本文在前两个章节中详细介绍了本研究的技术路线和相关的研究现状，在本章节中，将对智能医学影像处理研究在一般自然图像中的研究基础进行介绍，本文后续章节中提出的方法和实验评估将对本章介绍的基础方法和模型进行调用、改进和扩展。本章将对基于卷积神经网络的图像分类、基于生成对抗网络的图像合成、基于变分自编码器的图像合成、基于卷积神经网络的语义分割和基于卷积神经网络的目标检测这五项具体的智能医学影像处理研究依赖的基础性研究进行依次介绍。

3.1 基于卷积神经网络的图像分类

图像分类（Image Classification）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中图像所属类别的任务。卷积（Convolution）指的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的网络。通常，一个复杂的CNN中，卷积层与激活函数（Activation Function）、批归一化（Batch Normalization）、池化（Pooling）层、全连接（Full Connection）层等多层复合，数据层层计算并逐层传递到损失函数层，再通过基于随机梯度下降（SGD, Stochastic Gradient Descent）算法的反向传播机制，实现网络中卷积核的参数和其他可学习的参数不断地进行梯度更新，从而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 [77]提出，但直到2012年ILSVRC（ImageNet Large Scale Visual Recognition Challenge）[78]中CNN模型AlexNet [79]以高出10%的正确率力压第二名取得冠军，CNN的巨大潜力才逐步展现。此后CNN模型进入了飞速发展期，VGGNet [80]、GoogleNet [81]、ResNet [82]等经典网络相继提出，图像分类任务也随之快速发展。如图3.1所示，类似于VGGNet的CNN分类模型，接收图像作为输入，输

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 3.1: VGGNet模型结构设置 [80]。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。

出一个类别概率向量，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，通过最大概率类标取值函数 $\text{argmax}(\cdot)$ 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。

3.2 基于生成对抗网络的图像合成

生成对抗网络（GAN）[8]提出后，与CNN结合成为一种无监督学习的深度学习模型[51]。GAN包括两个主体部分，即一个生成器网络 G （Generator）和一个鉴别器网络 D （Discriminator），两个网络在训练过程中互相对抗， D 在对抗中学习对真实数据和生成数据的分布进行鉴别， G 在对抗中学习对真实数据的分布的模仿以生成逼真的数据。基于GAN的图像生成任务即通过训练生成器 G 来生成图像的生成任务。在普通的GAN中，生成器 G 以随机生成的噪声向量或矩阵 z 为输入，然后对噪声 z 进行层层变换，最后解码生成目标图像 x ；鉴别器 D 以图像 x 为输入，最后的输出为该图像是真实图像的概率 $D(x)$ ，若概率为100%则代表鉴别器判断输入的图像一定是真实的图像，输出为0代表鉴别器判断此图是与真实图像分布完全不同的图像。在对抗训练时，通过损失函数给定生成器 G 的学习目标是尽可能生成逼真的图像来使鉴别器 D 对生成图像的鉴别结果为真。而通过损失函数给定的鉴别器 D 的学习目标是尽可能对 G 生成的假图鉴别为假同时对真实的图像鉴别为真。损失函数约束 G 和 D 形成的对抗训练最终使得两个网络达到动态均衡，最终的平衡点称为纳什均衡点（Nash Equilibrium Point），平衡的标志是生成器生成的图像接近于真实图像分布，而鉴别器难以鉴别生成图像和真实图像的真假，对于给定的真实图像和生成图像的预测为真的概率接近于50%，即处于随机猜测水平。GAN的一个典型损失函数[8]如下：

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)}[\log(D(x))] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (3.1)$$

与一般的CNN相比，GAN包含两个独立的子网络，其特有的对抗性训练也与一般CNN的单一网络有监督训练过程不同。GAN中，鉴别器 D 可以视为生成器 G 的损失函数的一部分，即 G 的反向传播过程从 D 对 G 生成结果的鉴别结果开始， G 中网络参数更新所需的梯度来源于 D 而非CNN中那样直接来源于数据标签。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（RBM，Restricted Boltzmann Machine）和Generative Stochastic Networks（GSNs）[83]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；相比与一般CNN，GAN可以生成更加清晰、真实的数据，同时由于采用无监督学习，GAN既可以用于无监督学习任务，也可以用于半监督学习任务和混合监督学习任务。GAN的优势使得其应用场景十分丰富，比如图像风格迁移、超分辨率、草图补全、图像去噪等等。

ACGAN [20]在原始GAN结构的基础上进行了改进，使得生成器 G 可以生成多种不同类别的数据。相比较来说，原始GAN结构中生成器 G 只有随机噪声 z 作为输入，而ACGAN中生成器 G 的输入除了 z 还有一个类别输入 c 。对输出 x_g ，原始GAN结构中的鉴别器 D 输出的只有该图像真假鉴别结果 $d(x_g)$ ，而ACGAN中的鉴别器 D 除了真假鉴别结果 $d(x_g)$ 外还有一个分类鉴别结果 $c(x_g)$ 。对应的，ACGAN中的损失函数除了对抗性损失还有额外的类别一致性损失。图3.2中展示了ACGAN的结构和其他的GAN结构。

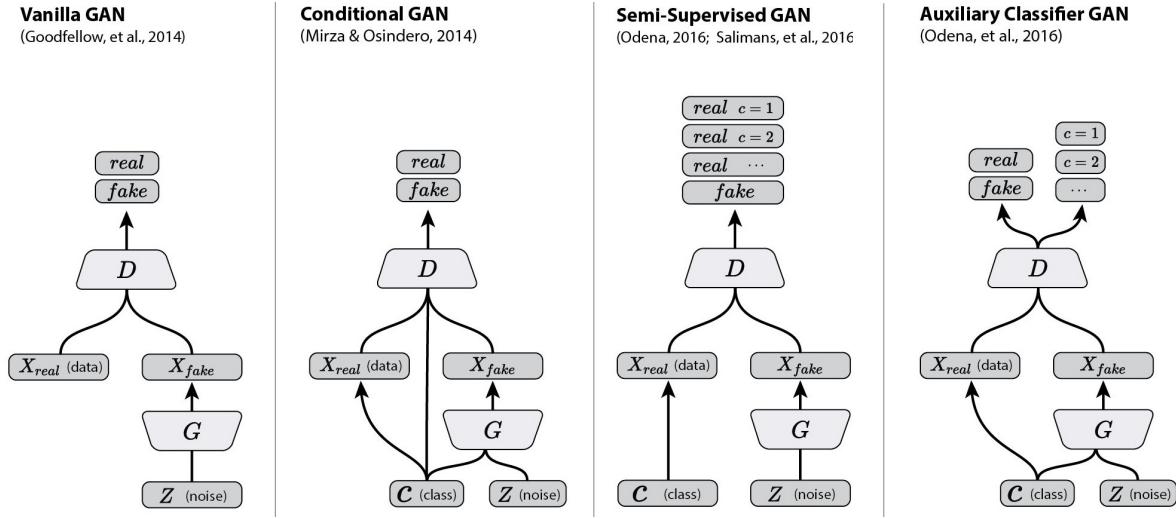


图 3.2: ACGAN模型结构¹。

3.3 基于变分自编码器的图像合成

变分自编码器（VAE，Variational Auto-Encoder）[84]是另一类重要的生成模型，

¹<https://twitter.com/ch402/status/793535193835417601>

它于2013年由Diederik P.Kingma和Max Welling最早提出，在2016年Carl Doersch的介绍论文 [85]发表后得到快速发展。VAE与GAN一样，是无监督学习最具前景的方法之一。VAE与GAN的目标是一致的，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，两者都假设了输入的 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生成服从目标数据分布的结果 $x = G(z)$ ，即这个模型能够将一个输入数据分布转换到目标数据的分布。一个VAE包含两个子网络，即一个编码器 E 和一个解码器 G ，编码器将数据分布的高级特征编码到一个低级特征空间得到特征图 z 。解码器接收数据的低级表征，然后解码还原数据的高级特征，即重建的数据 x_r 。

在标准VAE中，特征图 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 [84]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，VAE [84]文中对此进行了详细的推导和展开，由此推导得到的损失函数如下所示：

$$\min_{E,D} L(E, D) = \mathbb{E}_x[\|x - D(\tilde{z})\|_2^2], \quad (3.2)$$

$$\text{s.t. } \tilde{z} = \mu(x) + \exp(0.5 \times \sigma(x)) \times z, \quad (3.3)$$

$$[\mu(x), \sigma(x)] = E(x), \quad (3.4)$$

$$z \sim \mathcal{N}(0, 1^2). \quad (3.5)$$

3.4 基于卷积神经网络的语义分割

图像语义分割是计算机视觉领域的一个经典任务，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）。在基于深度学习的计算机视觉研究出现以前，随机森林和条件随机场等是常用的语义分割方式。2015年FCN [22]首次将深度学习应用在图像语义分割任务上，通过对全尺寸图像进行端到端的分割取得了十分显著的提升。此后U-net [24]在生物医学影像的语义分割任务上采用编码器-解码器结构并应用跳跃连接（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3.3所示，U-net接收一张图像输入，输出为分割结果的概率矩阵，通过 $\text{argmax}(\cdot)$ 函数即可得到分割掩膜标签。

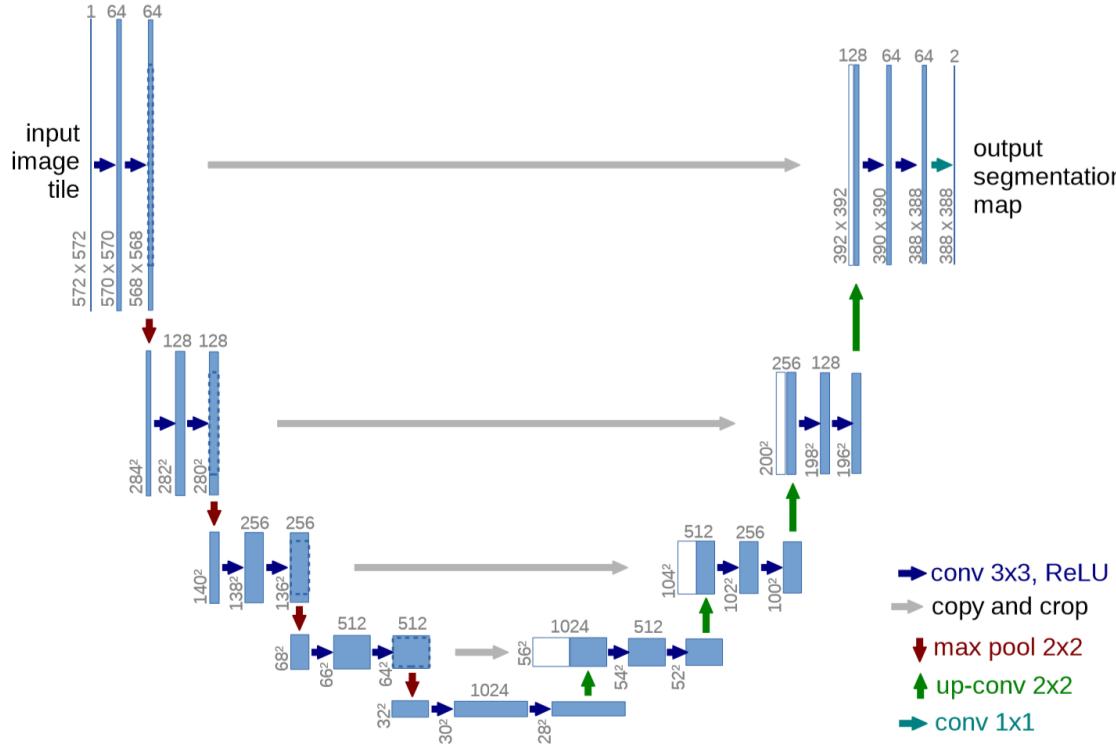


图 3.3: U-net模型结构 [24]。

3.5 基于卷积神经网络的目标检测

图像目标检测同样是计算机视觉领域经典任务之一。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许多突破性进展。当前的研究中存在两类研究思路略有不同的目标检测模型：一是基于候选区域（Region Proposal）的R-CNN [86]类双阶段模型方案，包括R-CNN，Fast R-CNN [87]，Faster R-CNN [88]等模型，它们均包含两个阶段，首先要前向推理生成目标候选框，其次再对候选框进行分类和回归对候选框进行筛选，得到的即为最终输出的检测框；以YOLO [89]、SSD [90]等模型为代表的另外一类单阶段模型方案，只需要应用单个卷积神经网络即可实现对多个目标的类别与位置的直接预测和输出。

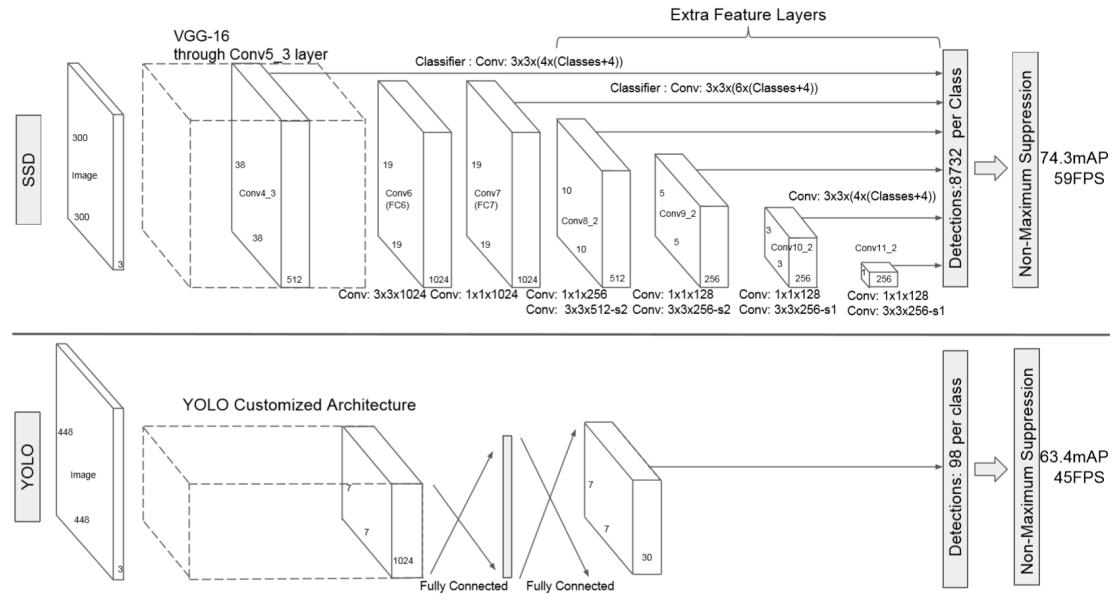


图 3.4: YOLO 和 SSD 模型结构的对比 [90]。

如图3.4所示，SSD模型基于VGG16 [80]模型进行修改，先将图像输入到预训练好的分类网络中来获得不同大小的特征图（Feature Map），通过抽取的多层次卷积层中的特征图构造出多个固定数量的不同尺寸的检测候选框（Anchor Box），然后与根据真实类别（背景也被算作一个类别）的检测框生成的同样数量的先验框计算检测和分类的损失，再对损失求导更新模型参数。预测时，同样先生成多个候选框，然后再将不同特征图获得的候选框结合起来，通过非极大值抑制（NMS，Non-Maximum Suppression）方法抑制掉一部分重叠或者不正确的候选框，生成最终的候选框集合，得到所求的包含目标位置、目标类别和类别置信度的目标检测结果。

第四章 医学影像结构特征图的提取和生成方法

本文在上一章中对智能医学影像处理研究基础中的几种主要的基于深度学习的图像处理任务和深度学习经典模型进行了介绍。本章针对第一章中介绍的医学影像合成研究中的首个挑战，基于第一章中介绍的研究思路和上一章中介绍的基础方法和模型，提出了一种为医学影像合成提供生理结构指导信息的结构特征图的提取与生成方法。本章还将对本文提出的结构特征图的提取和生成方法相比现有结构信息获取方法的改进和创新进行介绍。

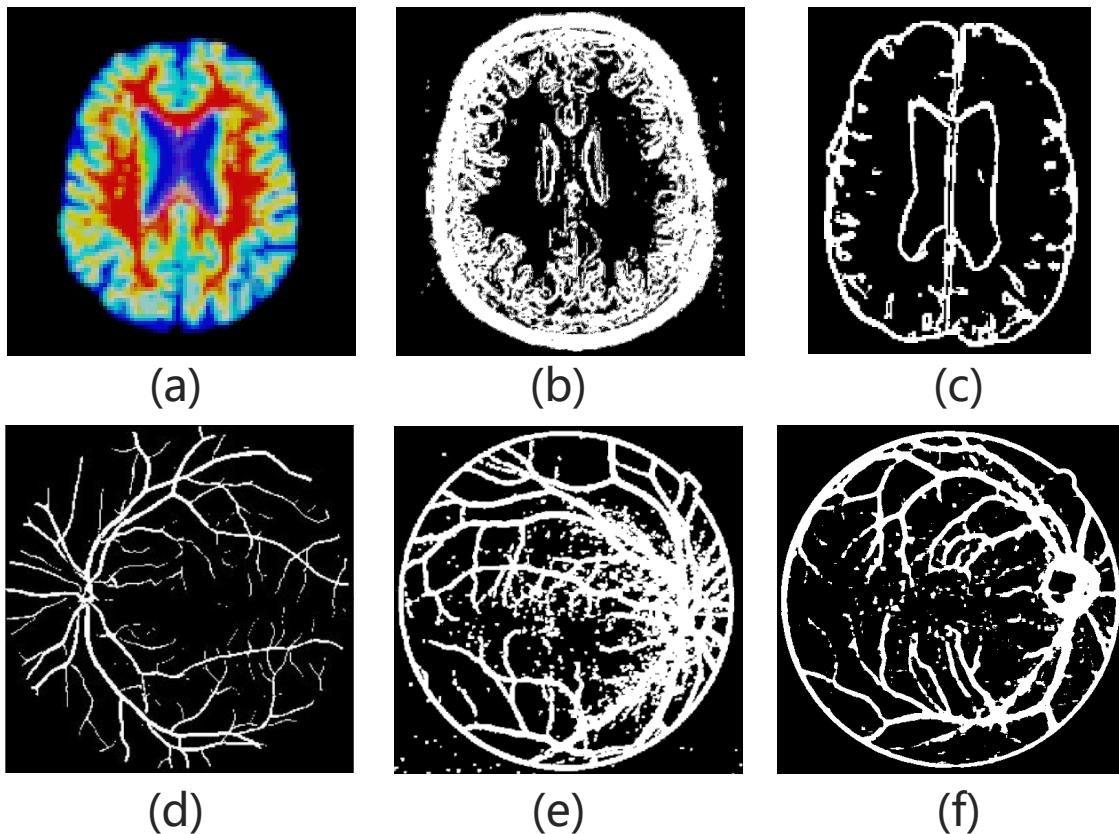


图 4.1: 各种提供生理结构信息的结构特征图。(a) 为脑组织分割标签 [42]，(b) 为SkrGAN [74]提取出的脑结构草图的反色效果图，(c) 为本文方法提取出的脑结构特征图，(d) 为视网膜血管分割标签 [62]，(e) 为SkrGAN [74]提取出的视网膜结构草图的反色效果图，(f) 为本文方法提取出的视网膜结构特征图

GAN直接从随机噪声中生成的医学影像很难生成真实合理的结构信息。如图4.1，

本文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（图4.1（d））可以看作是视网膜图像的结构特征图 [62]。结构特征图可以为合成医学影像提供必要的生成指导。在合成医学影像时，一些研究从脑组织分割标签 [42]（图4.1（a））获取了基本的脑结构信息。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，如图4.2。为避免这种额外的数据需求和训练代价，本研究设计了下述的从医学影像直接提取结构特征图的方法，该方法具有提取效果更具描绘能力、无需额外训练数据和提取训练等优点。

4.1 结构特征图提取方法

算法 1 结构特征图提取方法

输入：灰度图 x ; 像素阈值 α, β, γ ; 高斯核方差 σ_1, σ_2 .

输出：结构特征图 s .

- 1: $[ver, hor] = \text{sobel}(x)$
 - 2: $s_1 = \text{reduce_min}([ver, hor])$
 - 3: $s_2 = \text{reduce_max}([ver, hor])$
 - 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: $s = \text{medfilt}(s) \times (x > \gamma)$
-

在传统的数字图像处理方法中，Roberts算子 [91]、Prewitt算子 [92]、Sobel算子 [21]等都是常用的边缘检测算子。Sobel算子尤其适用于也最常用于处理医学影像。如

算法 1 中所示，本研究探索得出了一种从 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。

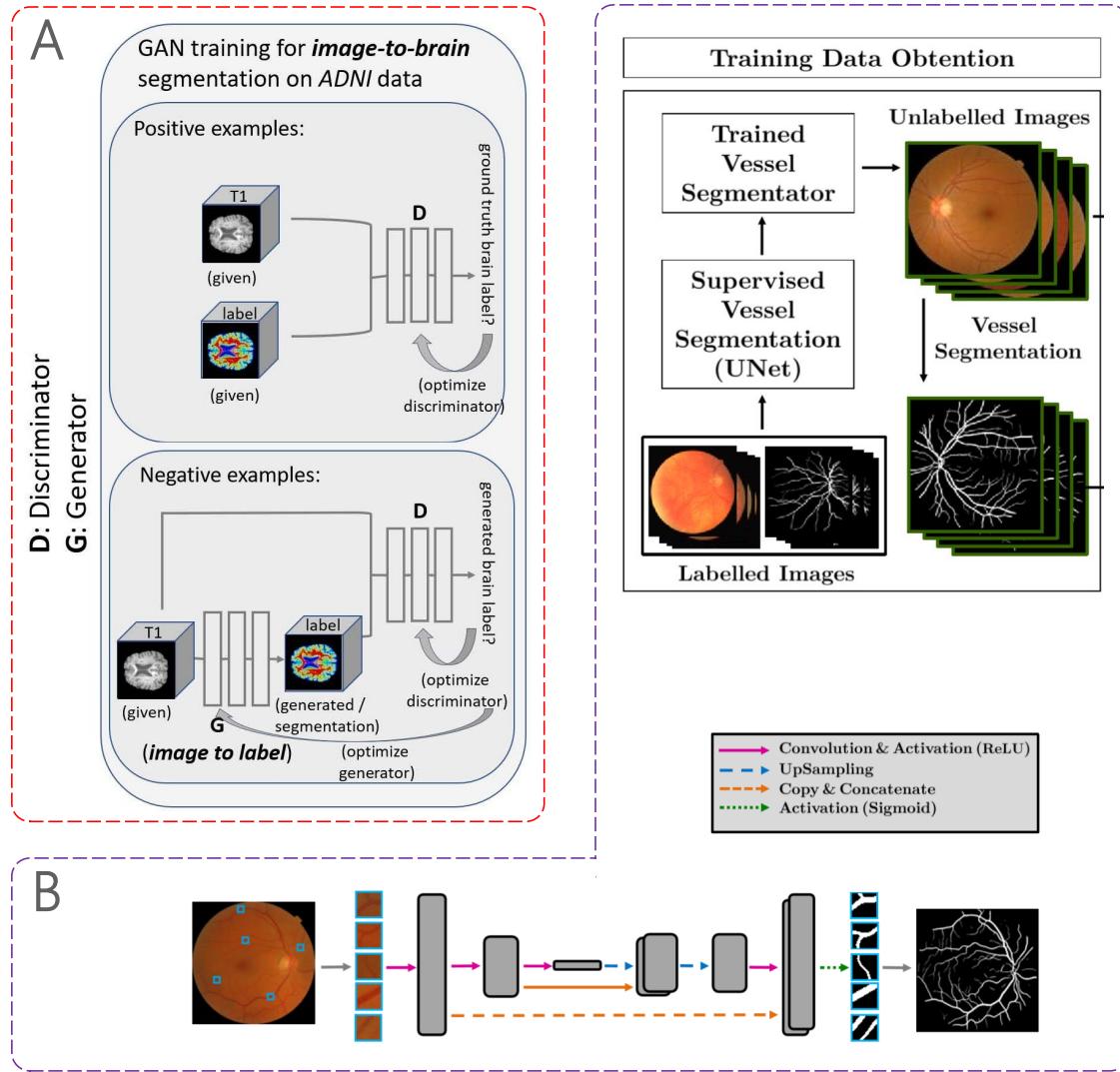


图 4.2: 当前两种需要训练的结构信息获取方法: 脑组织分割标签图提取方法 (A) [42] 和视网膜血管分割标签图提取方法 (B) [62, 93]。

算法 1 (Algorithm1(.)) 中, 使用 Sobel 检测算子 $\text{sobel}(\cdot)$ 从真实图像中提取水平和垂直边缘检测图 (如图 4.3 所示)。通过对得到的两个边缘检测图一起执行最大值规约操作 $\text{reduce_max}(\cdot)$ 和最小值规约操作 $\text{reduce_min}(\cdot)$, 可获得两个融合了水平垂直边缘检测图的结果图, 再对这两张图进行核尺寸为 3×3 的高斯模糊 [94] $\text{gaussian_blur}(\cdot)$ 以实现对线条轮廓的加粗, 然后将两张图与通过均值函数 $\text{mean}(\cdot)$ 计算出的平均像素值

作差，由此可去掉大部分的背景像素，只保留图中最突出的线条轮廓。再根据像素阈值对两个差值图进行二值化，然后对两个二值图求和后二值化可获得更完整的轮廓线条二值图。再进行一次高斯模糊可以对提取出的线条加粗，并能使有断点的线条相连成整体。最后再把结果和原始图的二值掩膜图相乘，可以完全去掉对应于原始图背景区域的噪声，再用 3×3 的中值滤噪函数 $\text{medfilt}(\cdot)$ 去掉剩余的孤立的噪点，即可得到需要的干净清晰又完整简洁的结构特征图。

在图4.1（a）中示例的脑组织分割标签的获取 [42]中，首先需要额外提供一个由脑MRI影像和配对的脑组织分割标签图组成的数据集，再在此数据集上进行如图4.2中A图所示的分割训练，获得一个脑组织分割器，然后再使用该脑组织分割器对另一个更大的数据集的脑MRI进行分割，得到可以给脑MRI合成提供指导信息的脑组织分割图，最后再使用分割得到的这些脑组织分割图合成与目标数据集中脑MRI类似的合成脑MRI。在图4.1（d）示例的视网膜血管分割标签图的获取 [62]中，同样首先需要额外提供一个由视网膜和配对的血管分割标签图组成的数据集，再在此数据集上进行如图4.2中B图所示的分割训练，获得一个视网膜血管分割器，再使用该视网膜血管分割器对目标数据集中的视网膜进行分割得到可以给视网膜合成提供指导信息的视网膜血管分割图，然后再使用分割得到的这些视网膜血管分割图训练视网膜血管生成器，最后再使用通过视网膜血管生成器合成的视网膜血管分割图进一步合成与目标数据集中视网膜类似的合成视网膜图像。

与本节的方法相比，上述两个研究中，脑MRI的合成需要额外的脑组织分割标签和分割训练，由于合成始终采用从真实影像提取的脑组织分割标签，其合成脑MRI数量和多样性将因此受到限制；视网膜的合成同样需要额外的视网膜血管分割标签数据和额外的训练，但该方法通过一项视网膜血管标签的合成训练，可以通过训练出的视网膜血管生成器获得无限多的合成视网膜血管标签，这可以为视网膜的合成提供不限数量和多样性充足的输入样本。

SkrGAN [74]的方法与本研究的思路有相似之处，其利用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并填充囊状结构得到了结构草图。和SkrGAN的方法相比，本文中的方法分别考虑了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。最后本文中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，本研究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 结构特征图的病灶结构移除

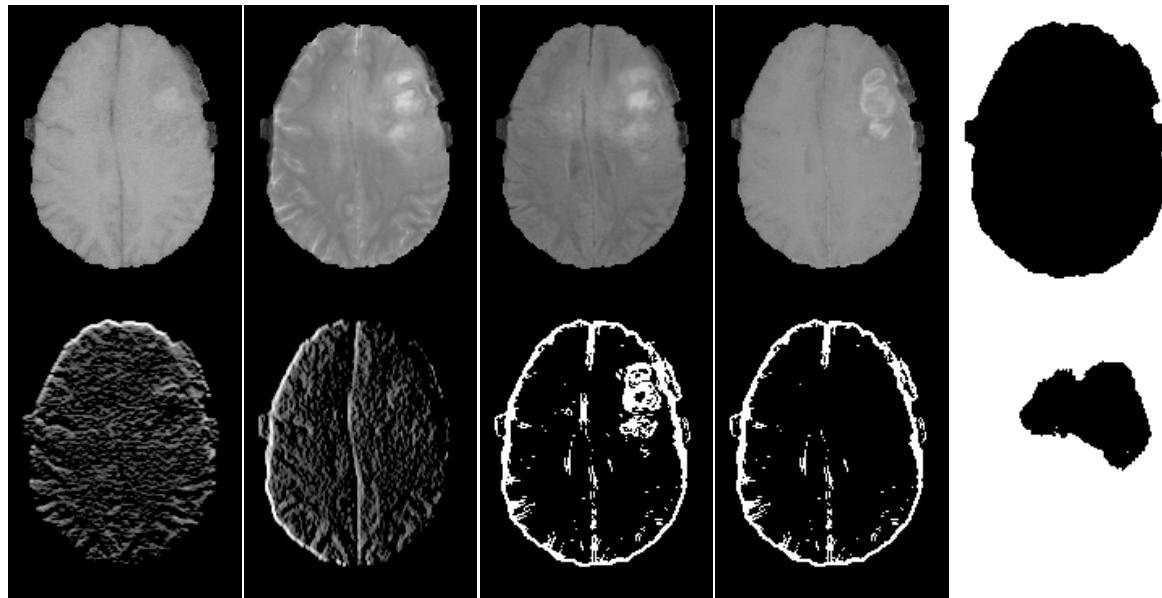


图 4.3: BRATS2015数据集中的多模态脑MRI和从中提取结构特征图产出的图像。其中第一行从左到右依次为MRI的四个模态T1、T2、T1c、Flair和从Flair提取出来的掩膜，第二行从左到右依次为从Flair采用Sobel算子提取出来的水平向和垂直向边缘检测结果图、从Flair提取出来的结构特征图、采用肿瘤掩膜去掉肿瘤病灶结构的结构特征图、肿瘤分割标签二值转换的肿瘤掩膜。

当没有足够的正常医学影像来产生结构特征图时，需要从有病灶的医学影像中提取结构特征图来丰富训练样本。对于许多有肿瘤、结节等病灶的医学影像，提取出的结构特征图很可能保留有病灶的结构信息，此时需要设法移除其中的病灶结构。相关脑MRI合成研究 [42]中正是由于没有合适的方法去除选用数据集中脑MRI的肿瘤病灶

的影响，因此不得不需要一个额外的正常脑MRI数据集来提供正常的脑组织分割标签图以合成高质量脑MRI。

如图4.3所示，对于肿瘤、结节等一般配有病灶位置标签且病灶结构只覆盖在局部范围的数据集，可先对病灶分割标签图 l 采用算法2（Algorithm2(\cdot)）执行转换，获得对应的病灶分割掩膜 m_l ，即 $m_l = \text{Algorithm2}(l)$ ；然后，通过掩膜 m_l 即可去除结构特征图 s 中的原始病灶部分的结构信息，公式表示如下：

$$s' = s \times m_l. \quad (4.1)$$

对于肺炎等病灶覆盖整体或无法对病灶进行较精准定位的数据集，有病灶的样本如无特别需求不宜用于提取结构特征图，应该直接弃用。

对有病灶信息的结构特征图，在对结构特征图处理进行上述病灶结构移除后，才可进行接下来的合成训练，以避免合成的结构特征图中出现病灶结构。对从已弃用有病灶样本的数据中提取的结构特征图可直接进入结构特征图生成训练环节。

4.3 结构特征图生成训练

算法2 掩膜提取方法

输入：灰度图 x ；像素阈值 α ；扩展像素宽度 p 。

输出：掩膜 m 。

- 1: $m = 1.0 - \text{ones} \times (x > \alpha)$
 - 2: $\text{size} = [x.\text{width} + p, x.\text{length} + p]$
 - 3: $m = \text{resize}(m, \text{size})$
 - 4: $m = \text{crop_padding}(m, p)$
 - 5: $m = \text{medfilt}(m)$
-

生成提供结构信息的输入时，相关脑MRI合成研究[42]仍需要输入真实图像以获取生成的脑组织分割图，这大大减少了合成数据的多样性。合成视网膜的研究[62]中，实现了基于VAE的方法从多维正态分布生成视网膜血管分布图。SkrGAN则采用GAN从随机噪声合成结构草图。以此为基础，本研究设计了一种结合VAE与GAN的混合网络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

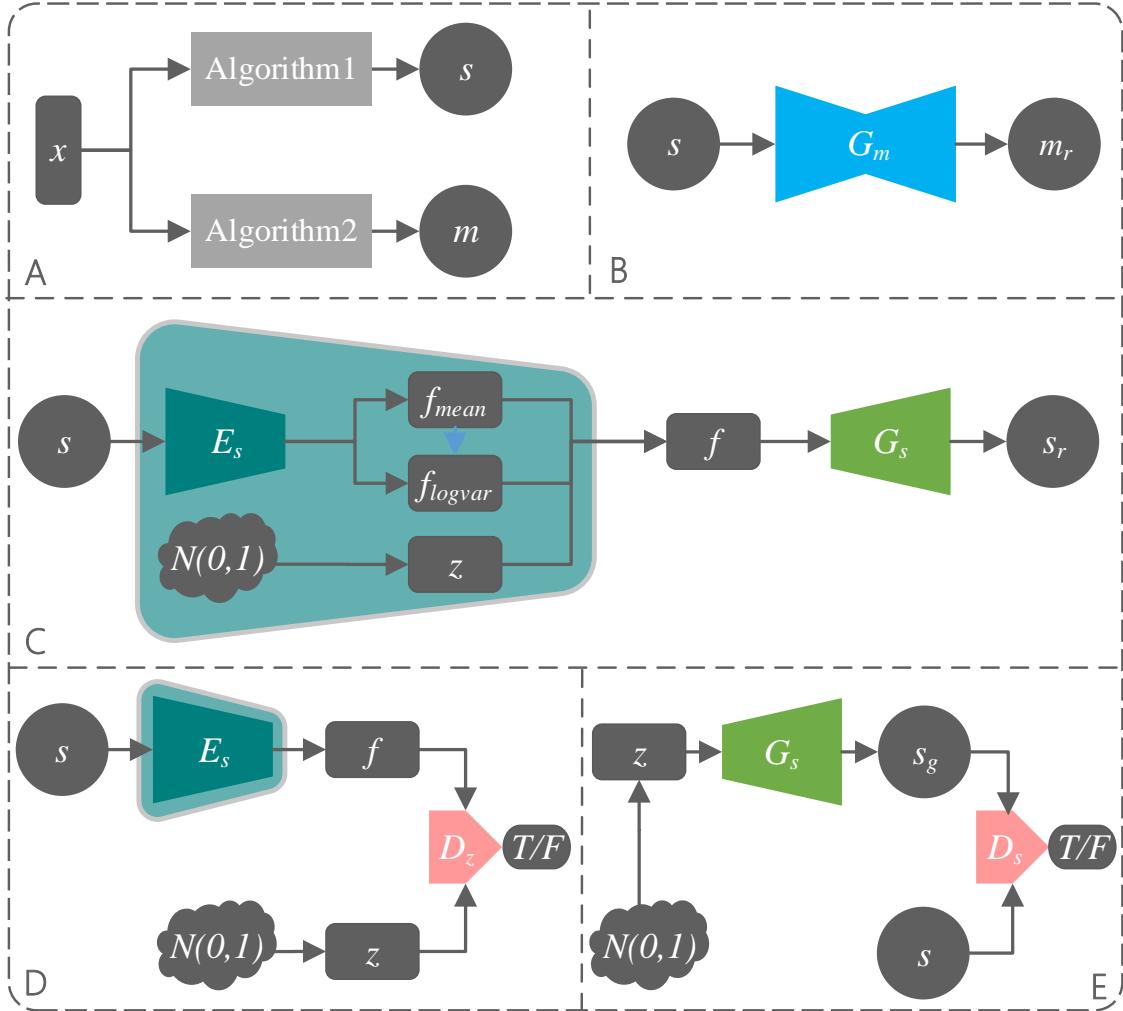


图 4.4: 结构特征图生成训练。A-E依次为结构特征图和掩膜的提取、掩膜的生成训练、VAE重建结构特征图训练、VAE编码器的对抗性训练和VAE解码器的对抗性训练。 x 是输入的真实影像， s 是结构特征图。 E_s 是VAE编码器，输出编码矩阵 f_{mean} 和 f_{logvar} 。 z 是来自多维正态分布 $\mathcal{N}(0, 1^2)$ 的随机噪声采样，而 f 是近似正态分布矩阵。 G_s 是VAE解码器， s_r 是重构的结构特征图，而 s_g 是生成的随机结构特征图。 D_s 是结构特征图鉴别器， D_z 是特征图分布鉴别器。 G_m 是掩膜生成器， m_r 是从 s 生成的掩膜。

多样性并且无需其他训练标签的结构特征图。此外，本研究的方法中还额外训练了一个掩膜生成器 G_m ，用于从结构特征图中获取目标器官区域的掩膜，生成的掩膜在后续环节中用于匹配病灶标签等。掩膜的生成训练与结构特征图的生成训练同步进行。在 G_m 训练期间，算法2提取的掩膜用作标签数据，其中 $\text{resize}(\cdot)$ 为最近邻插值函数， $\text{crop_padding}(\cdot)$ 为边距裁切函数。如图4.4所示，具体的训练过程如下：

- 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 ；
- 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；

如图4.5所示， E_s 、 G_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 在U-net的模型结构的基础上进行了调整适配。 E_s 为一个最后四层双线输出两个结果的正向VGG11（编码器）而 G_s 为一个反向VGG11（解码器），两者对接形成一个编解码结构的网络。掩膜生成器 G_m 的结构与下一章中模态生成器一致，但无需输入独热条件矩阵，详细结构如图5.4所示。其中，对所有模型的改进都包括将所有步长为2的池化操作改为步长为2的卷积操作，将所有步长为2的反卷积（Deconvolution）上采样改为在反卷积基础上与步长为2的最近邻插值上采样及后接的单层卷积堆叠，在所有除输出层外的卷积层之后执行实例归一化（Instance normalization）和Leaky ReLU非线性激活。各个网络的损失函数如下，其中 \mathbb{E} 为在指定数据分布上的期望（Expectation）函数：

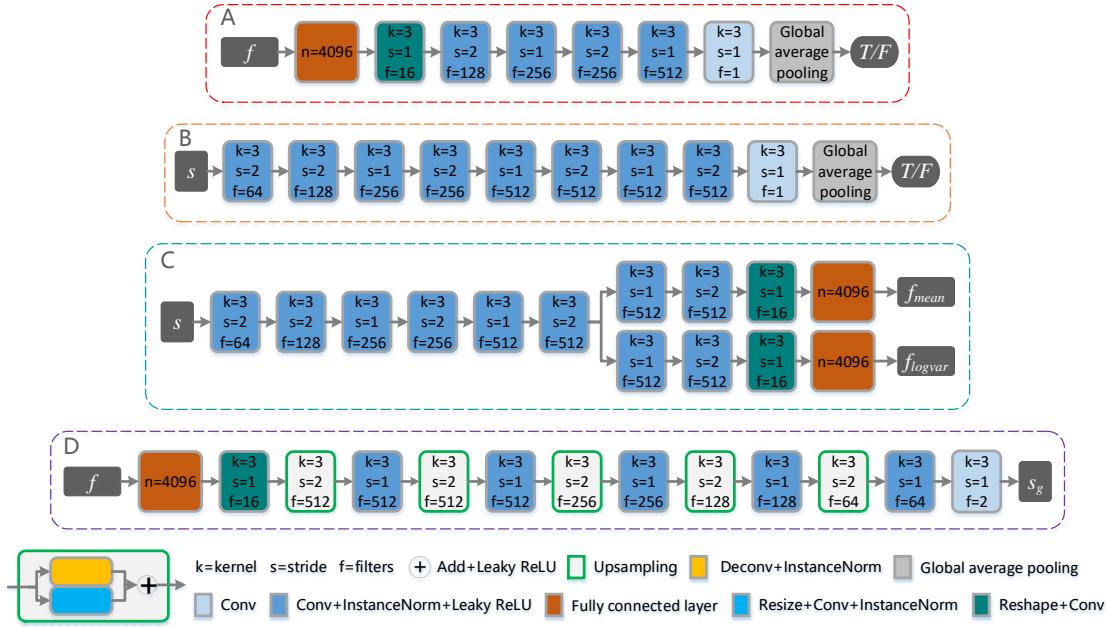


图 4.5: 结构特征图生成训练阶段各组件的网络结构。A-D依次为特征图分布鉴别器 D_z 、结构特征图鉴别器 D_s 、结构特征图编码器 E_s 、结构特征图生成器 G_s

- 阶段B: 掩膜生成损失

$$\mathcal{L}_m(G_m) = \mathbb{E}_{m,s}[\|m - m_r\|_2^2], \quad (4.2)$$

其中 $m_r = G_m(s)$ 。

- 阶段C: 结构特征图重建损失

$$\mathcal{L}_r(E_s, G_s) = \mathbb{E}_{s,f,m}[\|s - s_r\|_2^2 + \|m_r \times s_r\|_2^2], \quad (4.3)$$

其中 $s_r = G_s(f)$ 。

- 阶段D: 特征图分布对抗训练损失

$$\mathcal{L}_{d,z}(D_z) = \mathbb{E}_{s,z}[\|D_z(z) - 1\|_2^2 + \|D_z(f)\|_2^2], \quad (4.4)$$

$$\mathcal{L}_{g,z}(E_s) = \mathbb{E}_z[\|D_z(f) - 1\|_2^2], \quad (4.5)$$

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

$$\mathcal{L}_{d,s}(D_s) = \mathbb{E}_{s,z}[\|D_s(s) - 1\|_2^2 + \|D_s(s_g)\|_2^2], \quad (4.6)$$

$$\mathcal{L}_{g,s}(G_s) = \mathbb{E}_z[\|D_s(s_g) - 1\|_2^2 + \|m_g \times s_g\|_2^2], \quad (4.7)$$

其中 $s_g = G_s(z)$, $m_g = G_m(s_g)$ 。

4.4 结构特征图的过滤

算法 3 结构特征图过滤方法

输入: MAE阈值 δ .

输出: 合成的结构特征图 s .

```

1: do
2:    $s_g = G_s()$ 
3:    $m_g = G_m(s_g)$ 
4:    $s_g = \text{gaussian\_blur}(s_g)$ 
5:    $\text{contours} = \text{OpenCV.findContours}(s_g)$ 
6:    $m'_g = \text{OpenCV.drawContours}(s_g, \text{contours})$ 
7:    $\delta' = \text{MAE}(m'_g, m_g)$ 
8: while  $\delta' <= \delta$ 

```

对于一些目标器官主体轮廓为封闭曲线的结构特征图，由于在训练前对包含肿瘤等病灶的结构特征图进行病灶移除操作可能破坏其封闭性，因此在合成训练时产生一定程度的影响，再加上合成训练本身具有的随机性，这使得合成的结构特征图难免存在极少数合成质量较差结构特征图，一般表现为目标器官主体轮廓未很好闭合或部分结构残缺。为此，本研究设计了一种结构特征图过滤方法，如算法 3 所示。首先，对合成的结构特征图，使用掩膜生成器 G_m 生成结构特征图的掩膜 m 。然后，通过对合成结构特征图执行高斯模糊 $\text{gaussian_blur}(\cdot)$ 可以加粗图中线条，再使用开源算库OpenCV²提供的轮廓搜索算法 $\text{findContours}(\cdot)$ 和填充算法 $\text{drawContours}(\cdot)$ 来获得

²<https://opencv.org/>

合成结构特征图中所有闭合的轮廓并在其中填充像素，这样，通过一般算法获得了另一个合成结构特征图的掩膜 m' 。此时，通过计算两个掩模的平均绝对误差（MAE, Mean Absolute Error）可以判断对应合成结构特征图中的目标器官轮廓是否闭合。如果实际计算出的MAE值 δ' 低于设置的阈值 δ ，则意味着结构特征图的主体轮廓相对完整，可以保存此结构特征图；如果 δ' 高于设置的阈值则表明通过一般算法获得的结构特征图的掩膜的内部存在空心区域，与掩膜生成器生成的掩膜有很大的不同，这意味着合成的结构特征图的轮廓不是完整封闭的，因此需要重新生成。

4.5 结构特征图与随机噪声的融合

在使用结构特征图合成医学影像时，由于结构特征图是稀疏的二值图，过多的单值像素会使得模型难以训练，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，特别是在使用小数据集时，训练将尤其困难。通常的GAN在训练中，通过以随机噪声矩阵为输入来获得无限多样的输入样本。因此，本研究设计了如下的计算公式，向结构特征图中目标器官轮廓内的0值区域填充随机噪声，以融合结构特征图的结构信息和随机噪声的随机信息：

$$s' = s + z' \times (1 - m) \times (1 - s), \quad (4.8)$$

其中， z' 是从均匀分布 $\mathcal{U}(\alpha_1, \alpha_2)$ 采样的随机噪声，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 配对的二值掩膜图。如图6.1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视觉上更为接近，可以降低学习难度。

第五章 基于结构特征图的多模态医学影像合成方法

本章提出的解决方案将对第三章中介绍的基础方法和模型直接应用或进行简单适应性调整，但不再对应用的基础方法展开介绍。本章提出的方案致力于解决当前研究在合成多模态医学影像时仍存在的问题和不足，实现以结构特征图为输入的带病灶标签的配准多模态医学影像的合成。本章先对提出方案的整体架构和阶段划分进行介绍，随后对医学影像合成的主要阶段要解决的问题和详细的方法步骤进行介绍。

5.1 整体架构

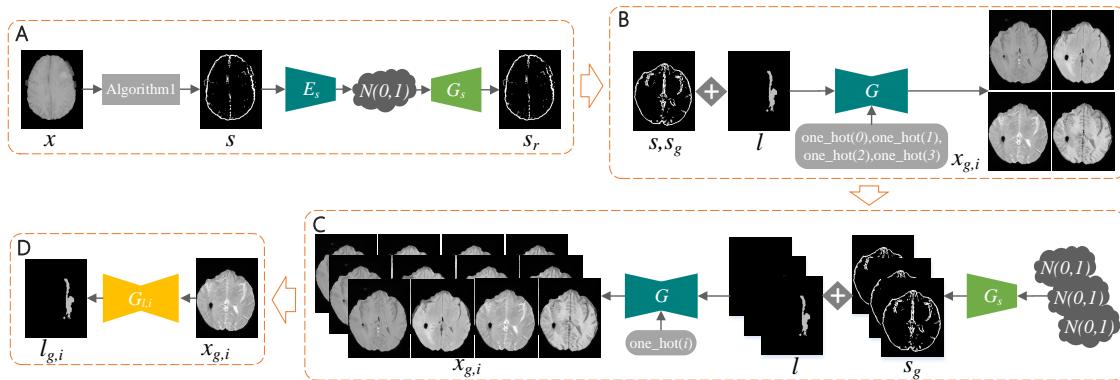


图 5.1: 整体架构。A是结构特征图的提取和生成阶段。B是多模态影像合成阶段。C是合成数据集构建阶段。D是合成数据可用性验证阶段。

图5.1为本研究完整方案的整体架构图，清晰展示了本研究合成医学影像完整方案的四个主要阶段及阶段之间的关联关系。结构特征图提取和生成阶段为第一个阶段，在上一章中进行了详细的介绍，主要的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任意数量的结构特征图为后续阶段模型的训练提供关键的输入数据。多模态影像合成阶段为第二个阶段，以上一阶段中的结构特征图为输入，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，同时通过病灶处理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确保实现

像素级配准。第三个阶段为合成数据集构建阶段，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。最后的合成数据可用性验证阶段，实现对合成医学影像数据在智能医学影像处理任务中的可用性进行评估验证，该阶段在下一章进行详细介绍。

5.2 多模态影像的合成

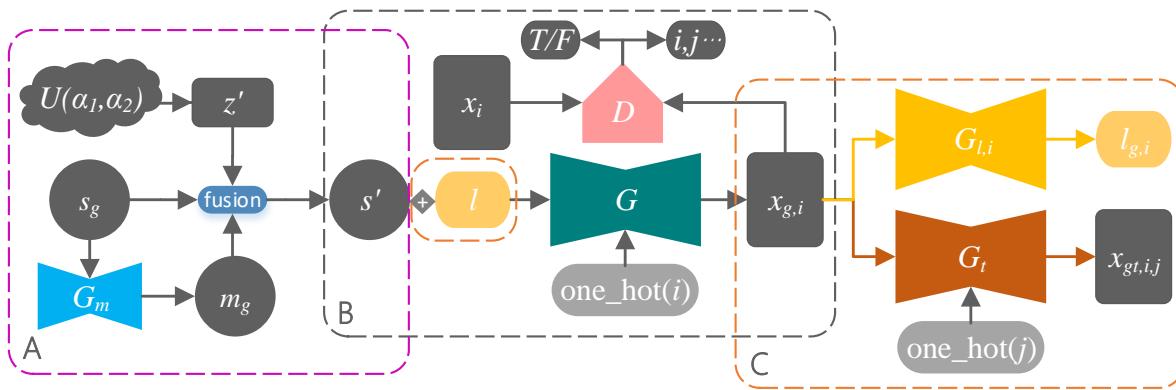


图 5.2: 多模态影像的合成。紫色框A中为结构特征图与随机噪声融合过程；灰色框B中为核心的医学影像合成过程；黄色框C中为可选的病灶合成与配准的指导和约束过程。 s_g 是随机合成的的结构特征图， m_g 是根据 s_g 生成的掩膜， z' 是从均匀分布 $\mathcal{U}(\alpha_1, \alpha_2)$ 采样的随机噪声， s' 是融合了噪声的结构特征图， l 是随机选择的真实病灶标签。one_hot(i)是模态*i*的独热矩阵表示。 G 是条件生成器， D 是鉴别器， x_i 是模态*i*的真实影像。 $G_{l,i}$ 是模态*i*的病灶处理器， $l_{g,i}$ 是病灶处理器从 $x_{g,i}$ 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5.2所示，首先，本节通过上一章中训练好的结构特征图生成器 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，本节使用一个接受独热（One-Hot）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态的合成图像 $x_{g,i}$ 。本节使用鉴别器 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。合成的多模态影像过程可以根据需要通过病灶处理器 $G_{l,i}$ 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。

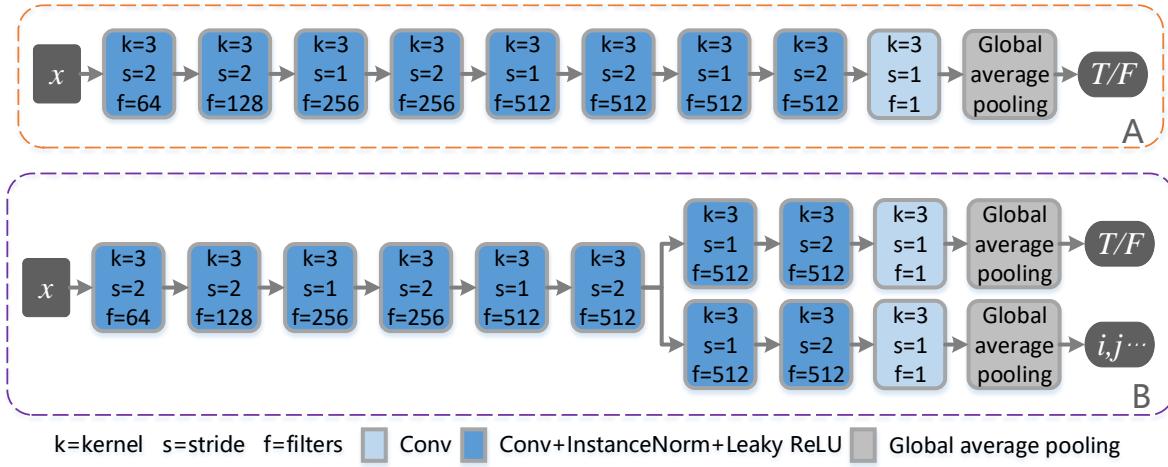


图 5.3: 鉴别器D的网络结构。A为合成单模态影像时采用的单线结构，B为合成多模态时采用的双线结构。

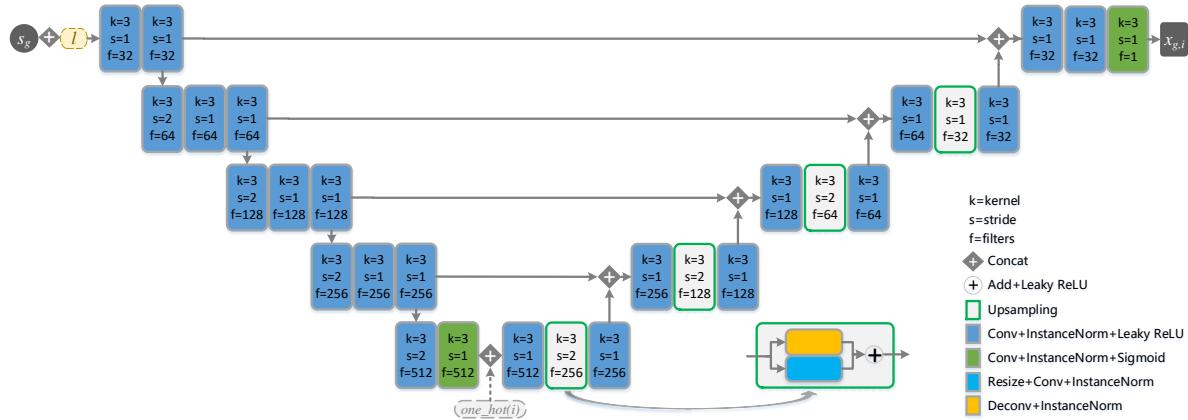


图 5.4: 生成器G的网络结构。虚线结构为可选过程。

D 在VGG11模型结构的基础上进行了调整适配，详细结构如图5.3所示。 G 在U-net的模型结构的基础上进行了调整适配，详细结构如图5.4所示。在单模态时， D 为单线结构， G 仅在网络第一层接收输入， G 与 D 组成一组基本GAN结构。在多模态时， D 为在最后三层输出两个结果的双线结构，即输出包括真假鉴别结果矩阵和类别鉴别结果矩阵， G 则以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 [20]结构。其中，对所有模型的改进都包括将所有步长为2的池化操作改为步长为2的卷积操作，将所有步长为2的反卷积上采样改为在反卷积基础上与步长为2的最近邻插值上采样及后接的单层卷积堆叠以平滑棋盘格效应（Checkerboard Artifacts），在所有除输出层外的卷积层之后执行实例归一化和Leaky ReLu非线性激活。

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 $D(x_i)$ 中的真假鉴别结果和模态类别鉴别结果，同样地， $d(x_{g,i})$ 、 $c(x_{g,i})$ 为 $D(x_{g,i})$ 中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

$$\begin{aligned} \mathcal{L}_d(D) = \mathbb{E}_{x,s_g,l} & \left[\sum_{i=0} (\|d(x_i) - 1\|_2^2 + \|d(x_{g,i})\|_2^2 + \right. \\ & \left. \|c(x_i) - i\|_2^2 + \|c(x_{g,i}) - i\|_2^2) \right], \end{aligned} \quad (5.1)$$

$$\mathcal{L}_g(G) = \mathbb{E}_{s_g,l} \left[\sum_i (\|d(x_{g,i}) - 1\|_2^2 + \|c(x_{g,i}) - i\|_2^2) \right], \quad (5.2)$$

其中 $x_{g,i} = G(\text{concat}([s', l]), \text{one_hot}(i))$ ， $s' = s_g + z' \times (1 - m_g) \times (1 - s_g)$ ， $m_g = G_m(s_g)$ ， z' 是从均匀分布 $\mathcal{U}(\alpha_1, \alpha_2)$ 采样的随机噪声且 $\alpha_1 = 0.5$ 、 $\alpha_2 = 0.6$ ， $\text{concat}(\cdot)$ 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

$$\mathcal{L}_{les}(G) = \mathbb{E}_{s_g,l} \left[\sum_{i=0} (\|l - l_{g,i}\|_2^2) \right], \quad (5.3)$$

其中 $l_{g,i} = G_{l,i}(x_{g,i})$ 。

- 模态配准损失

$$\mathcal{L}_{reg}(G) = \mathbb{E}_{s_g, l} \left[\sum_{j=0} \sum_{i=0, i \neq j} (\|x_{g,i} - x_{gt,j,i}\|_2^2) \right], \quad (5.4)$$

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

则多模态合成生成器的总损失为：

$$\mathcal{L}(G) = \mathcal{L}_g(G) + \mathcal{L}_{les}(G) + \mathcal{L}_{reg}(G). \quad (5.5)$$

从真实影像提取的结构特征图与真实医学影像可以实现自监督预训练，预训练过程可大大降低对抗性训练的难度，尤其是在小数据集上的训练中。预训练过程的损失函数如下：

$$\mathcal{L}_{pre}(G) = \mathbb{E}_{s,l} \left[\sum_{i=0} (\|x_{g,i} - x_i\|_2^2 + \|x_{g,i} \times m_i - x_i \times m_i\|_2^2) \right], \quad (5.6)$$

其中 $x_{g,i} = G(\text{concat}([s'_i, l_i]), \text{one_hot}(i))$, $s'_i = s_i + z' \times (1 - m_i) \times (1 - s_i)$, s_i 为采用算法 1 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 为采用算法 2 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 [74] 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 将自监督损失 \mathcal{L}_{pre} 与其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, 因而导致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较差。本文的方案中, 首先采用真实的结构特征图进行真实影像的重建预训练, 此时损失函数即为自监督损失 \mathcal{L}_{pre} , 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。这样既通过预训练加速了训练进程, 又确保了在对抗性训练中提高模型的泛化能力。

此外, 本研究的方法通过训练一个模型即可直接合成多个模态, 而 SkrGAN 和脑 MRI 合成 [42] 等方法中若要实现多模态合成则需要针对每个模态训练一个单独的模型。相比于 SkrGAN 等, 本研究的方法合成多模态时的优势具体包括: 一次性训练成倍减少了训练代价; 多模态生成模型可以在全部模态数据上进行训练学习, 模型的

训练样本量比SkrGAN等方法的模型的单模态训练样本量成倍增加，这使得模型在小数据集上表现更好；多模态的协同训练使得模型在接收同一个结构特征图输入时合成的多模态医学影像之间具有很好的配准度，其本身就比独立训练的多模型方法接收同一个输入合成的多模态影像配准度更高；多模态生成模型可以从同一输入同步合成多个模态，更便于对合成的多模态影像之间配准监督和病灶生成监督，多模型方法中实现同样的同步监督则需要成倍的内存开销和计算开销。总的来说，合成多模态影像时，本研究基于ACGAN架构的单模型方案相比于SkrGAN和脑MRI合成 [42]等多模型方案具有更多的优势。

5.3 无监督多模态影像的配准

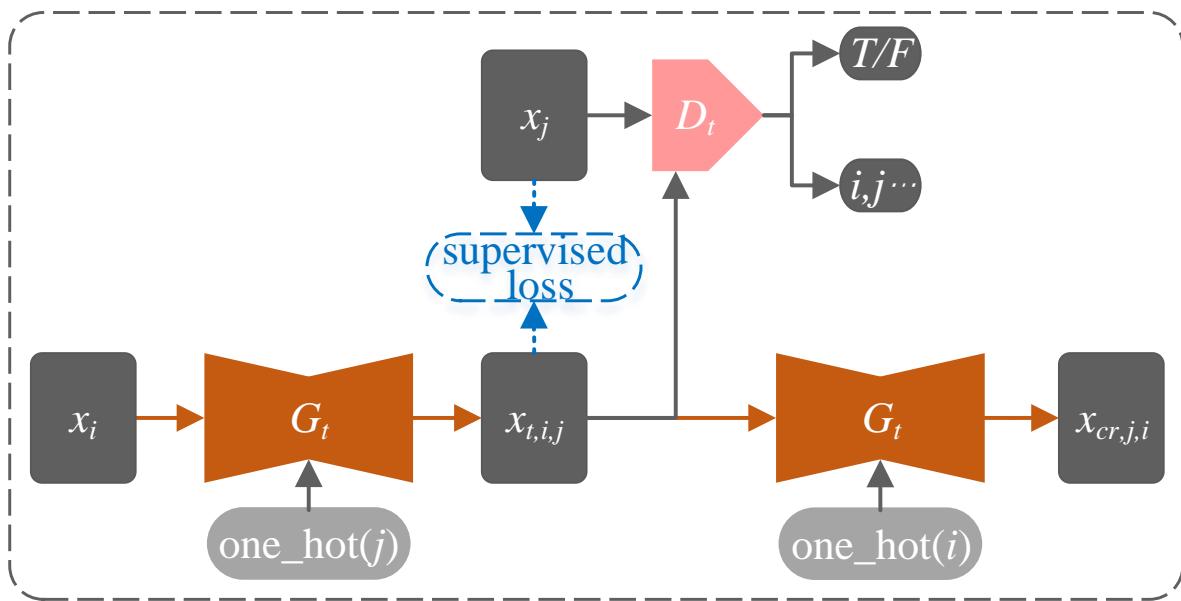


图 5.5: 模态转换器训练过程。其中内部虚线表示非必需过程。

当需要确保合成的不同模态的影像精确配准时，可以通过模态转换器对合成多模态影像进行模态转换一致性约束。模态转换器通过真实的多模态影像数据预先训练完成，如图5.5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）[14]。 D_t 与在VGG11模型结构的基础上进行了调整适配的 D 结构相同，详细结构如图5.3。 G_t 与

在U-net的模型结构的基础上进行了调整适配的G结构相同，详细结构如图5.4。各自的损失函数如下：

- 模态循环一致性损失

$$\mathcal{L}_{cycle}(G_t) = \mathbb{E}_x [\sum_{j=0} \sum_{i=0, i \neq j} (\|x_i - x_{cr,j,i}\|_2^2)], \quad (5.7)$$

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

- 模态转换合成影像对抗性训练损失

$$\begin{aligned} \mathcal{L}_{d,t}(D_t) = \mathbb{E}_{x,s,l} [\sum_{i=0} (\|d(x_i) - 1\|_2^2 + \|d(x_{t,i,j})\|_2^2 + \\ \|c(x_i) - i\|_2^2 + \|c(x_{t,i,j}) - i\|_2^2)], \end{aligned} \quad (5.8)$$

$$\mathcal{L}_{g,t}(G_t) = \mathbb{E}_{s,l} [\sum_{i=0} (\|d(x_{t,i,j}) - 1\|_2^2 + \|c(x_{t,i,j}) - i\|_2^2)]. \quad (5.9)$$

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

上述模态转换器的训练是无监督的，对多模态数据是否配准没有要求，因此基于模态转换的配准方法既可以采用未配准训练数据，同样也能接受本身已配准的多模态数据。对于本身已配准数据，可以在模态转换器的训练中增加如下的配准监督损失来加快模态转换器的训练过程：

$$\mathcal{L}_{sv}(G_t) = \mathbb{E}_x [\sum_{j=0} \sum_{i=0, i \neq j} (\|x_i - x_j\|_2^2)]. \quad (5.10)$$

5.4 病灶信息的添加和合成

当需要在合成的多模态影像中添加指定的病灶信息时，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。为了增加病灶的多样性，训练和使用时可以根据标签的类型对原始真实病灶标签集进行随机缩放、旋转、平移、翻转等常用的图像数据增强操作或直接随机生成，得到比真实病灶标签更加随机多样的随机病灶标签集 \mathbf{L} 。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，因此可以使用目标器官的掩膜 m 来过滤病灶标签。

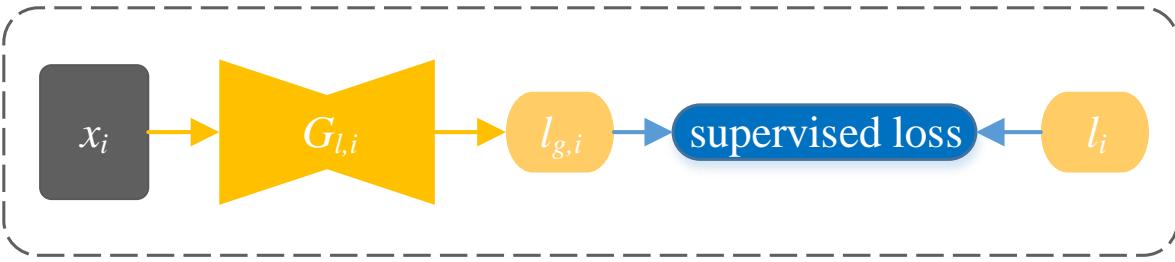


图 5.6: 病灶处理器训练过程示例。

算法 4 病灶标签选取方法

输入: 当前输入的合成结构特征图对应的掩膜 m_g ;随机病灶标签集 \mathbf{L} .

输出: 选取的病灶标签 l .

```

1: do
2:    $index = \text{randint}(0, \mathbf{L}.length)$ 
3:    $l = \mathbf{L}[index]$ 
4:    $flag = \text{reduce\_sum}(l \times m_g)$ 
5: while  $flag == 0$ 

```

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 。对应的筛选算法如算法 4 所示。

算法 4 中， $\mathbf{L}.length$ 表示随机病灶标签集 \mathbf{L} 中标签的数量， $\text{randint}(\cdot)$ 表示在指定范围内生成随机整数的函数， $\text{reduce_sum}(\cdot)$ 表示对输入矩阵的所有元素的求和函数。随机选取出的病灶标签 l 与结构特征图通过通道堆叠连接函数 $\text{concat}(\cdot)$ 融合，即：

$$s'' = \text{concat}([s', l]), \quad (5.11)$$

这样得到融合图既包含目标部位的基本解剖信息又包含指定的病灶信息，最后再以此融合图为输入合成医学影像。

如图 5.6 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，使用一个病灶处理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。病灶处理器用真实的影像和标签数据预先训练完成。除非病灶处理任务本身要求多模态输入等特殊需求外，每个模态应该均由独立的病灶处理器 $G_{l,i}$ 来指导合成而不应共用一个病灶处理器，以避免不同模态上病灶合成相互影响而出现一些模态合成

效果好另一些模态效果差的偏重现象。

在分割类病灶处理任务训练过程中，一个典型的均方根误差损失为：

$$\mathcal{L}_{seg}(G_{l,i}) = \mathbb{E}_{l,x}[\|l_i - l_{g,i}\|_2^2], \quad (5.12)$$

其中 l_i 是 x_i 对应的真实分割标签， $l_{g,i} = G_{l,i}(x_i)$ 是模型预测输出的分割标签。

在以softmax(\cdot)为最后输出层的分类病灶处理任务训练过程中，一个典型的交叉熵损失为：

$$\mathcal{L}_{class}(G_l) = \mathbb{E}_y[-\sum_c^n y_c \times \log(p_c)], \quad (5.13)$$

其中 $y_c \in \{0, 1\}$ 是医学影像 x 属于类别 c 的真实概率， p_c 为模型预测输出的类别为 c 的概率， n 是总共的类别数量。

在检测类病灶处理任务训练过程中，一个SSD检测模型 [90]的典型检测损失为：

$$\begin{aligned} \mathcal{L}_{detec}(G_l) = & \mathbb{E}_{r,l,g} \left[\sum_{i \in P}^n \sum_{m \in \{x_c, y_c, w, h\}} r_{ij}^k \times \text{smooth_l1}(l_i^m - \hat{g}_j^m) \right] - \\ & \alpha \mathbb{E}_{r,c} \left[\sum_{i \in P}^n r_{ij}^k \times \log(\hat{c}_i^k) + \sum_{i \in N} \log(\hat{c}_i^0) \right], \end{aligned} \quad (5.14)$$

其中， $r_{ij}^k \in \{0, 1\}$ 表示第 i 个默认框与类别 k 的第 j 个真实框的匹配结果， l 是预测出的检测框相对于默认框 d 的偏移量， \hat{g} 是真实框 g 相对于默认框 d 的偏移量， P 表示预测出的与真实框交并比（IoU, Intersection-over-Union）大等于设定阈值的预测框正样本集合， N 表示预测出的与真实框IoU低于设定阈值的预测框负样本集合， n 表示预测框正样本集合中预测框的数量， α 表示权重系数， c_i^k 表示第 i 个默认框对应的类别 k 的类别置信度， x_c, y_c, w, h 表示病灶真实检测框的中心点的横纵坐标和宽高，smooth_l1(\cdot)表示的Smooth L1函数和各变量间其他函数关系如下：

$$\text{smooth_l1}(a) = \begin{cases} 0.5a^2, & |a| < 1, \\ |a| - 0.5, & |a| \geq 1, \end{cases} \quad (5.15)$$

$$\begin{cases} \hat{g}_j^{x_c} = (g_j^{x_c} - d_i^{x_c})/d_i^w, \\ \hat{g}_j^{y_c} = (g_j^{y_c} - d_i^{y_c})/d_i^h, \\ \hat{g}_j^w = \log(\frac{g_j^w}{d_i^w}), \\ \hat{g}_j^h = \log(\frac{g_j^h}{d_i^h}), \\ \hat{c}_i^k = \frac{\exp(c_i^k)}{\sum_k \exp(c_i^k)}. \end{cases} \quad (5.16)$$

在实际应用时，损失函数应根据选取的病灶处理器的具体模型结构和病灶任务的实际需要设计，例如选用U-net作为病灶处理器执行分割任务时就可以使用U-net原始的加权交叉熵损失函数 [24]。本文的方案中，采用U-net、VGG11、SSD等模型作为病灶处理器时均可基本沿用模型原有的网络结构和损失函数方案，除对输入输出结果尺寸和模型参数量的调整外无需过多的调整适配。上述模型的详细结构见第三章的介绍。

5.5 合成数据集的构建

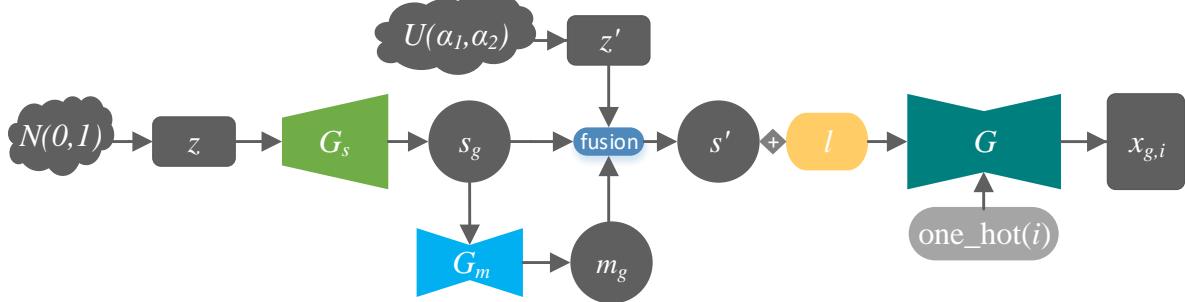


图 5.7: 合成数据集构建过程示意图。

如图5.7所示为在上述各个训练阶段完成之后的合成数据集的构建过程，具体步骤为：

- **合成结构特征图** 第一步，采用训练好的结构特征图生成器 G_s 从随机正态分布采样噪声 z 合成足量的结构特征图 s_g ；

- **合成结构特征图的掩膜** 第二步，采用训练好的掩膜生成器 G_m 从合成的结构特征图 s_g 生成对应的掩膜 m_g ；
- **过滤合成结构特征图** 第三步，采用第4.4节中的算法 3根据合成的结构特征图 s_g 及其生成器合成掩膜 m_g 和算法生成掩膜 m'_g ，来过滤掉少数不完整的结构特征图，此步骤应根据结构特征图情况选用；
- **融合结构特征图和随机噪声** 第四步，从均匀分布随机采样噪声 z' 通过第4.5节中公式 4.8与结构特征图 s_g 融合；
- **融合随机病灶标签** 第五步，采用第5.4节中的算法 4从构造的随机病灶标签集 L 中随机选择病灶标签，再通过公式 5.11融合结构特征图和选取的随机病灶标签，此步骤为可选步骤；
- **输入融合图和独热条件矩阵** 第六步，以上一步得到的融合图 s'' 为最初输入，然后根据需要设定指示生成模态的独热条件矩阵 $\text{one_hot}(i)$ ，默认直接对每一个融合图输入都循环输入全部独热条件矩阵以合成全部模态，但在单模态合成中无独热条件矩阵输入过程；
- **获得合成数据集** 第七步，循环执行上述所有步骤获得足量合成多模态医学影像 $x_{g,i}$ ，同时保存合成过程中产出的结构特征图 s_g 、掩膜 m_g 和病灶标签 l ，这样得到的全部数据就构成了需要的合成数据集。

由于可合成的结构特征图和医学影像是无数量限制的，在前述章节中，本研究对合成的结构特征图进行了过滤，在构建合成数据集时，除了上述通过掩膜对病灶标签进行过滤的步骤之外，实际应用中还可以对合成影像进行过滤，过滤时可以采用鉴别器提供的真假鉴别置信度或合成影像与真实影像数据集的平均相似度来对GAN的合成结果过滤，同时可通过病灶处理器提供的病灶标签还原程度来对包含合成病灶的合成影像过滤，最终得到需要的逼真程度高、病灶合成良好的由合成结构特征图、合成掩膜、病灶标签和合成多模态医学影像组成的合成数据集。

第六章 合成医学影像的性能评估

前两章详细介绍了本文解决方案中前三个阶段的方法和结合流程，本章将在多个数据集上对本文方法的评估实验进行介绍。本章将先对使用的数据集进行介绍，然后详细介绍本研究在各个数据集上的实验设置，包括训练设置和评估指标，之后将对本研究在各个数据集上的实验方案设计、实验结果进行介绍和分析并将本研究的合成影像与当前最优秀的合成影像进行量化对比与视觉对比评估，其中实验包括消融实验、合成医学影像整体质量的量化评估实验、合成医学影像的病灶有效性验证实验、合成医学影像的配准有效性验证实验和合成医学影像在智能医学影像处理任务中的可用性验证实验。本章评估实验中各项智能医学影像病灶处理任务中的病灶处理器与上一章中合成医学影像时采用的病灶处理器一致，均为在第三章中介绍的基础方法和模型上的简单适应性调整。

6.1 数据集

不同模态的医学影像从成像原理到视觉呈现都有诸多差异，一项通用的医学影像合成方案应能在常见的多种模态中表现良好才具有较好的鲁棒性。在带病灶的医学影像合成中，对病灶的评估也应考虑到常见的多类智能医学影像病灶处理任务。因此，本研究在六个公开数据集中对提出的方案进行了充分的实验验证，实验涵盖了RGB彩色照片、X-ray、MRI、CT等常见的医学影像成像模态，涵盖了脑部、视网膜、肺部截面、胸部等多个病变部位，涵盖了肿瘤分割、血管分割、肺炎分类、肺结节检测等多种智能医学影像病灶处理任务。这六个公开数据集的详细情况和数据预处理方式如下：

- **BRATS2015数据集³** [95]公开数据集BRATS2015包含T1、T2、T1c、Flair四个配准模态的脑部3D MRI。训练集每个模态有274张大小为 $155 \times 240 \times 240$ 的图及对应的274张肿瘤分割标签图。对训练集按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，对每个图像都进行标准化（Normalized），裁剪空白边框至尺寸为 184×144 。

³<https://www.smir.ch/BRATS/Start2015>

- **Kaggle Chest X-ray**数据集⁴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-ray灰度图，图像尺寸从 384×127 到 2772×2304 不等。在数据预处理时，对每个图像都进行标准化，尺寸缩放为 512×512 。

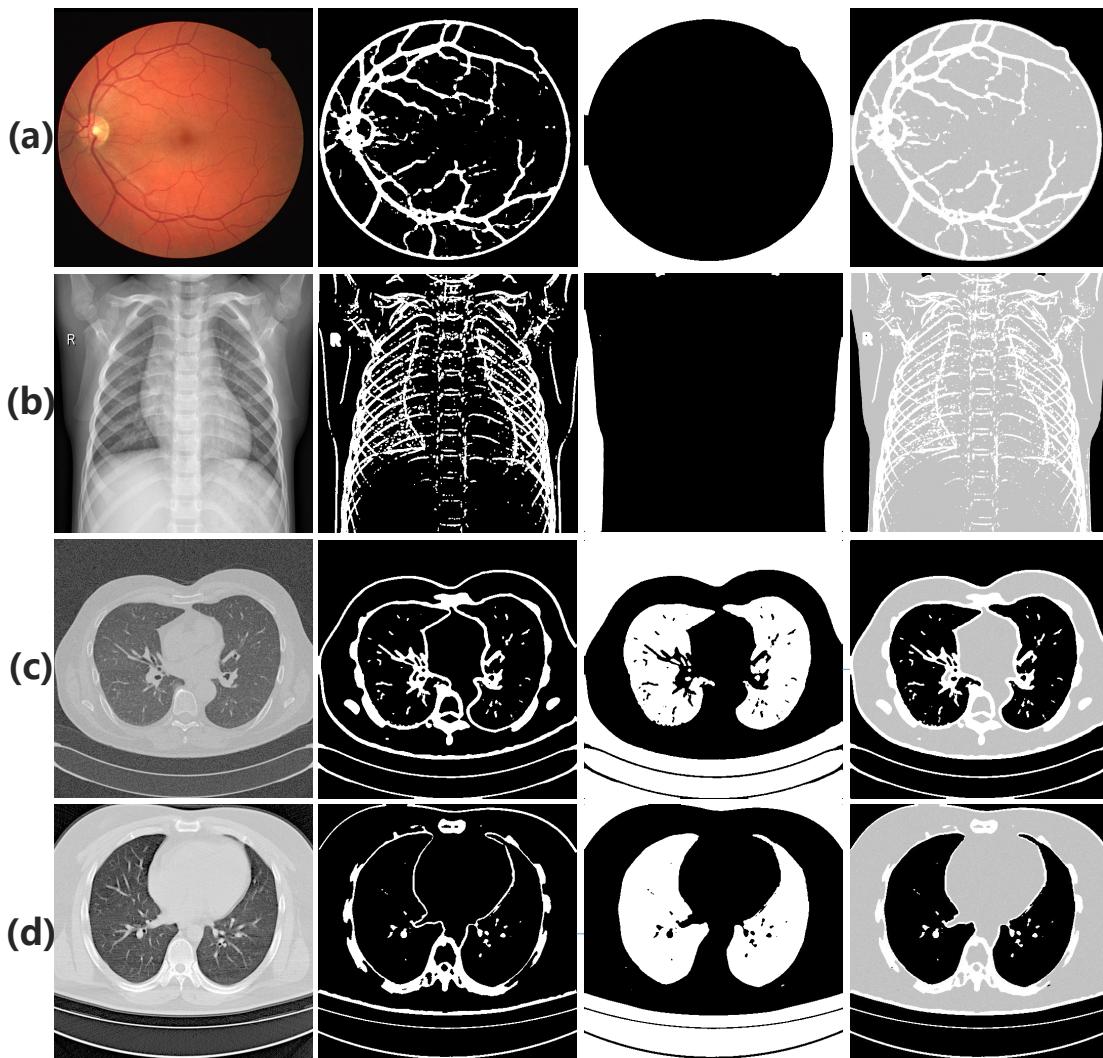


图 6.1: 各数据集中的医学影像和提取的结构特征图。(a) DRIVE视网膜图像数据集。(b) Kaggle Chest X-ray数据集。(c) Kaggle Lung CT数据集。(d) 天池全球数据智能大赛(2019)数据集。以上子图中从左到右依次为原图、结构特征图、掩膜、融合了随机噪声的结构特征图。

⁴<https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>

- **Kaggle Lung CT数据集⁵**该数据集包含267个胸部至腹部横向截面的2D CT扫描图，尺寸为 512×512 。在数据预处理时，对每个图像都进行标准化。
- **DRIVE视网膜图像数据集⁶**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和对应的视网膜血管分割标签，尺寸均为 565×584 ，在数据预处理时，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⁷**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，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（TC Lung CT）⁸**该数据集包含肺部3D CT扫描图共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，根据标注信息切取有标注的切片及其前后切片组成的3通道图，再对每个图像都进行标准化，尺寸缩放为 $512 \times 512 \times 3$ 。新的标签为中间通道对应的原始标签中病灶的位置标签。将检测标签可视化为检测框图后，对检测框进行填充，可得到用于与结构特征图融合输入的肺部病灶检测标签图。

图4.3中展示了BRATS2015数据集中脑MRI的四个模态、提取出的掩膜和结构特征图。图6.1中展示了DRIVE视网膜图像数据集、Kaggle Chest X-ray数据集、Kaggle Lung CT数据集和天池全球数据智能大赛(2019)数据集中的医学影像、提取出的掩膜和结构特征图以及融合噪声后的结构特征图。

6.2 实验设置

6.2.1 训练设置

本研究中实验采用的主要硬件环境为一个NVIDIA Tesla V100 GPU集群，软件环

⁵<https://www.kaggle.com/kmader/finding-lungs-in-ct-data/data/>

⁶<http://www.isi.uu.nl/Research/Databases/DRIVE/>

⁷<https://projects.ics.forth.gr/cvrl/fire/>

⁸<https://tianchi.aliyun.com/competition/entrance/231724/information>

境为Python3.6、CUDA8.0和TensorFlow1.13.2⁹。医学影像数据的读取和处理采用医学影像处理专业工具包SimpleITK¹⁰和矩阵运算工具包numpy¹¹。每个实验均经过在训练数据集上超过200个epoch（一个epoch为一次训练集的完全遍历）的充分的迭代训练并取得收敛。学习率为1e-5，无权重衰减，使用beta1为0.5的Adam优化器 [96]，批处理大小（Batch Size）为4（采用V100集群的多卡同步并行训练中，每张GPU上处理一张图）。在采用 G 进行生成训练和采用 G_t 进行转换训练时，考虑到输入输出的相似性，模型卷积层的参数初始化采用均值模糊卷积核的参数，即卷积核每个参数的初始值都设置为卷积核参数数量的倒数。其他模型采用均值为0，方差为0.1的正态分布随机采样作为初始参数。本研究中实验的评估结果均为2D图像上所有模态结果的平均值，每个实验的结果都为经过了四次成功收敛训练保留的最佳结果。

在基于不同的训练数据和需求时，本文中提出的方案可以对其中的一些方法进行选择性应用。本研究在多个数据集上有多个合成任务，各自具体的训练设置如下：

- **多模态脑肿瘤MRI的合成** 以BRATS2015训练集作为训练数据，以其肿瘤分割标签为输入的病灶标签，合成T1、T2、T1c、Flair四个MRI模态，使用无监督训练的模态转换器提供模态配准损失，使用一个U-net [24]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，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（如图4.3所示），在合成多模态MRI训练时，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，采用在真实影像及其结构特征图上进行的预训练模型。
- **眼底视网膜图像的合成** 以DRIVE视网膜图像训练集集和整个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模型。在视网膜血管分割实验中，通过缩放掩膜去除结构特征图的外层圆形轮廓，只保留圆内的血管线路，并以此血管线路图作为以该结构特征图合成的视网膜图像的血管分割粗标签。

⁹<https://www.tensorflow.org/>

¹⁰<http://www.simpleitk.org>

¹¹<https://numpy.org/>

- **肺部CT的合成** 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模型。
- **胸部X-ray的合成** 以Kaggle Chest X-ray训练集作为训练数据，结构特征图合成训练中弃用数据集中两类肺炎影像而只选用正常影像以确保合成的结构特征图无病变信息，并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VGG11 [80]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影像，采用在真实影像及其结构特征图上进行的预训练模型。
- **肺部低剂量CT的合成** 以天池全球数据智能大赛(2019)训练集作为训练数据，将肺部病灶检测标签转换为与结构特征图同尺寸的检测标签图（检测框填充为1像素，背景为0像素）后作为输入的病灶标签，以SSD [90]作为病灶检测器提供病灶生成指导损失，无模态配准损失，根据输入的肺部病灶检测标签图合成具有对应肺部病灶的单模态肺部低剂量CT，采用在真实影像及其结构特征图上进行的预训练模型。

6.2.2 评估指标

本研究以SkrGAN [74]中实现的当前最好质量的合成结果作为一个基准对比。本研究使用多尺度结构相似性（MS-SSIM， Multiscale Structural Similarity Index Measure），和Freshet Inception距离（FID， Freshet Inception Distance） [97]来评估合成医学影像的性能。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。本研究使用Dice Score [98]和均方根误差（MSE）来评估分割结果，前者指标越高表示性能越好，后者指标越低表示性能越好；使用敏感度（Sensitivity）、正确率（Accuracy）和ROC曲线下面积（AUC， Area Under the ROC Curve）来评估血管分割结果，三项指标均越高表示性能越好；使用正确率（Accuracy）来评估本研究的分类结果，指标越高表示性能越好；使用IoU 阈值为0.5的平均精度均值（mAP， Mean Average Precision） [99]来评估检测结果，指标越高表示

性能越好；使用峰值信噪比（PSNR，Peak Signal-to-Noise Ratio）、结构相似性（SSIM，Structural Similarity Index Measure）[100]来评估模态转换结果与真实模态的相似性，两项指标越高表示性能越好。

6.3 多模态医学影像合成方法的消融实验

表 6.1: BRATS2015上消融实验的实验设置

实验	输入结构特征图	输入随机噪声	模态配准损失 \mathcal{L}_{reg}	添加病灶标签和病灶生成损失 \mathcal{L}_{les}	使用掩膜过滤筛选病灶标签
A	✗	✓	✗	✗	✗
B	✓	✗	✗	✗	✗
C	✓	✓	✗	✗	✗
D	✓	✓	✓	✗	✗
E	✓	✓	✓	✓	✗
F	✓	✓	✓	✓	✓

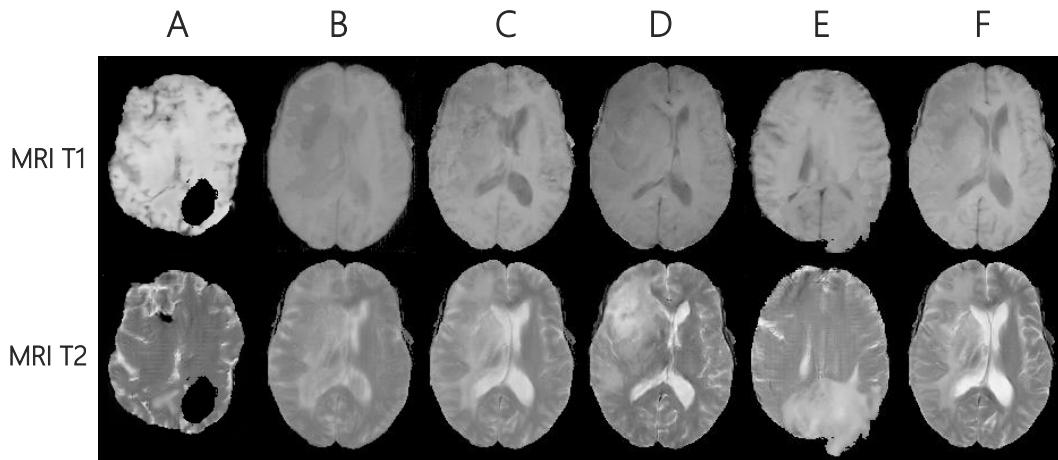


图 6.2: 消融实验的合成图像。B、C、E、F均采用了相同的结构特征图作为输入。

如表6.1所示，在BRATS2015数据集上，对合成多模态医学影像的方法进行的消融对比实验，用于验证在多模态合成阶段的各项方法的作用。首先，实验A以相同尺寸的随机噪声代替实验B中结构特征图作为MRI生成训练输入以验证添加结构特征

图的作用。实验**B**与**C**在相同的训练条件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对提高模型泛化能力并加快训练收敛的作用。实验**C**与**D**通过有无模态配准损失来验证该损失对边缘配准的矫正作用。实验**D**与**F**通过是否添加过滤后的病灶标签和有无病灶生成指导损失来验证这对病灶合成的指导作用。最后实验**E**与**F**通过是否使用掩膜对输入标签过滤来验证掩膜过滤的作用。

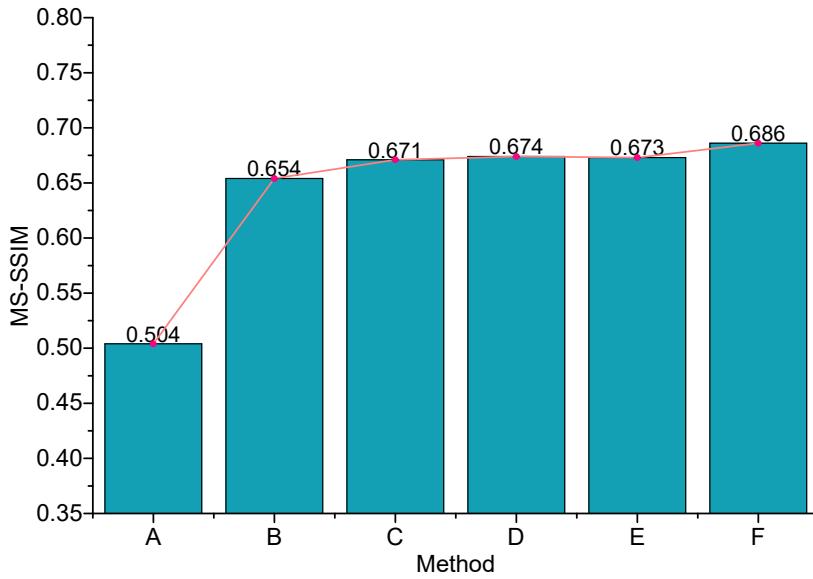


图 6.3: 脑肿瘤MRI数据集上消融实验的量化评估结果。

图6.2显示了消融对比实验中的合成图像的示例，其中模型**A**用随机噪声替换结构特征图且也因此无自监督预训练过程，由于没有结构特征图在脑轮廓上的约束，生成的图像符合MRI的特征，但与大脑的结构特征符合程度较低。模型**B**以结构特征图为输入但没有融合随机噪声，在与其他实验相同的训练epoch时合成质量较差，最终生成的图像的配准效果相对较差且未能合成明显的病灶信息。模型**C**没有模态配准损失，生成的图像的配准效果较弱，尤其在边缘部分有多处未能很好对齐。模型**D**没有病灶生成指导损失，可以看出生成的图像中的病变内容较为散乱随意和过度夸张，与输入的病灶标签不够吻合。模型**E**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**F**采用本文中完整的合成方案，合成了与输入病灶标签更吻合的病灶信息，配准度更高，合成图像的整体质量也更高。可以直观对比看出，本文合成方案的各项改进均对合成影像的真实程度都有提升作用，但在添加病灶标签且

使用病灶生成指导损失时，若不采用掩膜对病灶标签进行筛选，可能会生成如图中模型**E**类似的结构不合理的影像。

图6.3为消融实验量化评估结果的柱状图对比情况，不难看出，本文合成方案中的每一项方法都起到了针对性的提升效果。其中，采用结构特征图作为输入对合成图像的轮廓结构有直接且强烈的影响，因此对评估结果的提升最为突出。结构特征图融合随机噪声降低了整体合成难度而使得局部纹理细节的合成更丰富逼真、过滤病灶标签后添加病灶生成指导损失指导合成了更真实合理的病灶，因此这两项改进也对评估结果有明显提升。图标识数据中值得注意的是，不对输入的病灶标签进行过滤，合成的影像中不止会存在病灶结构不合理的影像，与真实数据集进行相似度评估的结果也会因此难以提升甚至下降。

6.4 合成医学影像数据与真实医学影像数据的相似性评估

如表6.2和表6.3所示，本研究对提出的方法在各个数据集上合成的医学影像数据的整体质量与真实医学影像数据进行了直接地相似性量化评估，并将结果与当前最先进的方法进行对比。表6.2中，与SkrGAN中的结果对比表明，在Kaggle Chest X-ray和Kaggle Lung CT两个数据集上，本文提出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 6.2: 不同数据集上合成图像质量的量化评估（一）

数据集	评估指标	Ours	SkrGAN [74]	DCGAN [51, 74]	ACGAN [20, 74]	WGAN [74, 101]	PGGAN [74, 97]
Kaggle 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
Kaggle 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6.3中，由于SkrGAN中两份数据集的不可获取或复现，本研究在DRIVE+FIRE彩色视网膜图像（Color Fundus）和BRATS2015两个与SkrGAN中所用的数据集类似的数据集上，对SkrGAN方法进行了复现，复现结果与表6.2中两个数据集上的结果指向的结论是一致的。BRATS2015数据集上，本文的方案可直接进行多模态合成，其余方案

表 6.3: 不同数据集上合成图像质量的量化评估 (二)

数据集	评估指标	Ours ⁺	Ours	SkrGAN*	GAN [#]
DRIVE+FIRE	MS-SSIM ↑	-	0.607	0.584	0.392
	FID ↓	-	30.13	37.91	227.41
BRATS2015 MRI	MS-SSIM ↑	0.692	0.686	0.653	0.504
	FID ↓	20.15	21.87	28.76	124.53
TC Lung CT	MS-SSIM ↑	-	0.676	0.667	0.543
	FID ↓	-	27.40	29.81	113.65

+ 表示合成肿瘤的评估，即根据病灶标签对合成数据集影像和原数据集影像中切割出的肿瘤的评估。

表示基础GAN的评估，即在DCGAN [51]的基础上将生成器加深为U-net，将鉴别器加深为VGG11。

* 表示复现的SkrGAN的评估结果。由于SkrGAN [74]中未给出草图详细算法或开源代码和鉴别器网络结构，合成多模态影像时，采用本文的算法提取的真实影像的结构特征图的二值反色图（0与1像素值调转）作为训练输入，以U-net为生成器、以VGG11为鉴别器、无病灶生成指导损失、无模态配准损失、无自监督预训练、采用自监督和对抗性混合损失。

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BRATS2015数据集上，添加的肿瘤分割标签作为病灶生成指导标签，评估时，根据病灶标签二值化后的掩膜对合成数据集中的影像与原数据集中的影像进行肿瘤切割，再对切割出的肿瘤单独评估，并与完整影像的评估结果对比。结果表明，采用本文方法得到的合成肿瘤与真实肿瘤具有更高的相似度，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6.3中彩色视网膜数据集上，SkrGAN中所用的自建数据集的数据量为本研究所用的数据集数据量的22倍多。本文方法以融合了噪声的结构特征图为输入、复现SkrGAN时采用本文算法提取的真实图像的结构特征图的二值反色图（0与1像素值调转）为输入，在采用同样的生成器和鉴别器、无额外的病灶生成指导损失和模态配准损失的情况下训练。从表中合成视网膜的评估结果来看，本文的方法取得了比复现的SkrGAN更好的评估结果，且质量高于基础GAN从随机噪声合成的图像。表6.3中天池肺部CT数据集上，本研究采用肺结节检测标签转化的二值标签图作为病灶标签输入，由于肺结节的结构十分细微且数据集样本量更充分，两者的评估结果十分接近，得益于输入时融合的随机噪声带来的对模型泛化能力的提升，本文方法的结果略高于

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的合成图像质量高于采用随机噪声为输入的基础GAN方法，相同条件下本文方法的合成图像整体质量优于当前最好的SkrGAN方法。

综合表6.2和表6.3的结果来看，采用结构特征图为输入的方法的合成图像的质量比采用随机噪声为输入的方法的合成图像的质量要好很多。训练中对结构特征图融合噪声处理相较于不处理或二值反色处理，得到的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。相较于原始随机噪声输入和SkrGAN的草图输入改进，本文方法采用的结构特征图输入，结合自监督预训练、病灶生成指导损失等多项改进使得最终合成的医学影像质量更高，与真实图像更接近。

6.5 合成医学影像的病灶有效性评估

6.5.1 脑肿瘤MRI合成

本实验在BRATS2015数据集上，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个模态的脑肿瘤MRI，其中每个模态采用一个训练好的采用U-net为模型的肿瘤分割器作为病灶处理器，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肿瘤病灶信息的合成，并通过模态转换器确保了合成的模态之间的配准。

表 6.4: 肿瘤分割器对不同测试数据的分割结果

训练数据	测试数据	MSE ↓	Dice Score ↑
真实	真实	0.027	0.915
真实	合成	0.098	0.838

6.5.2 对合成脑肿瘤MRI的肿瘤区域分割检验

如表6.4所示，本研究在BRATS2015训练数据集上训练了4个模态的肿瘤病灶分割器，并在BRATS2015测试数据集上对其进行了测试。然后，使用训练好的分割器对合成数据进行分割。从分割结果来看，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较高的相似度，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组织。由此看出，合成病灶是有效的，且能对肿瘤分割器产生实质影响。

图6.4展示了采用在BRATS2015训练集上训练并在测试集上表现优秀的分割器对合成的MRI的分割示例，不难看出，肿瘤分割器能够对合成的肿瘤进行识别和分割且分割出的肿瘤标签与输入的真实标签基本一致，这直观地表明了合成的MRI中的确根据输入的肿瘤标签生成了对应的肿瘤信息。

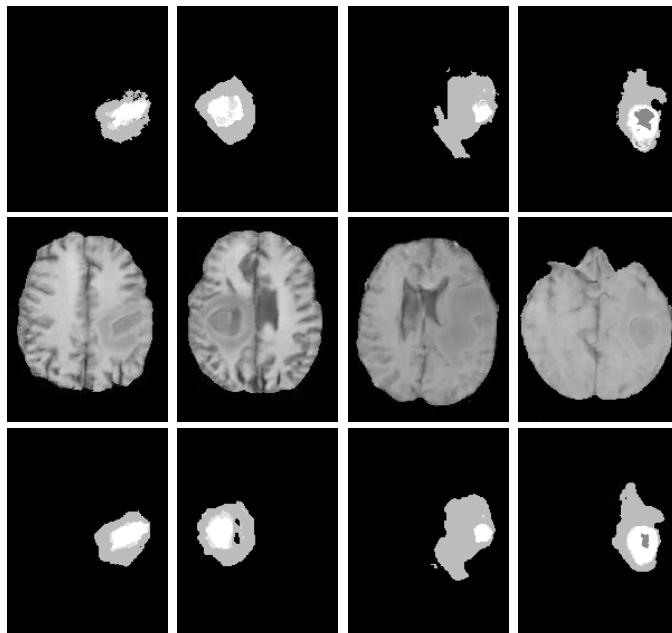


图 6.4: 脑肿瘤分割器对合成MRI的分割结果。从上至下三行依次为输入的肿瘤分割标签、合成的脑肿瘤MRI的T1模态、采用在真实数据集上训练并表现优秀的分割器对第二行合成MRI的分割结果。

6.6 合成多模态医学影像的配准有效性评估

本研究先在BRATS2015多模态数据集上合成了多模态脑MRI，合成训练时提供模态配准损失的模态转换器采用无监督方式训练，再将合成的多模态配准脑MRI用于新的模态转换器的有监督多模态转换训练，最后再在真实数据上进行模态转换测试，以验证合成多模态影像的配准有效性。由于当前最好的模态转换方法pGAN [102]只能进行两个模态的互转，而本研究需要实现任意多个模态的互相转换，因此本文

在pGAN方法采用的CycleGAN结构基础上进行了进一步调整和改进，得到了本文采用的模态转换方法。如表6.5（一）中结果所示，本文可实现任意数量模态转换的方法的转换效果与pGAN方法的转换效果十分接近。如表6.5（二）所示实验设置，本研究通过分别采用真实数据、合成数据和真实与合成混合的数据训练第5.3节中的模态转换器 G_t 。参照表中其他方法采用的有监督训练设置，各项转换实验中本研究模态转换方法采用了公式 5.10中的监督损失 $\mathcal{L}_{sv}(G_t)$ 。最后，在不同成分的数据上训练好的模型与当前最好的模态转换方法进行对比，以验证本研究的合成的多模态配准数据在对配准精度有极高要求的任务中的性能，实验结果如表6.5。

表 6.5: BRATS2015数据集上T1与T2模态互转实验的相似性评估结果.

(一) 不同模态转换方法采用真实数据训练后的转换效果评估.

	T1→T2	T2→T1	T1→T2	T2→T1
	PSNR ↑	PSNR ↑	SSIM ↑	SSIM ↑
REPLICA [102, 103]	24.64	24.49	0.924	0.917
Multimodal [72, 102]	25.09	23.78	0.939	0.935
pGAN [102]	27.19	25.80	0.946	0.940
本文转换方法	27.06	26.14	0.939	0.937

(二) 本文模态转换方法采用合成配准多模态脑MRI训练后的转换效果评估.

训练数据	测试	T1→T2	T2→T1	T1→T2	T2→T1
	数据	PSNR ↑	PSNR ↑	SSIM ↑	SSIM ↑
真实	真实	27.06	26.14	0.939	0.937
合成×10	真实	23.21	23.04	0.909	0.902
真实+合成×2	真实	30.85	29.48	0.960	0.958

从实验结果看，同在真实数据上训练的第5.3节中的模态转换方法与当前最好的模态转换方法pGAN [102]效果相当，这表明本研究的模态转换方法是比较优秀的。然后，再在10倍真实数据量的合成多模态配准脑MRI数据上训练了本研究方法中的模态转换器，结果表明，训练出的模型在测试集上的效果已经较为接近在真实数据上训练的模型，这意味着合成多模态影像不仅整体质量与真实影像较为接近，也意味着合成的多模态影像之间具有较高的配准度。在进一步对真实数据和合成数据随机混合的训

练中，混合训练模型取得了比仅使用真实数据训练的模型更好的转换结果，合成数据的使用大大提升了模型的转换效果，起到了数据增强的作用。由此表明，本研究的模态配准方案是效果良好的，合成的多模态影像具有较高的配准度，能在对配准度有较高要求的医学影像的任务中作为增强数据使用。

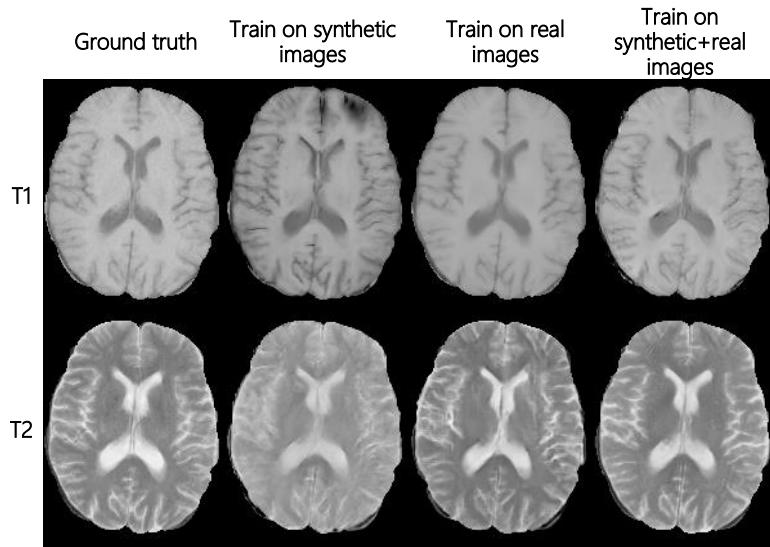


图 6.5: 脑MRI模态转换器的转换结果。第一行表示T1或从T2转换生成的T1模态，第二行表示T2或从T1转换生成的T2模态。

图6.5中三项转换结果与表6.5（二）中采用本文转换方法的三项实验对应，从转换结果看，只采用合成数据训练的转换器在测试集上的转换效果较差但也已经较为接近真实情况，使用真实数据训练的转换器效果比之更好，而混用合成数据和真实数据训练的转换器的转换效果是三项中表现最好的。这印证了上述合成多模态可以在要求配准的训练任务中起到增强作用的结论，验证了合成多模态具备比较良好的配准情况。

6.7 合成医学影像在智能医学影像病灶处理任务中的可用性评估

6.7.1 合成脑肿瘤MRI数据用于肿瘤区域分割任务

如表6.6所示，本研究将不同量的BRATS2015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分割训练，并在真实BRATS2015测试数据集上评估模型

的分割能力，分割模型为U-net。所有实验均经过了充分训练取得收敛。同时，实验中设置了三种数据混合模式：随机混合，先训练真实数据和先训练合成数据。

如表6.6所示，实验NO.3-NO.5的结果表明，训练中合成数据不能完全替代真实数据。NO.6-NO.8的结果表明，大量合成数据的预训练后再进行少量真实数据的微调训练取得的模型性能与完整真实数据的训练相似。NO.9-NO.11中，不同混合比例下合成数据的提升效果不同，但很难超过原本同样数量的真实数据的提升效果，整体上真实数据所占比例越大评估结果越接近完整真实数据集。NO.12-NO.17中，将不同数量的合成数据添加到真实数据中，结果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，但是当合成数据达到一定比例后继续增加时，提升效果会减弱。NO.18-NO.23中，比较了合成数据和通过一般数据增强方法生成的增强数据的增强效果。可以发现两者在数据量与增强效果之间的函数关系上相似但不相等。合成数据增强效果的上限和同等量合成数据产生的提升都高于一般增强数据。NO.24-NO.25与NO.15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，当合成数据用作增强数据与真实数据按照最合适的比例混合时，也可以实现最佳的增强效果。

通常，如果真实样本充足，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，则可以使用大量的合成数据进行预训练，然后再在少量真实数据进行微调训练，其结果可以与完整真实数据的结果相抗衡，此结论与其他脑MRI合成研究 [42]中的实验结论一致。此外，根据本文中实验的结果，合成的数据不建议单独用于模型的完全训练，也不建议用于补充训练以免破坏模型从真实数据上学习出的效果。

6.7.2 合成眼底视网膜数据用于视网膜血管分割任务

本研究使用DRIVE视网膜图像训练集+FIRE视网膜图像完整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。通过缩放的掩膜可以去除结构特征图的外层圆形轮廓，只保留圆形轮廓内的血管线路，获得的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分割粗标签。本研究以U-net为分割模型，使用DRIVE视网膜图像训练集+合成视网膜图像及其粗标签数据集，采用随机混合模式

表 6.6: BRATS2015数据集上的合成数据可用性验证实验.

NO.	真实数据量	合成数据量	增强数据量	混合方式	MSE ↓	Dice Score ↑
1	×1	0	0	-	0.027	0.915
2	×50%	0	0	-	0.041	0.900
3	0	×1	0	-	0.294	0.708
4	0	×2	0	-	0.253	0.736
5	0	×3	0	-	0.251	0.738
6	×10%	×1	0	\mathcal{S}	0.036	0.906
7	×10%	×2	0	\mathcal{S}	0.034	0.907
8	×10%	×3	0	\mathcal{S}	0.033	0.907
9	×20%	×80%	0	\mathcal{M}	0.038	0.904
10	×50%	×50%	0	\mathcal{M}	0.031	0.909
11	×80%	×20%	0	\mathcal{M}	0.028	0.914
12	×1	×20%	0	\mathcal{M}	0.025	0.921
13	×1	×50%	0	\mathcal{M}	0.020	0.939
14	×1	×80%	0	\mathcal{M}	0.021	0.937
15	×1	×1	0	\mathcal{M}	0.022	0.934
16	×1	×2	0	\mathcal{M}	0.023	0.931
17	×1	×3	0	\mathcal{M}	0.023	0.930
18	×1	0	×20%	\mathcal{M}	0.027	0.917
19	×1	0	×50%	\mathcal{M}	0.025	0.920
20	×1	0	×80%	\mathcal{M}	0.026	0.920
21	×1	0	×1	\mathcal{M}	0.026	0.919
22	×1	0	×2	\mathcal{M}	0.026	0.919
23	×1	0	×3	\mathcal{M}	0.027	0.918
24	×1	×1	0	\mathcal{R}	0.161	0.795
25	×1	×1	0	\mathcal{S}	0.020	0.940

\mathcal{M} : 随机混合 \mathcal{S} : 先训练合成数据 \mathcal{R} : 先训练真实数据

训练视网膜血管分割模型。再在DRIVE视网膜图像测试集上对训练好的模型分割测试，并与只使用DRIVE视网膜图像训练集的模型的分割结果对比，来验证合成眼底视网膜数据在视网膜血管分割任务中的可用性。由于本研究采用的数据集的数据量仅有SkrGAN自建视网膜数据集的二十分之一，因此尽管在结构特征图上本文的方法取得了提升，但最终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，本文方法的合成图像对分割Accuracy的提升与SkrGAN基本一致，在Sensitivity和AUC指标上略低。但如表6.7中SkrGAN*所示，当采用与本文方法相同的数据集对SkrGAN进行近似复现后对比时，本文的方法在各个指标上都表现更优。

表 6.7: 视网膜合成数据可用性验证实验

训练数据	测试数据	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 本文方法的合成图像	DRIVE测试集	0.8416	0.9518	0.9749

* 表示复现的SkrGAN的评估结果。由于SkrGAN [74]中未给出草图详细算法或开源代码和鉴别器网络结构，合成多模态影像时，采用本文的算法提取的真实影像的结构特征图的二值反色图（0与1像素值调转）作为训练输入，以U-net为生成器、以VGG11为鉴别器、无病灶生成指导损失、无模态配准损失、无自监督预训练、采用自监督和对抗性混合损失。

6.7.3 合成胸部X光线数据用于肺炎分类任务

本研究采用Kaggle Chest X-ray数据集合成了大量带有肺炎类别标签的胸部X-ray数据。合成时，由于只有单模态数据，因此无模态模态配准损失，输入的病灶标签为类别序号扩展成的与结构特征图同尺寸独热矩阵的肺炎类别标签图。再使用Kaggle Chest X-ray训练数据集+合成X-ray数据采用随机混合模式训练一个胸部X-ray肺炎分类模型，最后在Kaggle Chest X-ray测试集上对模型进行分类测试，并将该模型和只采

表 6.8: 胸部X-ray合成数据可用性验证实验

训练数据	测试数据	Accuracy ↑
Kaggle Chest X-ray训练集	Kaggle Chest X-ray测试集	0.804
Kaggle Chest X-ray训练集+2000张 合成图像	Kaggle Chest X-ray测试集	0.818

用Kaggle Chest X-ray数据集训练的模型的测试结果进行对比，分类模型采用VGG11。结果如表6.8所示，可见在分类任务中，本文方法在该数据集上的合成数据同样具有良好的可用性，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，因此分类任务中使用合成数据的提升效果相对有限。

6.7.4 合成肺部CT数据用于肺结节检测任务

表 6.9: 肺部CT合成数据可用性验证实验

训练数据	测试数据	mAP ↑
TC Lung CT训练集	TC Lung CT测试集	0.402
TC Lung CT训练集+20000张 合成图像	TC Lung CT测试集	0.410

本研究采用天池全球数据智能大赛(2019)数据集合成了大量带有肺结节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，因此无模态模态配准损失，输入的病灶标签为根据肺部病灶检测框坐标信息生成的对应的检测标签图，其中填充检测框得到的色块可以粗粒度的指示病灶合成的位置。本研究使用天池全球数据智能大赛(2019)训练数据集+合成肺部低剂量CT数据采用随机混合模式训练了一个肺部病灶检测模型，将该模型在天池测试数据上的测试结果与只采用天池全球数据智能大赛(2019)训练数据集训练的检测模型进行对比，检测模型采用SSD。结果如表6.9，这表明，在目标检测数据集上，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，但比分类标签指向性集中，因而在对应任务中，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。

6.8 合成影像的视觉评估

6.8.1 合成结构特征图多样性视觉评估

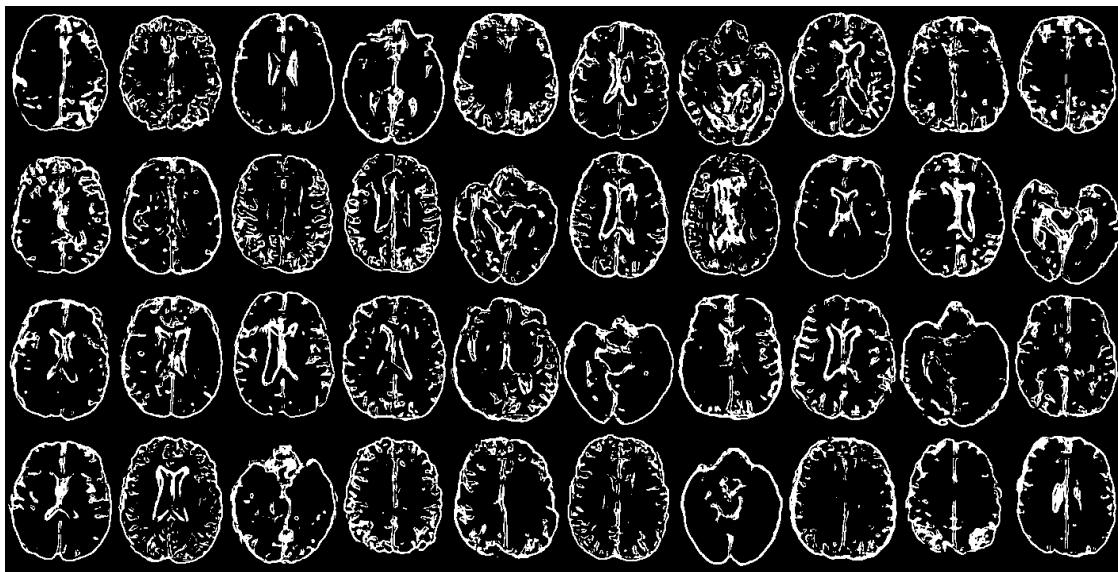


图 6.6: BRATS2015数据集上随机合成的结构特征图示例。

从正态分布随机采样得到如图6.6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。从图中可以看出，随机合成的结构特征图涵盖了脑部MRI以水平中线为轴的不同截面上的情况，这些样本在符合脑部生理结构的前提下在结构上更加千姿百态、丰富多彩，表现出了更为丰富的多样性。

6.8.2 结构特征图正态分布视觉评估

从正态分布进行顺序渐变采样，然后解码采样结果，获得如图6.7所示的相邻渐变的结构特征图分布。可以看出，图中不仅合成的结构特征图都符合脑部生理结构特征，而且无论是横向还是纵向又或者对角线上的直线上，各个结构特征图都在结构上依次呈现出了渐变的效果。

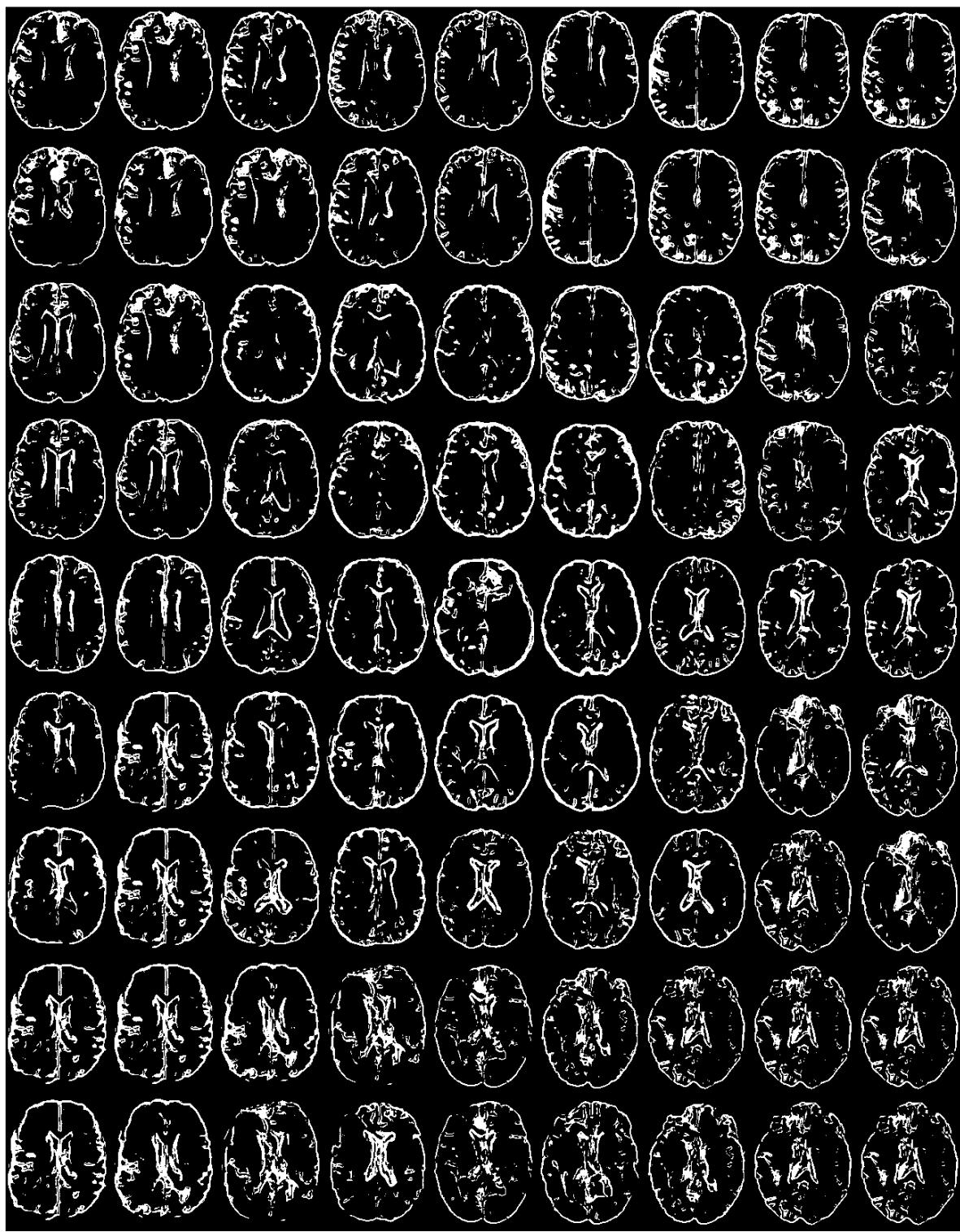


图 6.7: BRATS2015数据集上从正态分布顺序渐变采样合成的结构特征图示例。

6.8.3 合成医学影像视觉评估

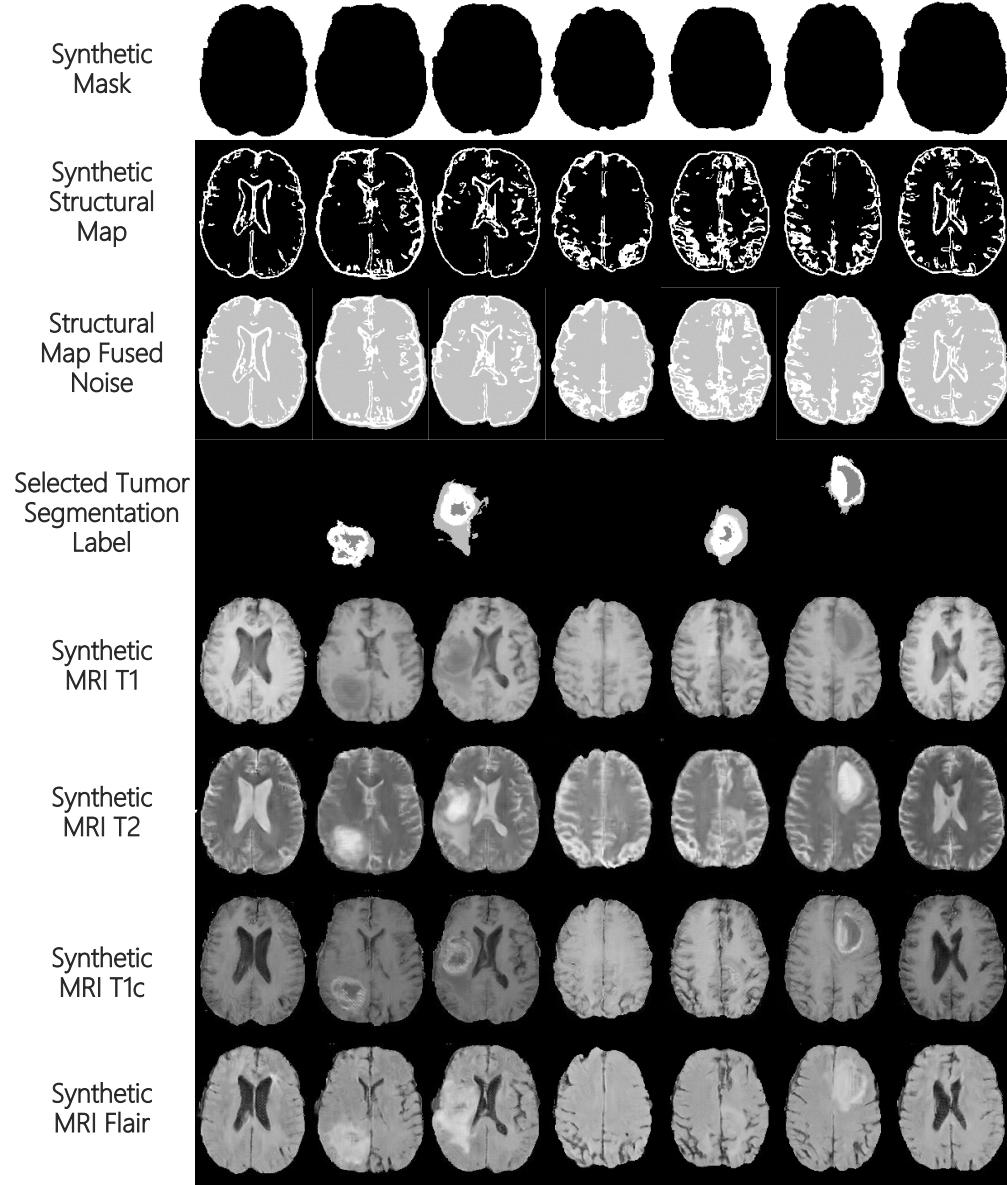


图 6.8: BRATS2015 上合成的结构特征图和多模态MRI。

图6.8中展示了在BRATS2015数据集上合成的一组掩膜、结构特征图、融合了噪声的结构特征图、选择的病灶标签和合成的四个模态的MRI。可以看出，合成的多模态MRI影像相互之间具备较高的配准度，同时还根据输入的病灶标签合成了对应的病灶信息。

表6.3中展示了本文的方法与其他方法在BRATS2015上的量化对比结果。图6.9中展示了SkrGAN中其他方法和本文的方法合成的脑MRI的视觉效果，由于SkrGAN中，采用的Neuromorphometrics脑MRI数据集¹²上筛选条件的未知，本研究采用在类似的数据集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6.9中（e）和（f）中不难看出，本文方法的结构特征图在更加简洁清晰、复杂度更低和限制更少的同时更具有合成指导性。从合成MRI来看，本文方法的合成的脑MRI更加清晰真实，服从输入的结构特征图的合成指导的同时展现了更多样的合成空间。

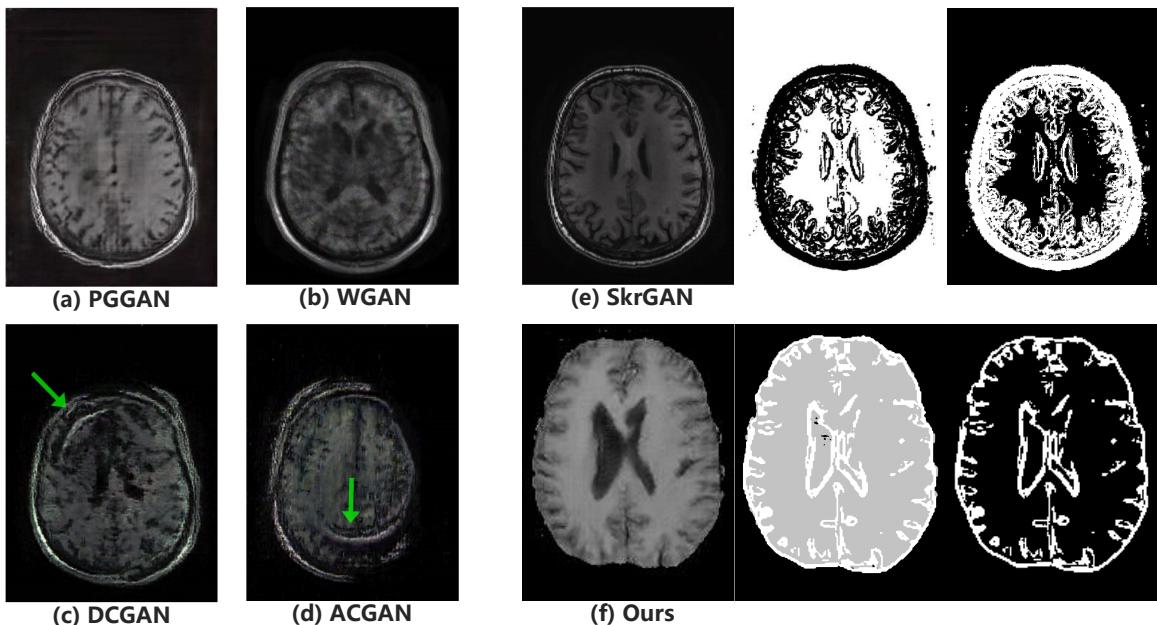


图 6.9: 脑部MRI合成效果对比。（a）PGGAN [74, 97]的合成图像。（b）WGAN [74, 101]的合成图像。（c）DCGAN [51, 74]的合成图像。（d）ACGAN [20, 74]的合成图像，（e）SkrGAN [74]的合成图像、输入的结构特征图和二值反色图。（f）本文方法的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。（a）-（e）为在Neuromorphometrics脑MRI数据集上满足一些未知条件筛选出的数据上进行训练的合成结果，（f）为在类似的BRATS2015上进行训练的合成结果。

图6.10中展示了本研究合成的带肿瘤病灶的脑MRI与当前最好的可合成肿瘤病灶脑MRI方法 [42]的视觉效果对比，除了与本研究一样在BRATS2015数据集上执行了

¹²<http://www.neuromorphometrics.com/>

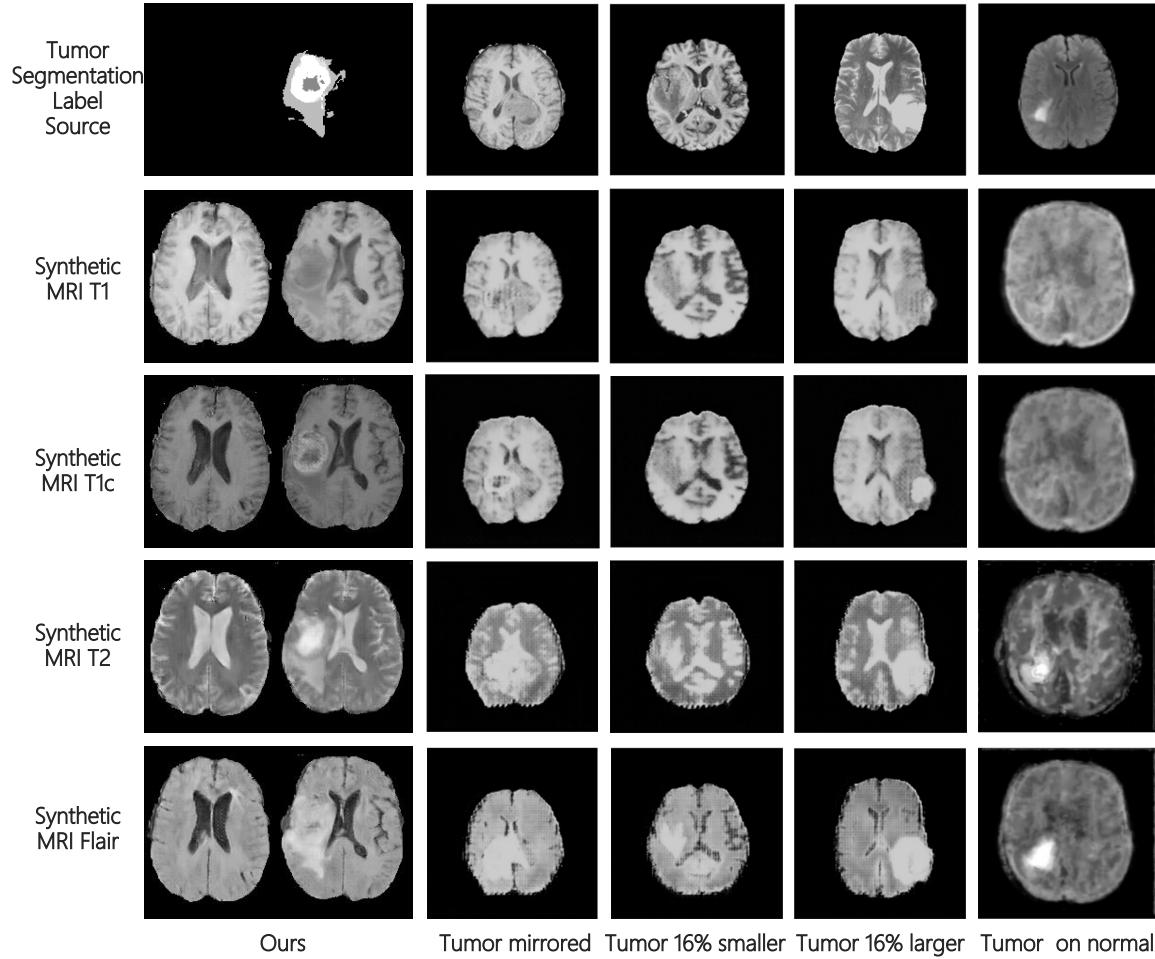


图 6.10: 带肿瘤病灶的脑部MRI合成效果对比。图中左边两列中，第一行为输入的病灶标签，后面四行为本文方法合成的无肿瘤病变和有肿瘤病变的四个模态的脑MRI；右边四列中，第一行为提供的肿瘤病灶标签对应的真实影像，后面四行为脑MRI合成研究 [42] 中对病灶标签进行镜像翻转、缩小、放大处理和无处理合成的有肿瘤病变的四个模态的脑MRI，其中前三列合成时的输入为采用一个训练好的脑组织分割器对第一行的属于BRATS2015数据集的真实影像分割得到的脑组织分割图，而第四列合成时的输入为采用相同脑组织分割器对ADNI数据集中的真实影像分割得到的脑组织分割图。

合成训练，该合成方法还需要额外的ADNI数据集¹³提供脑组织分割图。从图中可以看出，相比于本研究合成的脑MRI，该脑MRI方法 [42]合成的脑MRI整体质量较差：T1模态和T1c模态的区分度不明显；T2和Flair模态中棋盘格效应比较严重；合成的病灶存在超出脑轮廓不够符合真实生理逻辑的情况；合成病灶部分像素平坦无痕、缺乏符合生理结构的纹理细节，合成的病灶在不同模态中表现出的病变程度差异较大；以真实BRATS2015影像的脑组织分割图为输入合成的脑MRI与提供脑组织分割图的原图整体结构上一致度过高，相对缺乏了可变性和多样性；以真实ADNI影像的脑组织分割图为输入合成的脑MRI结构更加不清晰、细节比较混乱，更难以识别为脑部影像；以真实ADNI影像的脑组织分割图为输入合成的各个模态与实际模态的特征较为不符。从对比情况看出本研究的合成影像不仅整体质量更高，在合成病灶的真实程度、模态之间合成内容的一致性、不同模态与真实模态特征的吻合程度等方面都表现更加优秀。

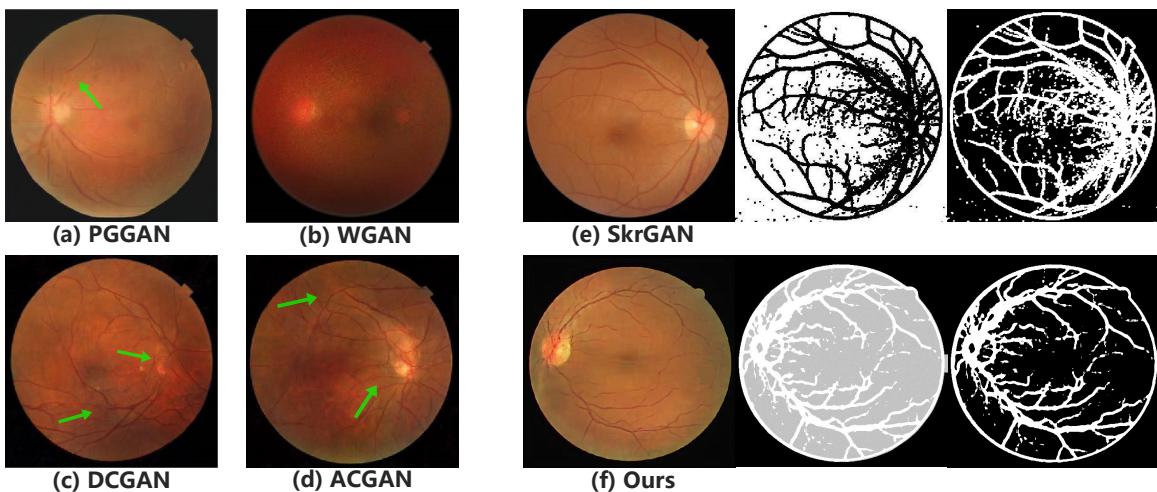


图 6.11：视网膜合成效果对比。（a）PGGAN [74, 97]的合成图像。（b）WGAN [74, 101]的合成图像。（c）DCGAN [51, 74]的合成图像。（d）ACGAN [20, 74]的合成图像。（e）SkrGAN [74]的合成图像、输入的结构特征图和二值反色图。（f）本文方法的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。（a）-（e）为在SkrGAN自建数据集上进行训练的合成结果，（f）为在类似的DRIVE+FIRE数据集上进行训练的合成结果。

¹³<http://adni.loni.usc.edu/>

表6.3中展示了本文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6.11中展示了本文方法在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SkrGAN中其他方法在其自建的6432张视网膜数据集上的合成的视网膜图像对比，本文方法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) 展示了本文方法的结构特征图和SkrGAN的草图的区别，本文方法的结构特征图血管线条走向更加自然真实、噪声更少。

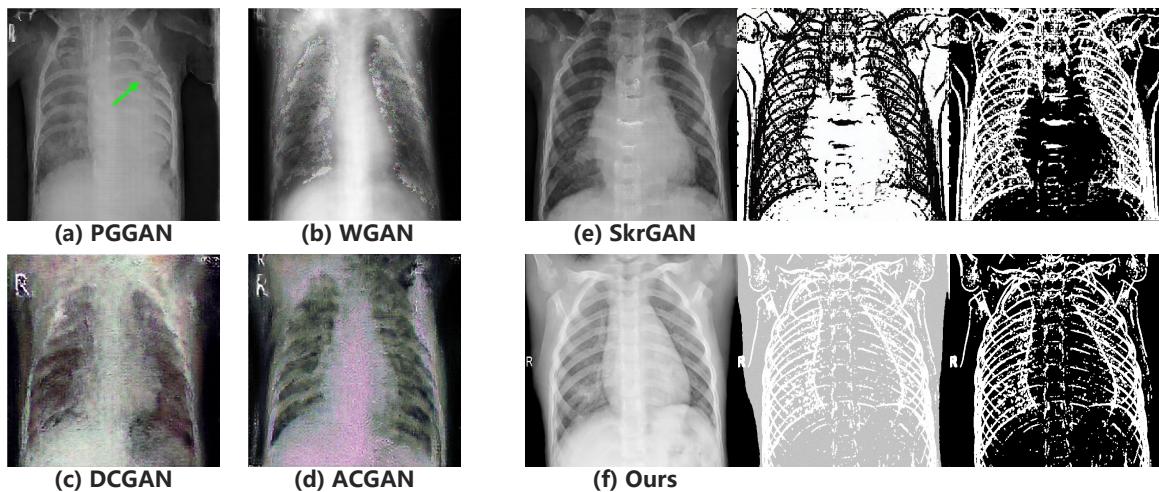


图 6.12: X-ray合成效果对比。(a) PGGAN [74, 97]的合成图像。(b) WGAN [74, 101]的合成图像。(c) DCGAN [51, 74]的合成图像。(d) ACGAN [20, 74]的合成图像。(e) SkrGAN [74]的合成图像、输入的结构特征图和二值反色图。(f) 本文方法的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。

图6.12中展示了本文方法在Kaggle Chest 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，本文方法的合成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) 展示了本文方法的结构特征图和SkrGAN的草图的区别，本文方法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，其中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6.13中 (f) 展示了本文方法在Kaggle Lung 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，本文方法的合成CT更加清晰真实。其中 (e) 和 (f) 展示了本文方法的结构特征图和SkrGAN的草图的区别，本文方法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，其中黑白交错界限不够明晰的同时又充满了更多噪点，因而显得整体更加杂乱无章。从合成的CT来看，SkrGAN的合成的肺中存在缺乏合理生理结构内容的问题，而本文方法的合成的肺中具有更加符合生理结构的气管血管等细微结构。

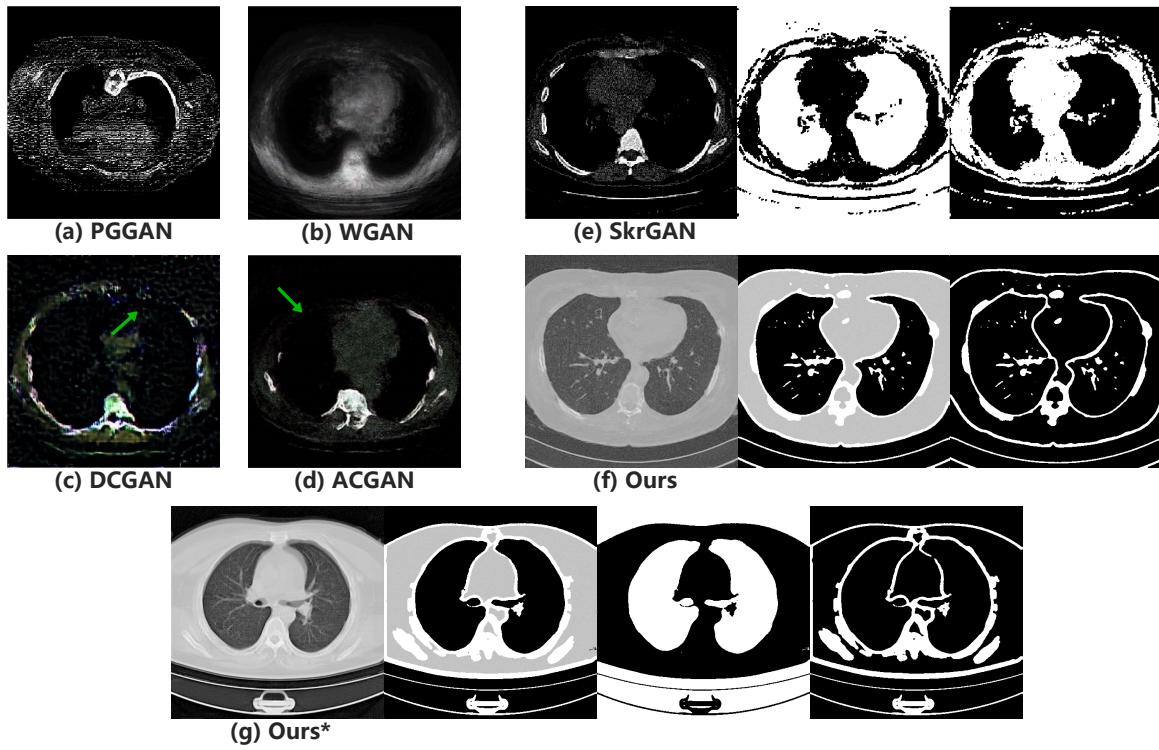


图 6.13: 肺部CT合成效果对比。(a) PGGAN [74, 97]的合成图像。(b) WGAN [74, 101]的合成图像。(c) DCGAN [51, 74]的合成图像。(d) ACGAN [20, 74]的合成图像。(e) SkrGAN [74]的合成图像、输入的结构特征图和二值反色图。(f) 在Kaggle Lung CT数据集上，本文方法的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。(g) 在TC Lung CT数据集上，本文方法的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。(f) - (g) 输出的合成图为标准化后的图像（像素值在0-1之间），视觉上与 (a) - (e) 中未处理的合成结果有一些差异。

图6.13中（g）展示了本文方法在TC Lung CT数据集上合成的结构特征图和CT影像，在具有更为充足的样本和具有对病灶细节进行损失指导的条件下，与SkrGAN和本文方法在Kaggle Lung CT数据集上合成的肺部影像相比，本文合成的天池肺部影像中结构信息更加丰富详实，整体效果更为逼真。

从上面的各项视觉对比评估不难看出，本研究方法相比于当前最好的合成方法合成的医学影像细节纹理上更加丰满精细，轮廓结构上更加真实合理，整体视觉效果与真实影像更为相似。

第七章 结语

在当前合成医学影像的研究中，依然存在复杂的生理结构信息难以合成、难以合成精确配准的多模态影像、难以确保合成有效的病灶信息、对合成病灶的有效性和合成影像的可用性很难量化评估等多项问题。本研究在多项优秀研究的基础上，进行了深入的分析、改进和创新，提出了一套同时解决其中多个问题的解决方案，并通过在多个数据集上的详细实验验证了提出的方案的可行性和有效性。

本文提出的整体解决方案实现了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。首先，本文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影像中提取解剖结构信息，相较于当前最好的草图提取方法，本研究中提取的结构特征图整体上更加干净简明、线条更加清晰、轮廓更加完整。在此基础上，本文提出一种结合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，本研究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，再采用基于ACGAN结构的条件生成器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影像，采用输入的病灶标签图提供合成病灶的指导信息的同时采用一个病灶处理器约束生成器合成指定的病灶内容，此外还可采用一个基于CycleGAN结构的条件转换器实现对合成多模态影像的配准约束。最后，本研究在多种医学影像数据集和多种病灶处理任务中间接评估了合成医学影像用作智能医学影像处理任务的预训练数据或增强数据的可用性、合成影像中病灶的有效性和合成多模态之间的配准程度，同时也使用通用相似性指标直接对合成影像数据集与真实影像数据集的相似性进行评估。多项实验的结果表明，本研究的合成方法在多种不同成像模态、成像部位和病灶种类的医学影像中表现鲁棒，合成的医学影像数据可以用作多种智能医学影像处理任务的预训练数据或增强数据，并能显著提高模型的泛化能力。

总的来说，本文提出的方案包括两个最主要的阶段，即先合成包含结构信息的结构特征图，再进一步合成多模态影像，并可以合成指定病灶信息和实现多模态影像的精确配准。本研究直接将合成数据用于训练智能医学影像处理任务的模型，然后根据模型的评估结果来间接评估合成医学影像的质量。主要的贡献简要总结如下

- 本文提出了一种结构特征图提取方法，可直接从医学影像中提取出更清晰简明的解剖结构信息，无需进行提取训练，再结合VAE和GAN的特点进一步提出的结构特征图生成方法，可从多维正态分布矩阵生成随机结构特征图。
- 本文提出了一种可从结构特征图合成多模态医学影像的方法，通过控制病灶的合成和多模态的配准合成任意数量的带病灶标签的配准多模态医学影像。
- 本文应用了更全面的评估方案，除了一般直接对合成影像数据集与真实影像数据集的相似性的评估外，还将具有合成病灶的合成医学影像用于不同的智能医学影像处理模型的训练，通过对训练好的模型的评估来间接评估合成影像的性能。

受限于硬件的内存，本研究仅进行了2D图像的合成，未尝试直接合成3D图像。此外，本研究中所采用的模型结构以经典的U-net和VGGNet为基础结构，主要是在数据处理方法、模型架构、损失函数和模型应用上的创新，并未在单个模型结构上进行更多的创新，也并未追踪应用最新的优秀模型、算子、微结构和训练技巧。在未来具有充分的时间和硬件资源的条件下，本研究可以继续在最新的优秀模型或算子的基础上进行应用、改进和创新，以验证本研究的方法在其他模型中的鲁棒性、提升方法的训练和推理效率、探索更高质量的医学影像生成方案和尝试直接的3D医学影像的合成。

参考文献

- [1] Huang S-W, Lin C-T, Chen S-P, Wu Y-Y, Hsu P-H, Lai S-H. AugGAN: Cross domain adaptation with GAN-based data augmentation[C] // Proceedings of the European Conference on Computer Vision (ECCV). 2018 : 718 – 731.
- [2] Zhong Z, Zheng L, Zheng Z, Li S, Yang Y. Camera style adaptation for person re-identification[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018 : 5157 – 5166.
- [3] Deng W, Zheng L, Ye Q, Kang G, Yang Y, Jiao J. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018 : 994 – 1003.
- [4] Wei L, Zhang S, Gao W, Tian Q. Person transfer gan to bridge domain gap for person re-identification[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018 : 79 – 88.
- [5] Liu L, Muelly M, Deng J, Pfister T, Li L-J. Generative Modeling for Small-Data Object Detection[C] // Proceedings of the IEEE International Conference on Computer Vision. 2019 : 6073 – 6081.
- [6] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C] // Advances in neural information processing systems. 2012 : 1097 – 1105.
- [7] Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[J]. arXiv preprint arXiv:1409.2329, 2014.
- [8] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets[C] // Advances in neural information processing systems. 2014 : 2672 – 2680.

- [9] Litjens G, Kooi T, Bejnordi B E, Setio A A A, Ciompi F, Ghafoorian M, Van Der Laak J A, Van Ginneken B, Sánchez C I. A survey on deep learning in medical image analysis[J]. *Medical image analysis*, 2017, 42 : 60 – 88.
- [10] Lee J-G, Jun S, Cho Y-W, Lee H, Kim G B, Seo J B, Kim N. Deep learning in medical imaging: general overview[J]. *Korean journal of radiology*, 2017, 18(4) : 570 – 584.
- [11] Shen D, Wu G, Suk H-I. Deep learning in medical image analysis[J]. *Annual review of biomedical engineering*, 2017, 19 : 221 – 248.
- [12] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016 : 2414 – 2423.
- [13] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C] // European conference on computer vision. 2016 : 694 – 711.
- [14] Zhu J-Y, Park T, Isola P, Efros A A. Unpaired image-to-image translation using cycle-consistent adversarial networks[C] // Proceedings of the IEEE international conference on computer vision. 2017 : 2223 – 2232.
- [15] Azadi S, Fisher M, Kim V G, Wang Z, Shechtman E, Darrell T. Multi-content GAN for few-shot font style transfer[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018 : 7564 – 7573.
- [16] Zhao B, Chang B, Jie Z, Sigal L. Modular generative adversarial networks[C] // Proceedings of the European conference on computer vision (ECCV). 2018 : 150 – 165.
- [17] Liang X, Zhang H, Lin L, Xing E. Generative semantic manipulation with mask-contrasting GAN[C] // Proceedings of the European Conference on Computer Vision (ECCV). 2018 : 558 – 573.

- [18] Choi Y, Choi M, Kim M, Ha J-W, Kim S, Choo J. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018 : 8789 – 8797.
- [19] Isola P, Zhu J-Y, Zhou T, Efros A A. Image-to-image translation with conditional adversarial networks[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017 : 1125 – 1134.
- [20] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier gans[C] // Proceedings of the 34th International Conference on Machine Learning-Volume 70. 2017 : 2642 – 2651.
- [21] Sobel I. Camera models and machine perception[R]. [S.l.] : Computer Science Department, Technion, 1972.
- [22] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015 : 3431 – 3440.
- [23] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12) : 2481 – 2495.
- [24] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C] // International Conference on Medical image computing and computer-assisted intervention. 2015 : 234 – 241.
- [25] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille A L. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint arXiv:1412.7062, 2014.
- [26] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille A L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected

- crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4) : 834 – 848.
- [27] Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.
- [28] Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation[C] // Proceedings of the European conference on computer vision (ECCV). 2018 : 801 – 818.
- [29] Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation[C] // Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017 : 11 – 19.
- [30] He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn[C] // Proceedings of the IEEE international conference on computer vision. 2017 : 2961 – 2969.
- [31] Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017 : 2881 – 2890.
- [32] Lin G, Milan A, Shen C, Reid I. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017 : 1925 – 1934.
- [33] Yang M, Yu K, Zhang C, Li Z, Yang K. Denseaspp for semantic segmentation in street scenes[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018 : 3684 – 3692.
- [34] Wang P, Chen P, Yuan Y, Liu D, Huang Z, Hou X, Cottrell G. Understanding convolution for semantic segmentation[C] // 2018 IEEE winter conference on applications of computer vision (WACV). 2018 : 1451 – 1460.

- [35] Li H, Xiong P, Fan H, Sun J. Dfanet: Deep feature aggregation for real-time semantic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019 : 9522 – 9531.
- [36] Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H. Dual attention network for scene segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019 : 3146 – 3154.
- [37] Liu C, Chen L-C, Schroff F, Adam H, Hua W, Yuille A L, Fei-Fei L. Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019 : 82 – 92.
- [38] He J, Deng Z, Zhou L, Wang Y, Qiao Y. Adaptive pyramid context network for semantic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019 : 7519 – 7528.
- [39] Zhang C, Lin G, Liu F, Yao R, Shen C. Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019 : 5217 – 5226.
- [40] Miao S, Piat S, Fischer P, Tuysuzoglu A, Mewes P, Mansi T, Liao R. Dilated fcn for multi-agent 2d/3d medical image registration[C] // Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [41] Van Nguyen H, Zhou K, Vemulapalli R. Cross-domain synthesis of medical images using efficient location-sensitive deep network[C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. 2015 : 677 – 684.
- [42] Shin H-C, Tenenholtz N A, Rogers J K, Schwarz C G, Senjem M L, Gunter J L, Andriole K P, Michalski M. Medical image synthesis for data augmentation and anonymization using generative adversarial networks[C] // International workshop on simulation and synthesis in medical imaging. 2018 : 1 – 11.

- [43] Huo Y, Xu Z, Bao S, Assad A, Abramson R G, Landman B A. Adversarial synthesis learning enables segmentation without target modality ground truth[C] // 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). 2018 : 1217 – 1220.
- [44] Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W, Webb R. Learning from simulated and unsupervised images through adversarial training[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017 : 2107 – 2116.
- [45] Liu M-Y, Breuel T, Kautz J. Unsupervised image-to-image translation networks[C] // Advances in neural information processing systems. 2017 : 700 – 708.
- [46] Kim T, Cha M, Kim H, Lee J K, Kim J. Learning to discover cross-domain relations with generative adversarial networks[C] // Proceedings of the 34th International Conference on Machine Learning-Volume 70. 2017 : 1857 – 1865.
- [47] Zhu Y, Elhoseiny M, Liu B, Peng X, Elgammal A. A generative adversarial approach for zero-shot learning from noisy texts[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018 : 1004 – 1013.
- [48] Zhang H, Patel V M. Densely connected pyramid dehazing network[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018 : 3194 – 3203.
- [49] Gong Y, Karanam S, Wu Z, Peng K-C, Ernst J, Doerschuk P C. Learning compositional visual concepts with mutual consistency[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018 : 8659 – 8668.
- [50] Xian W, Sangkloy P, Agrawal V, Raj A, Lu J, Fang C, Yu F, Hays J. TextureGAN: Controlling deep image synthesis with texture patches[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018 : 8456 – 8465.
- [51] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

- [52] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.
- [53] Perarnau G, Van De Weijer J, Raducanu B, Álvarez J M. Invertible conditional GANs for image editing[J]. arXiv preprint arXiv:1611.06355, 2016.
- [54] Anoosheh A, Agustsson E, Timofte R, Van Gool L. ComboGAN: Unrestrained scalability for image domain translation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018 : 783 – 790.
- [55] Royer A, Bousmalis K, Gouws S, Bertsch F, Mosseri I, Cole F, Murphy K. XGAN: Unsupervised image-to-image translation for many-to-many mappings[G] // Domain Adaptation for Visual Understanding. [S.l.] : Springer, 2020 : 33 – 49.
- [56] Burgos N, Cardoso M J, Guerreiro F, Veiga C, Modat M, McClelland J, Knopf A-C, Punwani S, Atkinson D, Arridge S R, others. Robust CT synthesis for radiotherapy planning: application to the head and neck region[C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. 2015 : 476 – 484.
- [57] Huang Y, Shao L, Frangi A F. Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017 : 6070 – 6079.
- [58] Vemulapalli R, Van Nguyen H, Kevin Zhou S. Unsupervised cross-modal synthesis of subject-specific scans[C] // Proceedings of the IEEE International Conference on Computer Vision. 2015 : 630 – 638.
- [59] Kamnitsas K, Baumgartner C, Ledig C, Newcombe V, Simpson J, Kane A, Menon D, Nori A, Criminisi A, Rueckert D, others. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks[C] // International conference on information processing in medical imaging. 2017 : 597 – 609.

- [60] Fan Z, Sun L, Ding X, Huang Y, Cai C, Paisley J. A segmentation-aware deep fusion network for compressed sensing MRI[C] // Proceedings of the European Conference on Computer Vision (ECCV). 2018 : 55 – 70.
- [61] Anirudh R, Kim H, Thiagarajan J J, Aditya Mohan K, Champley K, Bremer T. Lose the views: Limited angle CT reconstruction via implicit sinogram completion[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018 : 6343 – 6352.
- [62] Costa P, Galdran A, Meyer M I, Niemeijer M, Abràmoff M, Mendonça A M, Campilho A. End-to-end adversarial retinal image synthesis[J]. IEEE transactions on medical imaging, 2017, 37(3) : 781 – 791.
- [63] Iglesias J E, Konukoglu E, Zikic D, Glocker B, Van Leemput K, Fischl B. Is synthesizing MRI contrast useful for inter-modality analysis?[C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. 2013 : 631 – 638.
- [64] Zhang Z, Yang L, Zheng Y. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network[C] // Proceedings of the IEEE conference on computer vision and pattern Recognition. 2018 : 9242 – 9251.
- [65] Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, Shen D. Medical image synthesis with context-aware generative adversarial networks[C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. 2017 : 417 – 425.
- [66] Osokin A, Chessel A, Carazo Salas R E, Vaggi F. GANs for biological image synthesis[C] // Proceedings of the IEEE International Conference on Computer Vision. 2017 : 2233 – 2242.
- [67] Yi X, Babyn P. Sharpness-aware low-dose CT denoising using conditional generative adversarial network[J]. Journal of digital imaging, 2018, 31(5) : 655 – 669.
- [68] Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra M K, Zhang Y, Sun L, Wang G. Low-dose CT image denoising using a generative adversarial network with Wasserstein

- distance and perceptual loss[J]. IEEE transactions on medical imaging, 2018, 37(6) : 1348 – 1357.
- [69] Wolterink J M, Leiner T, Viergever M A, Išgum I. Generative adversarial networks for noise reduction in low-dose CT[J]. IEEE transactions on medical imaging, 2017, 36(12) : 2536 – 2545.
- [70] You C, Li G, Zhang Y, Zhang X, Shan H, Li M, Ju S, Zhao Z, Zhang Z, Cong W, others. CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE)[J]. IEEE Transactions on Medical Imaging, 2019, 39(1) : 188 – 203.
- [71] Lyu Q, You C, Shan H, Wang G. Super-resolution MRI through deep learning[J]. arXiv preprint arXiv:1810.06776, 2018.
- [72] Chartsias A, Joyce T, Giuffrida M V, Tsafaris S A. Multimodal MR synthesis via modality-invariant latent representation[J]. IEEE transactions on medical imaging, 2017, 37(3) : 803 – 814.
- [73] Joyce T, Chartsias A, Tsafaris S A. Robust multi-modal MR image synthesis[C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. 2017 : 347 – 355.
- [74] Zhang T, Fu H, Zhao Y, Cheng J, Guo M, Gu Z, Yang B, Xiao Y, Gao S, Liu J. SkrGAN: Sketching-rendering Unconditional Generative Adversarial Networks for Medical Image Synthesis[C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. 2019 : 777 – 785.
- [75] Kingma D P, Welling M. Auto-Encoding Variational Bayes[J]. international conference on learning representations, 2014.
- [76] Rezende D J, Mohamed S, Wierstra D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models[J]. international conference on learning representations, 2014 : 1278 – 1286.

- [77] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11) : 2278 – 2324.
- [78] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, others. Imagenet large scale visual recognition challenge[J]. International journal of computer vision, 2015, 115(3) : 211 – 252.
- [79] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C] // Advances in neural information processing systems. 2012 : 1097 – 1105.
- [80] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [81] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015 : 1 – 9.
- [82] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016 : 770 – 778.
- [83] Alain G, Bengio Y, Yao L, Yosinski J, Thibodeau-Laufer E, Zhang S, Vincent P. GSNS: generative stochastic networks[J]. Information and Inference: A Journal of the IMA, 2016, 5(2) : 210 – 249.
- [84] Kingma D P, Welling M. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.
- [85] Doersch C. Tutorial on variational autoencoders[J]. arXiv preprint arXiv:1606.05908, 2016.

- [86] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation[C] //Proceedings of the IEEE conference on computer vision and pattern recognition. 2014 : 580–587.
- [87] Girshick R. Fast R-CNN[J]. international conference on computer vision, 2015 : 1440–1448.
- [88] Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks[C] // Advances in neural information processing systems. 2015 : 91–99.
- [89] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016 : 779–788.
- [90] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg A C. SSD: Single shot multibox detector[C] // European conference on computer vision. 2016 : 21–37.
- [91] Roberts L G. Machine perception of three-dimensional solids[D]. [S.l.] : Massachusetts Institute of Technology, 1963.
- [92] Prewitt J M. Object enhancement and extraction[J]. Picture processing and Psychopictorics, 1970, 10(1) : 15–19.
- [93] Costa P, Galdran A, Meyer M I, Abramoff M D, Niemeijer M, Mendonca A M, Campilho A. Towards Adversarial Retinal Image Synthesis.[J]. arXiv preprint arXiv:1701.08974, 2017.
- [94] Wink A M, Roerdink J B. Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing[J]. IEEE transactions on medical imaging, 2004, 23(3) : 374–387.
- [95] Menze B, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber M-A, Arbel T, Avants B, Ayache N,

- Buendia P, Collins L, Cordier N, Corso J, Criminisi A, Das T, Delingette H, Demiralp C, Durst C, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glockner B, Golland P, Guo X, Hamamci A, Iftekharuddin K, Jena R, John N, Konukoglu E, Lashkari D, Antonio Mariz J, Meier R, Pereira S, Precup D, Price S J, Riklin-Raviv T, Reza S, Ryan M, Schwartz L, Shin H-C, Shotton J, Silva C, Sousa N, Subbanna N, Szekely G, Taylor T, Thomas O, Tustison N, Unal G, Vasseur F, Wintermark M, Hye Ye D, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. The multimodal brain tumor image segmentation benchmark (BRATS)[J]. IEEE transactions on medical imaging, 2014, 34(10) : 1993 – 2024.
- [96] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [97] Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation[J]. arXiv preprint arXiv:1710.10196, 2017.
- [98] Dice L R. Measures of the amount of ecologic association between species[J]. Ecology, 1945, 26(3) : 297 – 302.
- [99] Everingham M, Van Gool L, Williams C K I, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) Challenge[J]. International Journal of Computer Vision, 2010, 88(2) : 303 – 338.
- [100] Wang Z, Bovik A C, Sheikh H R, Simoncelli E P. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4) : 600 – 612.
- [101] Arjovsky M, Chintala S, Bottou L. Wasserstein Generative Adversarial Networks[J]. international conference on machine learning, 2017 : 214 – 223.
- [102] Dar S U, Yurt M, Karacan L, Erdem A, Erdem E, Çukur T. Image synthesis in multi-contrast MRI with conditional generative adversarial networks[J]. IEEE transactions on medical imaging, 2019, 38(10) : 2375 – 2388.

- [103] Jog A, Carass A, Roy S, Pham D L, Prince J L. Random forest regression for magnetic resonance image synthesis[J]. Medical Image Analysis, 2017, 35 : 475 – 488.

参考文献

附录一 攻读硕士学位期间取得的科研成果

发表的论文：

- [1]. **Yili Qu**, Yaobin Ke, Wei Yu. A Solution for Input Limit in CNN Due to Fully-Connected Layer[C]//2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS). IEEE, 2018: 611-616.
- [2]. **Yili Qu**, Chufu Deng, Wanqi Su, Ying Wang, Yutong Lu, Zhiguang Chen, Nong Xiao. Multimodal Brain MRI Translation Focused on Lesions[C]//2020 12th International Conference on Machine Learning and Computing (ICMLC). ACM, 2020.
- [3]. Wanqi Su, **Yili Qu**, Chufu Deng, Ying Wang, Fudan Zheng, Zhiguang Chen. Generative Adversarial Network By Discrete Wavelet Transform to Denoise Low-Dose CT Images[C]//27th IEEE International Conference on Image Processing (ICIP). IEEE, 2020.
- [4]. Wei Yu, **Yili Qu**, Yutong Lu. A Hybrid MapReduce Implementation of PCA on Tianhe-2[C]//Journal of Physics: Conference Series. IOP Publishing, 2019, 1168(5): 052013.

专利：

- [1]. 瞿毅力, 王莹, 苏琬棋, 邓楚富, 卢宇彤, 陈志广. 生成配准的带病灶分割标签的多模态MRI的方法、系统及介质, 已公开发明专利, 专利公开号CN110544275A, 2019.
- [2]. 瞿毅力, 苏琬棋, 邓楚富, 王莹, 卢宇彤, 陈志广. 基于条件生成对抗网络的多模态MRI转换方法、系统及介质, 已公开发明专利, 专利公开号CN110544239A, 2019.
- [3]. 瞿毅力, 苏琬棋, 邓楚富, 王莹, 卢宇彤, 陈志广, 肖依. 基于模块化GAN的多模态MRI与多模态CT的转换方法、系统及介质, 已公开发明专利, 专利公开号CN110689561A, 2020.
- [4]. 卢宇彤, 瞿毅力, 郑馥丹, 陈志广. 一种基于进化算法的卷积神经网络结构搜索方法及系统, 已公开发明专利, 专利公开号CN109299142A, 2019.
- [5]. 卢宇彤, 瞿毅力, 陈志广. 可使具有全连接层的CNN接受不定形状输入的方法及系统, 已公开发明专利, 专利公开号CN109583584A, 2019.
- [6]. 苏琬棋, 瞿毅力, 邓楚富, 王莹, 陈志广, 卢宇彤. 基于GAN的多模态低剂量CT转换高剂量CT的方法、系统及介质, 已公开发明专利, 专利公开号CN110559009A, 2019.

参与课题:

- [1]. 基于超级计算机的大数据处理支撑平台研究. 国家自然科学基金委员会项目. 项目批准号: 61872392. 负责人: 陈志广. 依托单位: 中山大学. (项目组主要成员)
- [2]. 基于生成对抗网络的多模态医学影像合成. 中山大学优秀研究生创新发展项目. 批准号: 19lgjyjs61. 依托项目: 面向典型应用的虚拟数据空间验证与优化, 依托项目类别: 国家重点研发计划课题, 依托项目负责人: 陈志广. 项目负责人: 苏琬棋. 承担单位: 中山大学. (项目组主要成员)
- [3]. 深度学习在定量降水预报统计后处理上的应用. 中山大学优秀研究生创新发展项目. 批准号: 19lgjyjs60. 依托项目: 面向大数据的高性能计算支撑系统研究, 依托项目类别: 广东省自然科学基金研究团队项目. 依托项目负责人: 肖依. 项目负责人: 徐风扬. 承担单位: 中山大学. (项目组主要成员)
- [4]. 面向典型应用的虚拟数据空间验证与优化. 国家重点研发计划课题. 课题编号: 2018YFB0203904. 所属项目: 高性能计算虚拟数据空间. 课题负责人: 陈志广. 课题承担单位: 中山大学. (项目组其他研究人员)

致谢

感谢卢宇彤导师、王青导师和陈志广老师的悉心指导，感谢国家超级计算广州中心提供的各项资源和服务，感谢苏琬棋师妹、吕翾师妹、郑馥丹师姐、邓楚富师弟、黎红波同学、徐风杨同学、柯耀斌同学、王莹同学和余炜师姐在我硕士期间的诸多帮助和陪伴！感谢实验室所有老师和同学给我的帮助！感谢家人始终如一的充分理解和无私支持，家人永远是我最坚强的后盾！一路走来，家人、老师、同学等周围人们的帮助与支持成就了今天的我，我时刻铭记于心、知恩图报！

今年是极其特殊的一年，在研究过程中我遇到了许多客观的困难，但我和许多的中国人民一起，克服了自身的困难，一起赢得了一场胜利！在此，我要特别感谢新冠肺炎疫情期间全国奋斗在抗疫一线的白衣天使、抗疫战士！感谢湖北人民，特别是武汉人民的坚守付出！感谢祖国和人民同心协力的守护！感谢国际社会给中国提供的援助和支持！同时，我也要向一直在医疗物资生产一线加班加点的父母表达敬意！

最后，愿中华民族伟大复兴早日实现，愿祖国繁荣昌盛人民幸福，愿世界和平无病无灾！我愿为这丰富多彩、生机盎然的祖国大地的建设努力贡献自己的绵薄之力，为构建人类命运共同体和共产主义事业奋斗终身！

瞿毅力

二零二零年五月