

# 中山大学硕士学位论文

## 基于GAN的多模态医学影像合成

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

医学图像数据的采集和标注一直是医学图像处理任务中面临的挑战，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客观的方式来验证和评估其性能。

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，无需配准训练数据即可从随机矩阵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充分验证了合成病灶的有效性和合成影像的可用性。主要工作有：

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动编码器的训练，最后得到的模型实现了从正态分布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影像合成方案，将随机生成的结构特征图选取的病灶标签融合后再合成多模态医学影像，通过模态转换器实现合成的多模态之间的转换一致性约束确保合成的多模态之间的配准，通过病灶处理器实现对合成影像的病灶标签还原约束确保合成影像根据输入的标签生成对应的病灶。

3.本研究使用两项指标对合成影像的质量进行直接评估，同时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性能。结果表明合成影像可用于预训练和数据增强并能显著提升模型的泛化能力。

**关键词：**医学影像、图像合成、多模态配准、病灶、生成对抗网络

# Synthesis of Multimodal Medical Images Based on GAN

Major: Software Engineering

Name: Yili Qu

Supervisor: Prof. Yutong Lu

## Abstract

The collection and annotation of medical image data have always been a challenge in many data-driven medical image processing tasks, especially for those based on registered multimodal medical image data. This 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, an objective way is needed to verify and 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 even if registration training data are unavailable. Experiments on multiple datasets have comprehensively verified the effectiveness of synthetic lesions and the availability of synthetic images. The main contribution is as follows:

1. This study proposed a structural map extraction and synthesis method based on 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 normal distribution.

2. This study proposed a multimodal registration medical images synthesis scheme that

## Abstract

---

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.In this study, two indicators were used to directly evaluate the quality of the synthesized images. Meanwhile, the performance of the synthesized images was indirectly evaluated by applying the synthesized images to the training of intelligent medical image processing tasks and evaluating the trained models. The results shown that the synthesized image can be used for pre-training and data augmentation and can significantly improve the generalization ability of the model.

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

## 目录

摘要 . . . . .	I
<b>Abstract</b> . . . . .	II
<b>第一章 绪论</b> . . . . .	1
1.1 研究背景和目的 . . . . .	1
1.2 研究内容和技术路线 . . . . .	5
1.3 本文的创新点 . . . . .	6
1.4 本文的结构 . . . . .	7
<b>第二章 国内外研究现状</b> . . . . .	8
2.1 图像合成的研究现状 . . . . .	8
2.2 医学影像合成的研究现状 . . . . .	9
2.3 小结 . . . . .	11
<b>第三章 基础方法</b> . . . . .	12
3.1 基于CNN的图像分类 . . . . .	12
3.2 生成对抗网络 . . . . .	14
3.3 变分自编码器 . . . . .	15
3.4 语义分割 . . . . .	16
3.5 目标检测 . . . . .	18
<b>第四章 多模态医学影像的合成方法</b> . . . . .	19
4.1 整体架构 . . . . .	19
4.2 结构特征图的提取和生成 . . . . .	20
4.3 多模态影像的合成 . . . . .	25
4.4 多模态影像的配准 . . . . .	28
4.5 病灶信息的添加和合成 . . . . .	29
4.6 合成数据集的构建 . . . . .	29

<b>第五章 合成医学影像的性能评估 . . . . .</b>	<b>31</b>
5.1 数据集 . . . . .	31
5.2 实验设置 . . . . .	33
5.3 多模态医学影像合成方法的消融实验 . . . . .	36
5.4 多个数据集上与其他方法的评估对比 . . . . .	38
5.5 合成病灶的有效性评估 . . . . .	40
5.6 合成医学影像的可用性评估 . . . . .	42
5.7 合成影像的视觉评估 . . . . .	46
<b>第六章 结语 . . . . .</b>	<b>53</b>
<b>参考文献 . . . . .</b>	<b>55</b>
<b>附录一 学术成果 . . . . .</b>	<b>68</b>
<b>致谢 . . . . .</b>	<b>70</b>

# 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，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，许多机构不得不拒绝公开自己的数据。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。

随着生成对抗网络的强大生成能力在一系列研究中逐渐展现，面对这种医学影像数据集稀少、数据量小的现状，利用合成的医学影像数据来有效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，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，无法实现的模态之间相互配准。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，其最有价值的地方在于其中包含的病灶信息。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。

在上述的问题和挑战中，涉及到的核心技术背景和概念包括深度学习、医学影像及其模态、多模态图像的配准、病灶和标注、图像合成，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（Deep Learning）是机器学习（Machine Learning）领域中一个新的研究方向，它的出现使得机器学习更接近于最初的目标——人工智能（Artificial Intelligence）。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）模型 [1]、循环神经网络（RNN）模型 [2]、生成对抗网络（GAN）模型 [3]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的前景 [4–6]。
- **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）、CT成像、PET成像、B超成像、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医学影像的智能处理任务的训练和学习中，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被称为配准（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的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神经网络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。
- **病灶和标注** 病灶（Lesion）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，医学影像的训练数据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（Annotate）得到可以用于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，医生们的本职工工作常常十分繁重，大量的标注任务需要医生长期牺牲空闲时间逐步完成。
- **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的目标图像的过程，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转换的图像转换。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）[3]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[7–10]、人像转换[11–14]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。

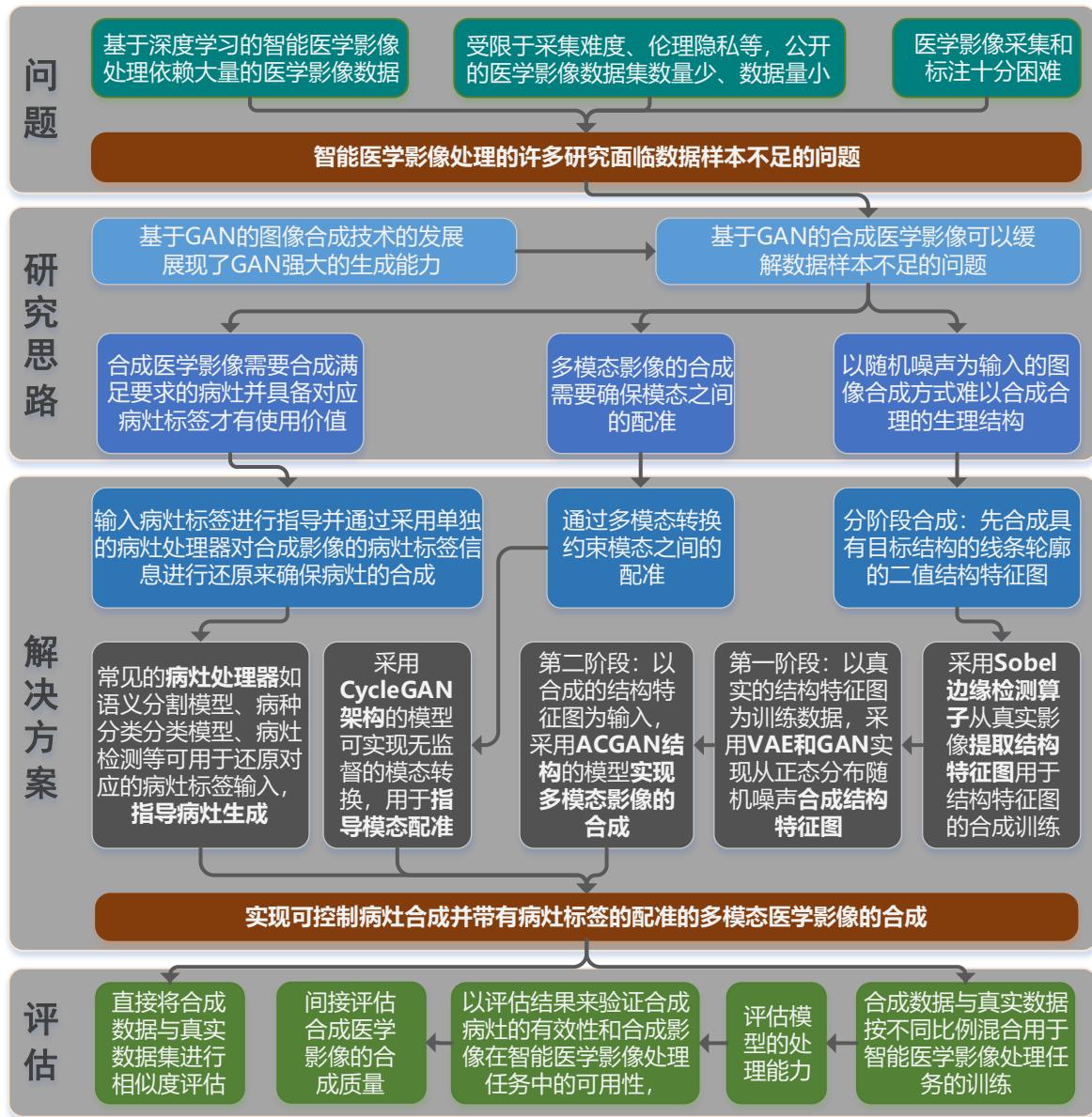


图 1.1: 研究内容。

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少、要求配准的训练数据、合成病灶不可控、合成数量有限制、合成生理结构不合理、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

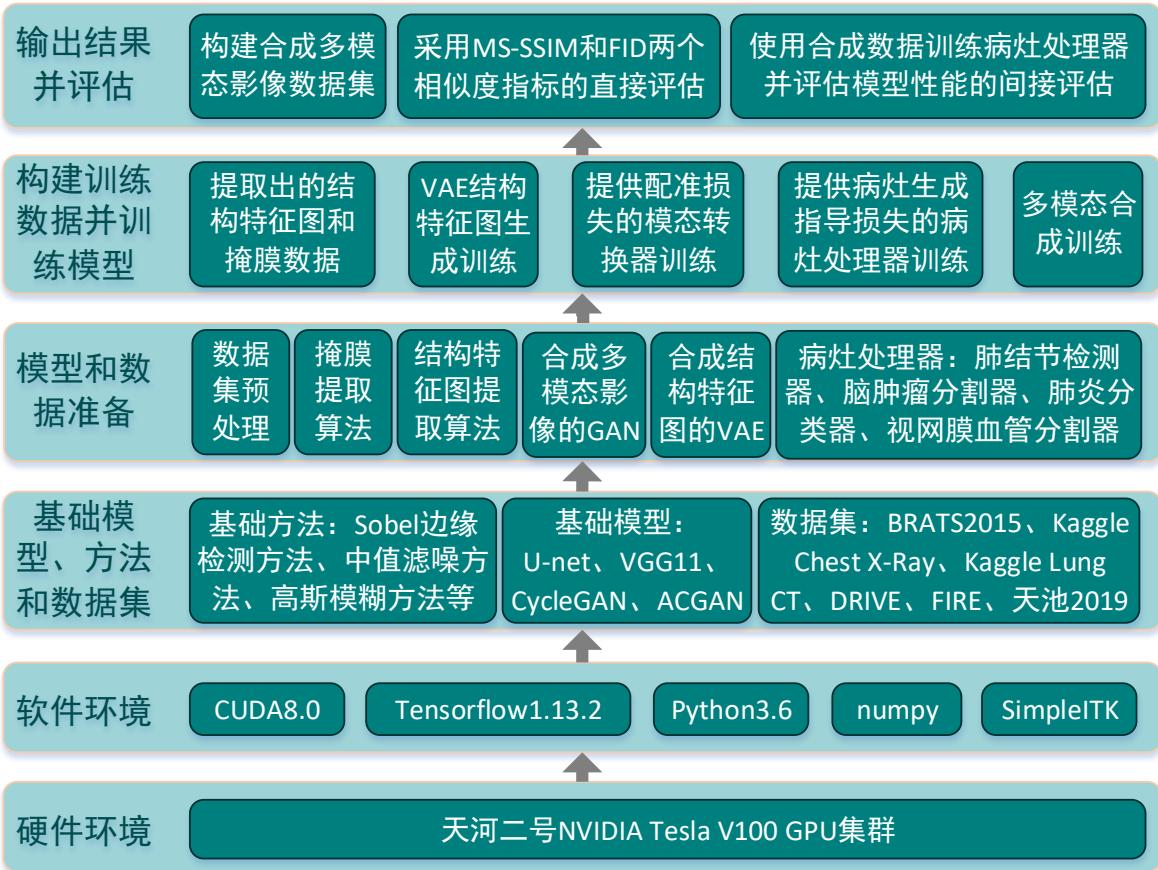


图 1.2: 技术路线。

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的线条轮廓二值图，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 [15]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 [9]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。

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，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实现对生成器根据输入合成指定的病灶内容的约束。

对合成的多模态医学影像，本研究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的合成质量的评估。

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，图中简要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

针对本研究面临的上述挑战，本文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，无需配准训练数据即可从随机矩阵合成具有指定病灶的配准多模态医学影像，并在多个数据集上充分验证了合成病灶的有效性和合成影像的可用性。合成影像可用于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，以此可有效缓解智能医学影像处理任务中数据不足的问题。本文具体的创新点如下：

- 本文提出了一种基于Sobel算子（Sobel Operator）[16]的结构特征图提取和生成方法，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，本文提取的结构特征图更加干净简洁、线条清晰、完整合理。
- 本文提出了一种基于变分自动编码器和生成对抗网络的结构特征图合成方法，可从随机正态分布矩阵合成任意数量的、具有充分多样性的结构特征图和对应

的器官掩膜。

- 本文提出了一种可控制病灶合成的医学影像合成方案，输入结构特征图和选取的病灶标签，再通过病灶处理器从合成影像中还原出病灶标签信息，由此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。
- 本文提出了一种多模态配准影像合成方案，合成多模态影像时通过模态转换器提供的合成多模态之间的转换一致性损失约束，来确保合成多模态的配准。
- 本文提出了一种合成病灶和合成医学影像的评估方案，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，间接地评估了合成病灶的有效性和合成影像的可用性。

## 1.4 本文的结构

本文有六章，每章的内容如下：

**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。**介绍国内外的研究现状并进行简单的分析与总结

**第三章：基础方法。**介绍本文后续章节在方法和实验中将会运用的基础方法，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。

**第四章：多模态医学影像的合成。**本章是本文方法的完整描述，首先对本文方法整体架构进行了介绍，再依次对结构特征图的提取和生成方法、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。

**第五章：合成医学影像的性能评估。**本章节介绍对合成的医学影像的评估方法和实验结果，包括数据集介绍、评估指标介绍、消融实验、合成影像的量化评估、合成病灶的有效性评估、合成数据集在智能医学影像处理任务上的可用性评估。

**第六章：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

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到了充分的发展，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。本研究基于无数前人的优秀研究基础，从国内外的研究现状深入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。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。

### 2.1 图像合成的研究现状

上一章对图像合成的概念进行了介绍和解释，本节对其研究现状继续详细介绍。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FCN [17]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，例如 SegNet [18]、U-Net [19]、DeepLab系列 [20–23]、Fully Convolutional DenseNet [24]、E-Net [25]、EncNet [26]、Mask R-CNN [27]、PSP-Net [28]、RefineNet [29]、G-FRNet [30]、DecoupledNet [31]、DenseASPP [32]、ResNet DUC [33]、DFANet [34]、DANet [35]、Auto Deeplab [36]、APCNet [37]、CANet [38] 等，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从结构上被模式化成一个编码器和一个解码器连接起来的生成器结构。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 [7,8,39,40]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 [41–43]。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 [14,44–48]，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 [11–14]、风格迁移 [9,10]和草图补全 [14,49]等。在结构设计和训练方法上，DCGAN [50]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 [15]在GAN最重要的变体CGAN [51]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的前景。之后CycleGAN [9]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 [52]、StarGAN [13]、ContrastGAN [12]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 [11]、ComboGAN [53]和XGAN [54]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，尤其是在语义分割、人像合成等领域，图像合成方法都取得了充分深入的发展和应用，涌现出了大量的优秀研究成果。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分的可行性，这一方面开拓了图像合成方法的应用领域，同时也能够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 [55]、稀疏编码 [56,57]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 [39,40]。

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 [58]，重建 [59,60]、合成 [41,43,61,62]、转换 [40,58,63–68]和超分辨率 [69,70]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。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 [63]，脑部MRI转换合成CT图像 [64]，低剂量肺部CT合成高剂量CT [66–68]等一些通过无辐射的MRI或低辐射的低剂量CT影像合成高辐射的高剂量CT来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可以提高治疗的可行性 [55]。这些的医学影像模态转换合成的研究中，很多研究仅针对两模态之间的转换 [40, 55, 57, 58, 63–65]，对于多个模态的互相转换合成的研究则相对较少 [41, 71, 72]。在多模态影像合成的研究中，有研究 [71]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，而在一些医学图像的转换配准的研究 [39]中CNN又详细展示了在模态配准方向的潜力，于是相关研究 [72]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MRI合成，但其限制了输出为单一模态。

与模态转换合成方式不同，最近的另一些研究则试图通过合成数据来缓解医学影像数据样本稀少的难题，如脑MRI的合成 [41]、视网膜的合成 [61]、多种不同部位和不同模态的单模态医学影像的合成 [73]。

其中，脑MRI合成研究 [41]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结构分割图，不仅需要额外的数据标签和额外训练一个脑结构分割网络，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额外的损失约束，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。

一项对于视网膜合成的研究 [61]通过变分自编码器(VAE) [74, 75]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，从合成数量上没有限制，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SkrGAN [73]与本研究思路不谋而合，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，并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没有适合的使用场景。

### 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、可控可验证的病灶的合成和客观的合成数据的可用性评估。

## 第三章 基础方法

本文在前两个章节中简要介绍了研究的技术路线和相关研究基础，在本章节中，将对本研究使用的基础方法进行介绍，本文后续章节中提出的方法将对本章提到的方法进行调用、改进和扩展。本章将对基于CNN的图像分类、生成对抗网络、变分自动编码器、语义分割、目标检测进行依次介绍。

### 3.1 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中图像所属类别的任务。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，Convolutional Neural Networks）是在传统神经网络的基础上应用了卷积操作的网络。通常，一个复杂的卷积神经网络中，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，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）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 [76]提出，但直到2012年ILSVRC（ImageNet Large Scale Visual Recognition Challenge）[77]中CNN模型AlexNet [78]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潜力才逐步展现。此后CNN模型进入了飞速发展期，VGGNet [79]、GoogleNet [80]、ResNet [81]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 $\text{argmax}()$ 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。

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 ( $224 \times 224$ RGB image)					
conv3-64	conv3-64 <b>LRN</b>	conv3-64 <b>conv3-64</b>	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 <b>conv3-128</b>	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 <b>conv1-256</b>	conv3-256 conv3-256 <b>conv3-256</b>	conv3-256 conv3-256 <b>conv3-256</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

图 3.1: VGGNet模型结构设置 [79]。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。

## 3.2 生成对抗网络

生成对抗网络（GAN， Generative Adversarial Networks） [3]提出后，与卷积神经网络结合成为一种无监督学习的深度学习模型 [50]。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。通常，生成器 $G$ 接收一个随机的噪声 $z$ ，通过这个噪声生成图像 $x$ ；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 $D(x)$ ，如果概率为1，就代表鉴别器判断输入的图像100%是真实的图像，而输出为0，就代表鉴别器判断此图是与真实图像完全不同的图像。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。这样， $G$ 和 $D$ 的训练过程就是一个动态的“对抗”学习过程，最终的对抗平衡点为博弈论中的纳什均衡点（Nash Equilibrium Point）。损失函数约束两者形成的对抗训练最终使得两个网络达到动态均衡，其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基本接近 0.5，即处于随机猜测水平。GAN的一个典型损失函数如下：

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

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）[82]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 [15]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。

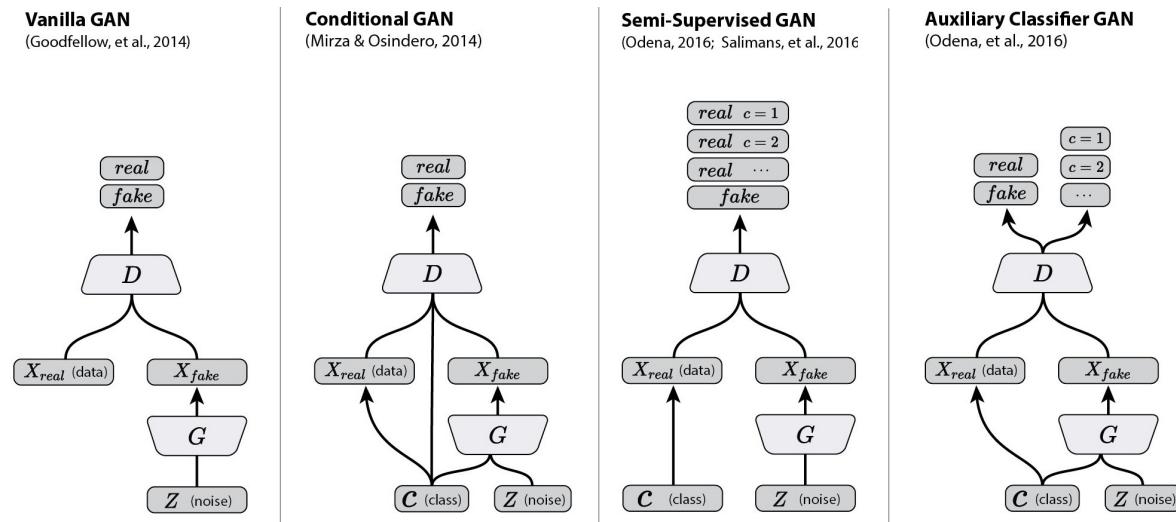


图 3.2: ACGAN模型结构<sup>1</sup>。

### 3.3 变分自编码器

变分自编码器（VAE，Variational Auto-Encoder）[83]是另一类重要的生成模型，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[84]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

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

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表征空间，低级表征被称作本征向量（Latent Vector），即 $z$ 。解码器接收数据的低级表征，然后解码还原数据的高级表征，即重建的数据 $x_r$ 。

在标准VAE中，低级表征 $z$ 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 [83] 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 [83] 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 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）。在基于CNN的计算机视觉研究出现以前，TextonForest和随机森林分类器是常用的语义分割方式。2015年全卷积神经网络（FCN）[17]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 [19]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 $argmax()$ 函数即可得到分割掩膜标签。

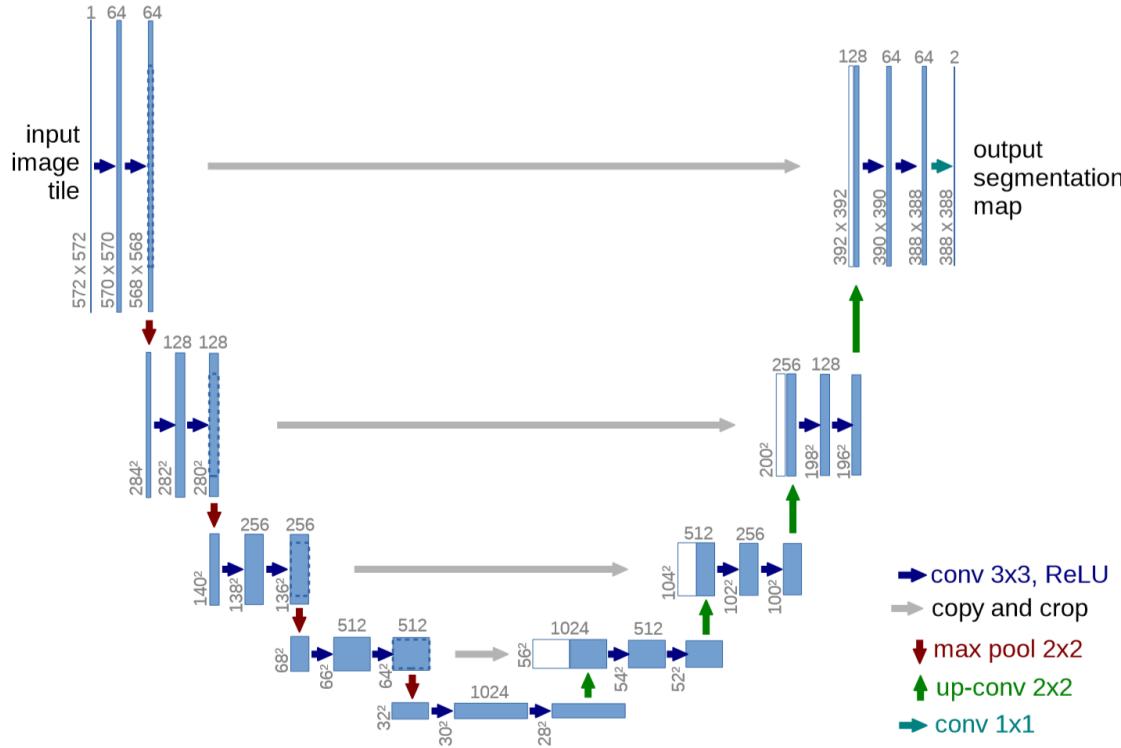


图 3.3: U-Net模型结构 [19]。

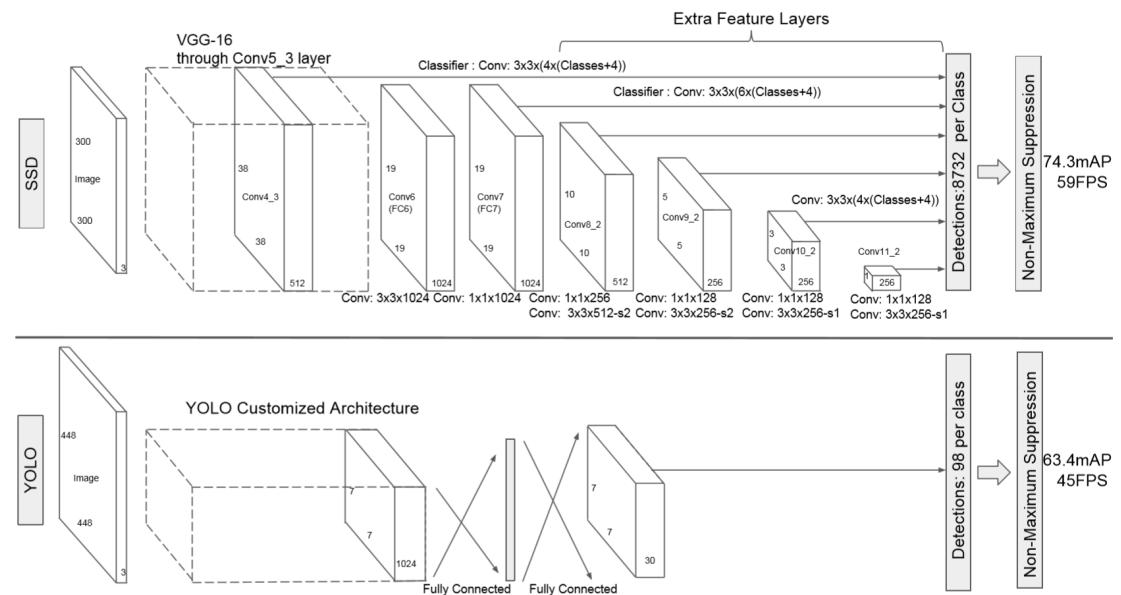


图 3.4: YOLO和SSD模型结构的对比 [85]。

### 3.5 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。当前比较流行的模型可以分为两类：一类基于Region Proposal的R-CNN [86]类模型（R-CNN, Fast R-CNN [87], Faster R-CNN [88]等），包含两个阶段，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以筛选出最终的检测框；另一类是YOLO [89]、SSD [85]等单阶段算法模型，仅用单个卷积神经网络即可实现直接预测不同目标的类别与位置。

如图3.4所示，SSD模型基于VGG16 [79]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）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 多模态医学影像的合成方法

本文在第一章中介绍了本研究面临的三个主要的挑战和对应的研究思路与研究内容，在第二章中分析了国内外的研究现状和仍存在的问题，然后在上一章节中对几种主要的基于深度学习的图像处理任务进行了介绍。本章对本研究基于前述的研究思路和基础方法，在现有研究的基础上的发展、改进和创新进行介绍。本章提出的解决方案将对上一章介绍的基础方法直接应用，而不再展开介绍。本章提出的方案致力于解决当前研究仍存在的问题和不足，实现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，随后依次介绍各个阶段解决的问题和详细的方法步骤。

### 4.1 整体架构

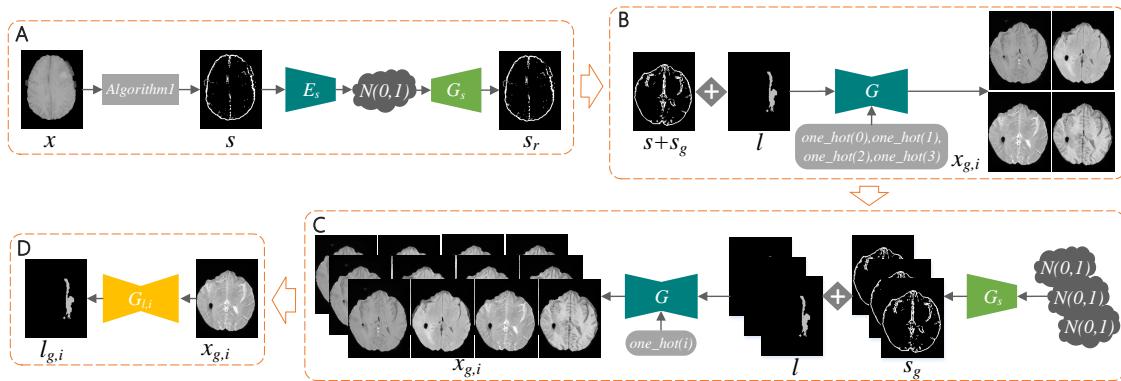


图 4.1: 整体架构。A是结构特征图的提取和生成阶段。B是多模态影像合成阶段。C是合成数据集构建阶段。D是合成数据可用性验证阶段。

图 4.1所示包括四个主要阶段。在结构特征图提取和生成阶段，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。多模态影像合成阶段，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。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，合成数据可用性验证阶段，对合成数据在智能医学影像处理任务中的可用性进行评估验证（评估阶段见下一章）。

---

**Algorithm 1** Structural Map Extraction

- 
- 1: Input a real image  $x$ , pixel threshold  $\alpha$  and  $\beta$  and  $\gamma$ , gaussian kernel variance  $\sigma_1$  and  $\sigma_2$
  - 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:  $s = s \times (x > \gamma)$
- 

## 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真实的结构信息。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，视网膜血管分布图可以看作是视网膜图像的结构特征图 [61]。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，一些研究从组织分割标签 [41]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，本研究设计了一种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，无需额外训练数据的优点。

### 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 [90]、Prewitt算子 [91]、Sobel算子 [16]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图像。如算法 1 中所示，本研究探索了一种从 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。

算法 1 中，使用 Sobel 检测算子 `sobel()` 从真实图像中提取水平和垂直边缘检测图。每个边缘检测图通过执行最大规约 `reduce_min()` 和最小规约 `reduce_min()`，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  $3 \times 3$  的高斯模糊 `gaussian_blur()` 以实现对线条轮廓的加粗，然后将两张图与通过均值函数 `mean()` 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  $3 \times 3$  的中值滤噪函数 `medfilt()` 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。如图 5.1，一些有病灶结构信息的结构特征图，可以使用病灶分割标签转换的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。

SkrGAN [73] 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，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。

### 4.2.2 结构特征图生成训练

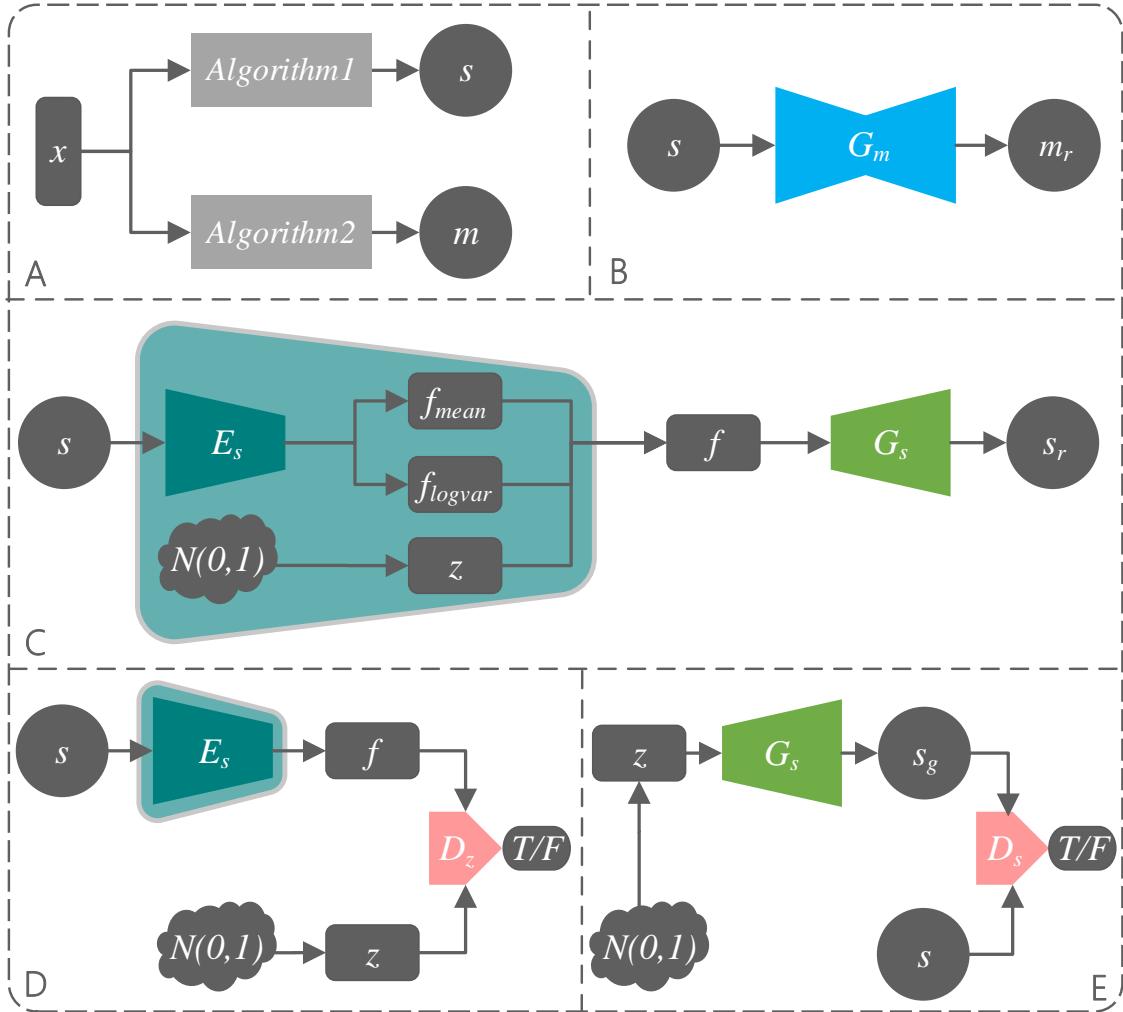


图 4.2: 结构特征图生成训练。A-E依次为结构特征图和掩膜的提取、掩膜的生成训练、VAE重建结构特征图训练、VAE编码器的对抗性训练和VAE解码器的对抗性训练。 $x$ 是输入的真实影像， $s$ 是结构特征图。 $E_s$ 是VAE编码器，输出编码矩阵 $f_{mean}$ 和 $f_{logvar}$ 。 $z$ 是来自多维正态分布 $\mathcal{N}(0, 1^2)$ 的随机噪声采样，而 $f$ 是近似正态分布矩阵。 $G_s$ 是VAE解码器， $s_r$ 是重构的结构特征图，而 $s_g$ 是生成的随机结构特征图。 $D_s$ 是结构特征图鉴别器， $D_z$ 是特征图分布鉴别器。 $G_m$ 是掩膜生成器， $m_r$ 是从 $s$ 生成的掩膜。

生成提供结构信息的输入时，脑MRI合成研究 [41]仍需要输入真实图像以获取生

---

**Algorithm 2** Mask Extraction

---

```

1: Input a real grayscale image  $x$ , pixel threshold  $\alpha$ , 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:  $m = \text{medfilt}(m)$ 

```

---

成的脑结构分割图，这大大减少了合成数据的多样性。合成视网膜的研究 [61] 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多样性并且无需其他训练标签的结构特征图。此外，生成器 $G_m$ 被训练用于从大脑结构特征图中获取目标器官区域的掩膜，以供之后用于匹配病灶标签。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 $\text{resize}()$ 为最近邻插值函数， $\text{crop\_padding}()$ 为边距裁切函数。如图 4.2 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；

$E_s$ 、 $G_s$ 、 $D_z$ 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。 $E_s$ 为一个最后两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。其中，对所有模型的改进都包括将所有步长为2的池化（Pooling）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）上采样改为步长为2的最近邻插值上采样后接一个卷积层，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。各个网络的损失函数如下：

- 阶段B: 掩膜生成损失

$$\mathcal{L}_m(G_m) = \mathbb{E}_{m,s}[\|m - m_r\|_2^2], \quad (4.1)$$

其中  $m_r = G_m(s)$ 。

- 阶段C: 结构特征图重建损失

$$\mathcal{L}_r(E_s, G_s) = \mathbb{E}_{s,f,m}[\|s - s_r\|_2^2 + \|m_r \times s_r\|_2^2], \quad (4.2)$$

其中  $s_r = G_s(f)$ 。

- 阶段D: 特征图分布对抗训练损失

$$\mathcal{L}_{d,z}(D_z) = \mathbb{E}_{s,z}[\|D_z(z) - 1\|_2^2 + \|D_z(f)\|_2^2], \quad (4.3)$$

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

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

$$\mathcal{L}_{d,s}(D_s) = \mathbb{E}_{s,z}[\|D_s(s) - 1\|_2^2 + \|D_s(s_g)\|_2^2], \quad (4.5)$$

$$\mathcal{L}_{g,s}(G_s) = \mathbb{E}_z[\|D_s(s_g) - 1\|_2^2 + \|m_g \times s_g\|_2^2], \quad (4.6)$$

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

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0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：

$$s' = s + z' \times (1 - m) \times (1 - s), \quad (4.7)$$

其中， $z'$ 是从均匀分布 $\mathcal{U}(\alpha_1, \alpha_2)$ 采样的随机噪声，默认 $\alpha_1, \alpha_2$ 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5.2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，降低了学习难度。

## 4.3 多模态影像的合成

如图 4.3所示，首先，本节通过训练好的 $G_s$ 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，本节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。

$D$ 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， $G$ 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， $D$ 为在最后两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 [15]结构。其中，对所有模型的改进都包括将所

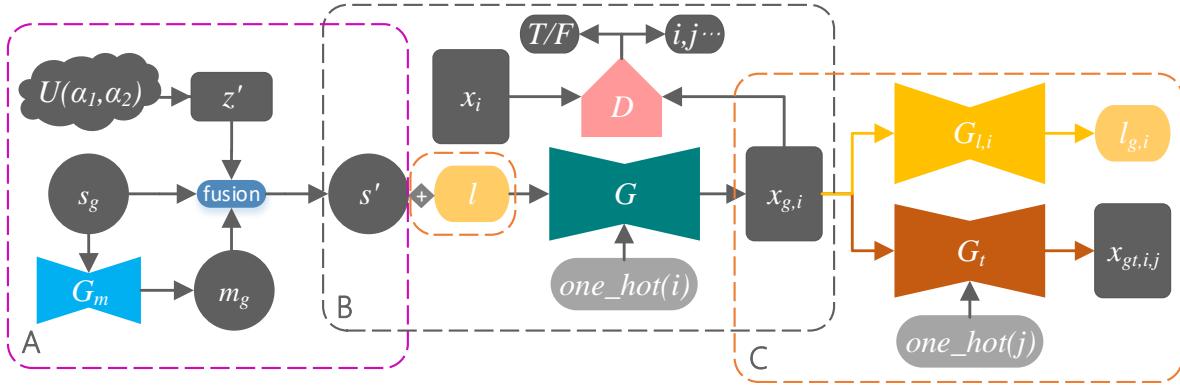


图 4.3: 多模态影像的合成。紫色框A中为结构特征图与随机噪声融合过程; 灰色框B中为核心的医学影像合成过程; 黄色框C中为可选的病灶合成与配准的指导和约束过程。 $s_g$ 是随机合成的的结构特征图,  $m_g$ 是根据 $s_g$ 生成的掩膜,  $z'$ 是从均匀分布 $U(\alpha_1, \alpha_2)$ 采样的随机噪声,  $s'$ 是融合了噪声的结构特征图,  $l$ 是随机选择的真实病灶标签。 $one\_hot(i)$ 是模态*i*的独热矩阵表示。 $G$ 是条件生成器,  $D$ 是鉴别器,  $x_i$ 是模态*i*的真实影像。 $G_{l,i}$ 是模态*i*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。

有步长为2的池化 (Pooling) 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步长为2的最近邻插值上采样再接一个卷积层, 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

$$\begin{aligned} \mathcal{L}_d(D) = & \mathbb{E}_{x,s_g,l} \left[ \sum_{i=0} (\|d(x_i) - 1\|_2^2 + \|d(x_{g,i})\|_2^2 + \right. \\ & \left. \|c(x_i) - i\|_2^2 + \|c(x_{g,i}) - i\|_2^2) \right], \end{aligned} \quad (4.8)$$

$$\mathcal{L}_g(G) = \mathbb{E}_{s_g,l} \left[ \sum (\|d(x_{g,i}) - 1\|_2^2 + \|c(x_{g,i}) - i\|_2^2) \right], \quad (4.9)$$

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 $U(\alpha_1, \alpha_2)$ 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

$$\mathcal{L}_{les}(G) = \mathbb{E}_{s_g, l} \left[ \sum_{i=0} (\|l - l_{g,i}\|_2^2) \right], \quad (4.10)$$

其中  $l_{g,i} = G_{l,i}(x_{g,i})$ 。

- 模态配准损失

$$\mathcal{L}_{reg}(G) = \mathbb{E}_{s_g, l} \left[ \sum_{j=0} \sum_{i=0, i \neq j} (\|x_{g,i} - x_{gt,j,i}\|_2^2) \right], \quad (4.11)$$

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:

$$\mathcal{L}(G) = \mathcal{L}_g(G) + \mathcal{L}_{les}(G) + \mathcal{L}_{reg}(G). \quad (4.12)$$

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:

$$\mathcal{L}_p(G) = \mathbb{E}_{s,l} \left[ \sum_{i=0} (\|x_{g,i} - x_i\|_2^2 + \|x_{g,i} \times m_i - x_i \times m_i\|_2^2) \right]. \quad (4.13)$$

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 [73]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。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。

#### 4.4 多模态影像的配准

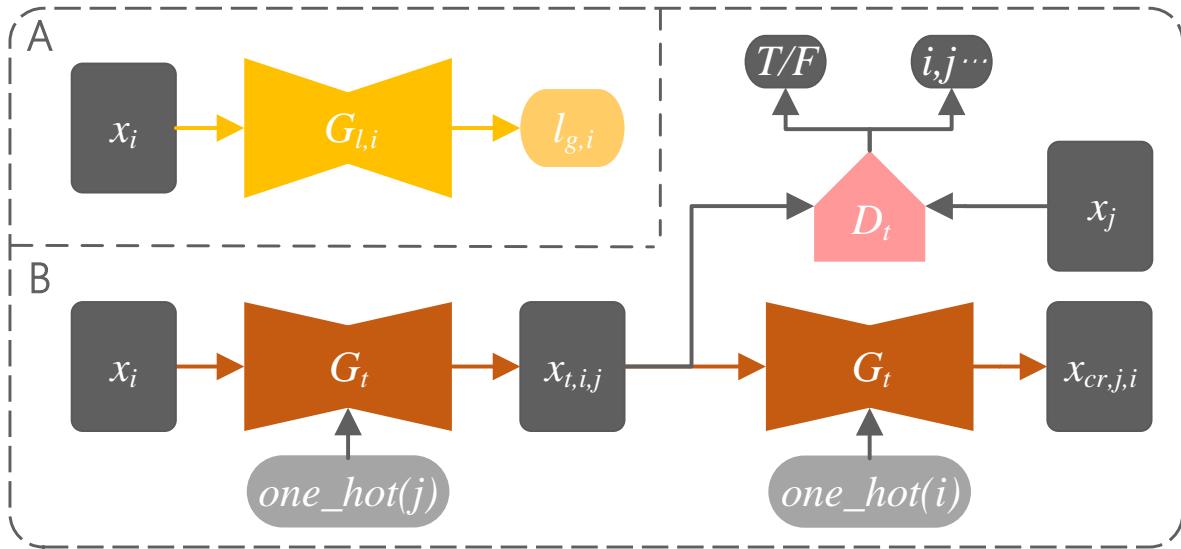


图 4.4: 病灶处理器训练 (A) 和模态转换器训练 (B)。

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 4.4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）[9]， $D_t$ 与在VGG11模型结构的基础上进行了调整适配的 $D$ 结构相同， $G_t$ 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，各自的损失函数如下：

- 模态循环一致性损失

$$\mathcal{L}_{cycle}(G_t) = \mathbb{E}_x [\sum_{j=0} \sum_{i=0, i \neq j} (\|x_i - x_{cr,j,i}\|_2^2)], \quad (4.14)$$

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

$$\begin{aligned}\mathcal{L}_{d,t}(D_t) = \mathbb{E}_{x,s,l} & \left[ \sum_{i=0} (\|d(x_i) - 1\|_2^2 + \|d(x_{t,i,j})\|_2^2 + \right. \\ & \left. \|c(x_i) - i\|_2^2 + \|c(x_{t,i,j}) - i\|_2^2) \right],\end{aligned}\quad (4.15)$$

$$\mathcal{L}_{g,t}(G_t) = \mathbb{E}_{s,l} \left[ \sum_{i=0} (\|d(x_{t,i,j}) - 1\|_2^2 + \|c(x_{t,i,j}) - i\|_2^2) \right]. \quad (4.16)$$

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。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。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 4.4 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。每个模态均由独立的病灶处理器  $G_{l,i}$  来指导合成，典型的训练过程损失为：

$$\mathcal{L}_{seg}(G_{l,i}) = \mathbb{E}_{l,x} [\|l_i - l_{r,i}\|_2^2], \quad (4.17)$$

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处理器和病灶任务的实际需要设计，例如选用 U-net 作为病灶处理器时就可以使用 U-net 原始的加权交叉熵损失函数 [19]。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额外的调整适配。

## 4.6 合成数据集的构建

如图 4.5 所示，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

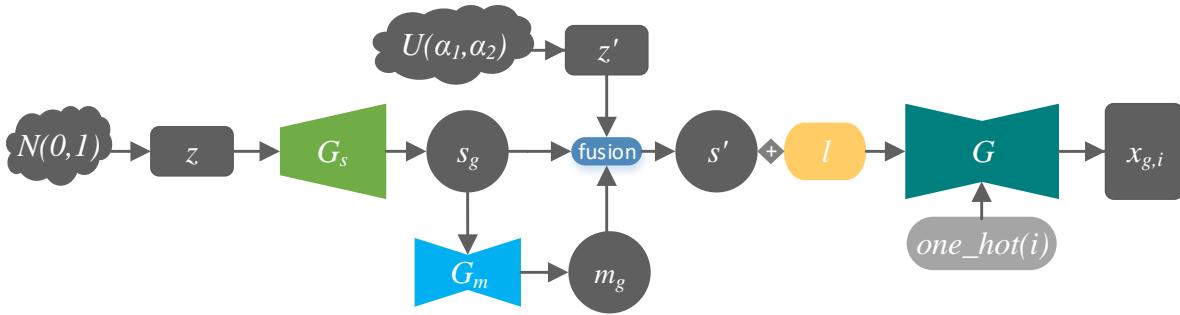


图 4.5: 构建合成数据集。

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，再将两者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，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、翻转或直接随机生成，以获得随机病灶标签集。生成的结构特征图与从随机病灶标签集中随机选择的标签融合时，像训练阶段一样，可以通过掩膜生成器 $G_m$ 从结构特征图中获取的掩膜来筛选出合适的标签。最后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，实际应用时，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，过滤时可以采用鉴别器提供的真假置信度来对GAN的合成结果过滤和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。

## 第五章 合成医学影像的性能评估

上一章节详细介绍了本文解决方案中各个阶段的方法和结合流程，本章节中将在多个数据集上对本文方法进行实验验证。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视觉对比评估，其中实验包括消融实验，多个数据集上与其他方法的对比实验，合成病灶有效性验证实验和合成数据可用性验证实验。本章的评估实验中采用了第三章中介绍的基础方法。

### 5.1 数据集

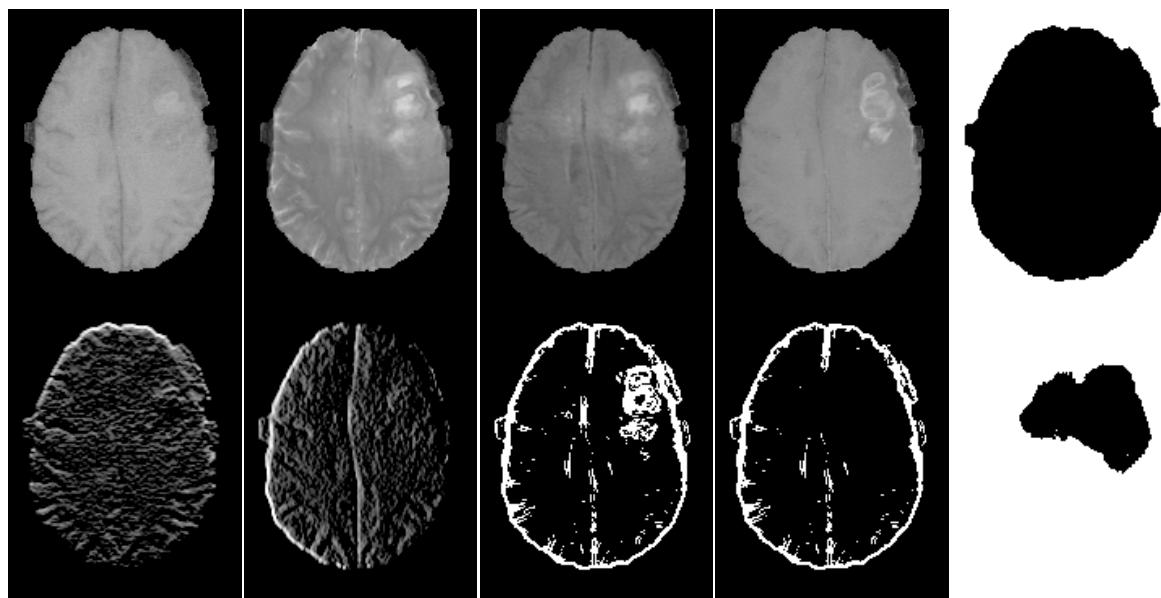


图 5.1: BRATS2015数据集中医学图像和结构特征图提取过程产出的图像。其中子图依次为T1、T2、T1c、Flair四个模态的MRI、从Flair提取出来的掩膜、从Flair采用Sobel算子提取出来的水平向和垂直向边缘检测结果图、从Flair提取出来的结构特征图、采用肿瘤掩膜去掉肿瘤病灶结构的结构特征图、肿瘤分割标签二值转换的肿瘤掩膜。

本研究在六个公开数据集中对提出的方案进行了实验验证，涵盖了彩色照片、X-

ray、MRI、CT等常见的成像模态，涵盖了脑部、视网膜、肺部截面、胸部等多个成像部位，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处理任务。这六个公开数据集的详细情况和预处理方式如下：

- **BRATS2015数据集<sup>2</sup>** [92]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 $184 \times 144$ 。
- **Kaggle Chest X-ray数据集<sup>3</sup>**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 $384 \times 127$ 到 $2772 \times 2304$ 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 $512 \times 512$ 。
- **Kaggle Lung CT数据集<sup>4</sup>**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 $512 \times 512$ 。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<sup>5</sup>**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 $565 \times 584$ ，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 $512 \times 512$ 。
- **FIRE视网膜图像数据集<sup>6</sup>**该数据集包含268张 $2912 \times 2912$ 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 $512 \times 512$ 。
- **天池全球数据智能大赛(2019)数据集 (TC Lung CT)**<sup>7</sup>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

<sup>2</sup><https://www.smir.ch/BRATS/Start2015>

<sup>3</sup><https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>

<sup>4</sup><https://www.kaggle.com/kmader/finding-lungs-in-ct-data/data/>

<sup>5</sup><http://www.isi.uu.nl/Research/Databases/DRIVE/>

<sup>6</sup><https://projects.ics.forth.gr/cvrl/fire/>

<sup>7</sup><https://tianchi.aliyun.com/competition/entrance/231724/information>

大泡, 5-索条, 31-动脉硬化或钙化, 32-淋巴结钙化, 33-胸膜增厚)。本研究中的实验仅进行了其中肺结节病灶的生成和检测。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肺结节的位置标签。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检测标签图。

图 5.1 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 5.2 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)数据集中的医学图像、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。

## 5.2 实验设置

### 5.2.1 训练设置

本研究中实验采用的主要硬件环境为一个NVIDIA Tesla V100GPU集群, 软件环境为Python3.6、CUDA8.0和TensorFlow1.13.2<sup>8</sup>。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 [93], 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, 即卷积核初始化参数采用以前层通道数的倒数为均值, 以 $1e-5$ 为方差的正态分布随机采样。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至少4次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数据集作为训练数据, 以其肿瘤分割

---

<sup>8</sup><https://www.tensorflow.org/>

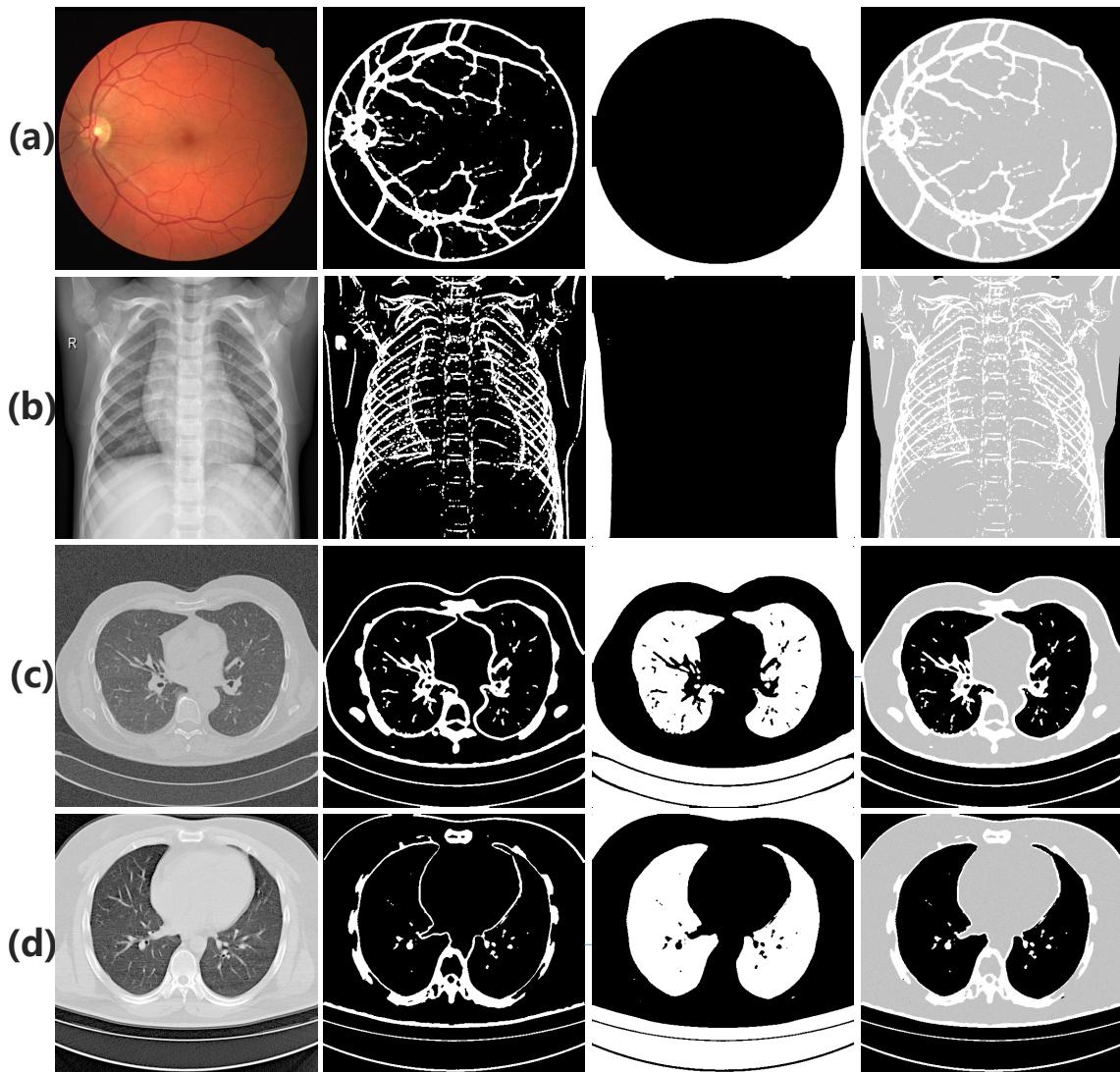


图 5.2: 各数据集中的医学图像和提取的结构特征图。(a) DRIVE 视网膜图像数据集。(b) Kaggle Chest X-ray 数据集。(c) Kaggle Lung CT 数据集。(d) 天池全球数据智能大赛(2019)数据集。以上子图中从左到右依次为原图、结构特征图、掩膜、融合了随机噪声的结构特征图。

标签为输入的病灶标签，合成T1、T2、T1c、Flair四个MRI模态，使用模态转换器提供模态配准损失，使用一个U-net [19]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 5.1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。
- **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数据集作为训练数据，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 [79]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)数据集作为训练数据，将肺结节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 [85]作为病灶检测器提供病灶生成指导损失，无模态配准损失，根据输入的肺结节检测标签图合成具有对应肺结节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。本研究仅进行了该数据集中肺结节病灶的单类病灶生成和单类别检

测实验。

### 5.2.2 评估指标

本研究以SkrGAN [73]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），和Freshet Inception距离（FID） [94]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 [95]和均方误差（MSE）来评估分割结果，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，使用正确率（Accuracy）来评估本研究的分类结果，使用平均精度（AP，Average Precision）来评估检测结果。

## 5.3 多模态医学影像合成方法的消融实验

表 5.1: 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	√	√	√	√	√

如表 5.1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影响。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**通过是否适宜掩膜对输入标签过滤来验证掩膜过滤的作用。

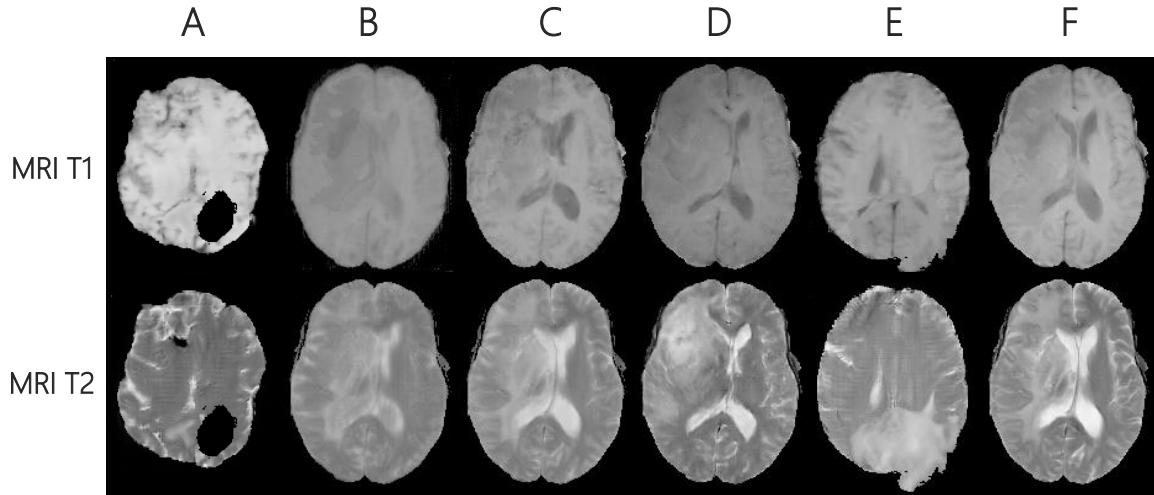


图 5.3: 消融实验的合成图像。B、C、E、F均采用了相同的结构特征图作为输入。

图5.3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，但不符合大脑的结构特征。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C没有模态配准损失，生成的图像的配准效果不是很好，尤其在边缘细节处。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散乱随意、过渡夸张，与输入的病灶标签不吻合。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吻合的病灶信息，配准度高，合成图像的整体质量也最高。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5.4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力，其中，采用结构特征图作为输入对评估结果的提升最为突出，其次是对结构特征图融合随机噪声和在对病灶标签过滤之后添加

病灶生成指导损失产生的提升。图中的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。

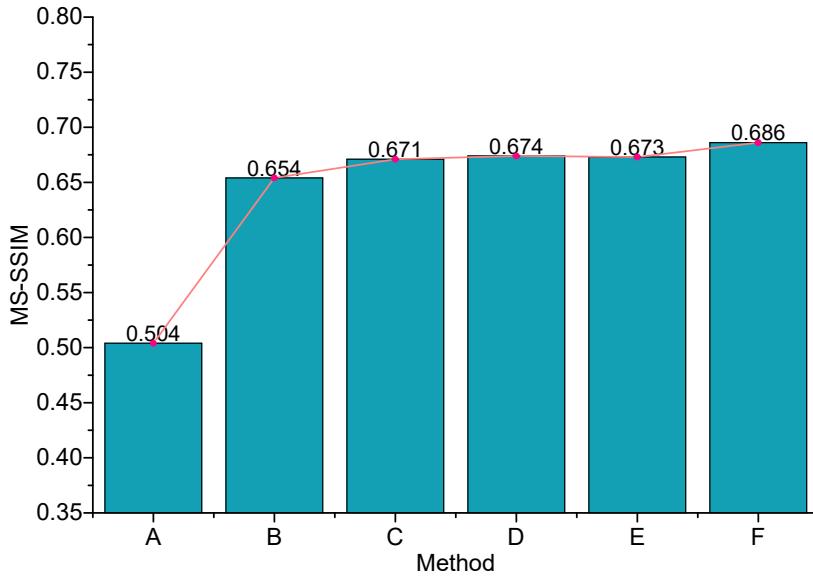


图 5.4: 脑肿瘤MRI数据集上消融实验的量化评估结果。

## 5.4 多个数据集上与其他方法的评估对比

如表 5.2 和 表 5.3 所示，本研究对提出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 5.2 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 5.3 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 5.2 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，本文方法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 5.3 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、无额外的病灶损失和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，且质量远高于基础 GAN 从随机噪声合成的图像。表 5.3 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远高于采用随机噪声为输入的基础 GAN 方法，相同条件下本文的方法合成图像整体质量优于当前最好的 SkrGAN 方法。

综合表 5.2 和表 5.3 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转处理，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 SkrGAN 的草图和其他随机噪声输入，本文方法采用的结构特征图、自监督预训练、病灶损失等多种举措使得本文方法的合成医学影像质量更高，与真实图像更接近。

表 5.2：不同数据集上合成图像质量的量化评估（一）

Dataset	Metric	Ours	SkrGAN [73]	DCGAN [50,73]	ACGAN [15,73]	WGAN [73,96]	PGGAN [73,94]
Kaggle Chest	MS-SSIM ↑	<b>0.597</b>	0.506	0.269	0.301	0.401	0.493
X-ray	FID ↓	<b>102.5</b>	114.6	260.3	235.2	300.7	124.2
Kaggle Lung	MS-SSIM ↑	<b>0.473</b>	0.359	0.199	0.235	0.277	0.328
CT	FID ↓	<b>66.91</b>	79.97	285.0	222.5	349.1	91.89

表 5.3: 不同数据集上合成图像质量的量化评估 (二)

Dataset	Metric	Ours <sup>+</sup>	Ours	SkrGAN*	GAN#
<b>DRIVE+FIRE</b>	MS-SSIM ↑	-	<b>0.607</b>	0.584	0.392
	FID ↓	-	<b>30.13</b>	37.91	227.41
<b>BRATS2015 MRI</b>	MS-SSIM ↑	<b>0.692</b>	<b>0.686</b>	0.653	0.504
	FID ↓	<b>20.15</b>	<b>21.87</b>	28.76	124.53
<b>TC Lung CT</b>	MS-SSIM ↑	-	<b>0.676</b>	0.667	0.543
	FID ↓	-	<b>27.40</b>	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 [50]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 [73]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。

## 5.5 合成病灶的有效性评估

### 5.5.1 脑肿瘤MRI合成

本实验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 5.4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。

表 5.4: 肿瘤分割器对不同测试数据的分割结果

Training Dataset	Test Dataset	MSE	Dice Score
real	real	<b>0.027</b>	<b>0.915</b>
real	synthetic	0.098	0.838

图 5.5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。

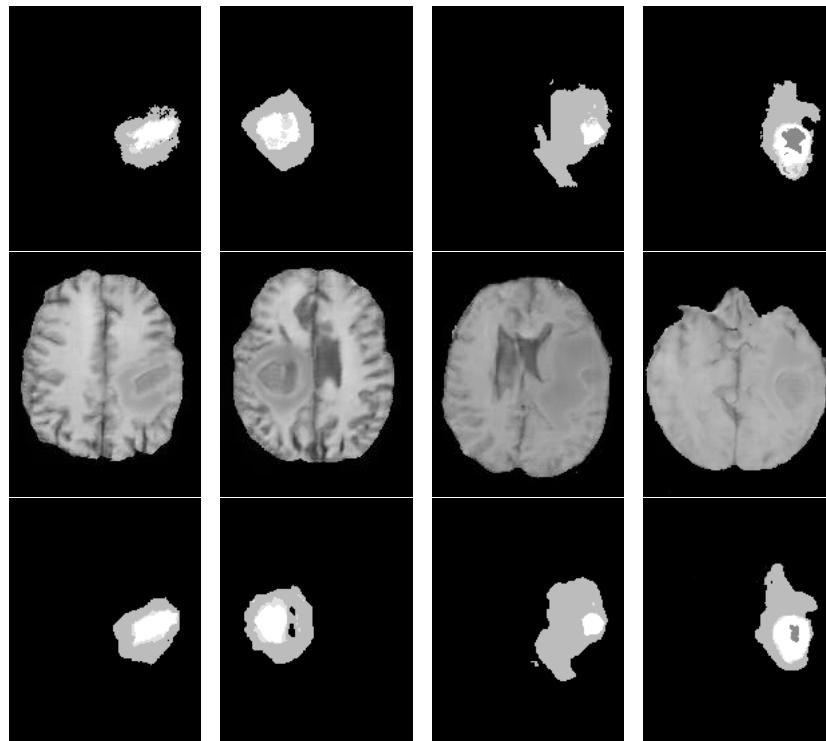


图 5.5: 脑肿瘤分割器对合成MRI的分割结果。从上至下三行依次为输入的肿瘤分割标签、合成的脑肿瘤MRI的T1模态、采用在真实数据集上训练并表现优秀的分割器对第二行合成MRI的分割结果。

## 5.6 合成医学影像的可用性评估

### 5.6.1 合成脑肿瘤MRI数据用于肿瘤区域分割任务

如表 5.5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。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 5.5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，但都不会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（Data Augmentation）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 [41]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真实数据训练出的效果。

表 5.5: BRATS2015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	-	0.027	0.915
2	$\times 50\%$	0	0	-	0.041	0.900
3	0	$\times 1$	0	-	0.294	0.708
4	0	$\times 2$	0	-	0.253	0.736
5	0	$\times 3$	0	-	0.251	0.738
6	$\times 10\%$	$\times 1$	0	$\mathcal{S}$	0.036	0.906
7	$\times 10\%$	$\times 2$	0	$\mathcal{S}$	0.034	0.907
8	$\times 10\%$	$\times 3$	0	$\mathcal{S}$	0.033	0.907
9	$\times 20\%$	$\times 80\%$	0	$\mathcal{M}$	0.038	0.904
10	$\times 50\%$	$\times 50\%$	0	$\mathcal{M}$	0.031	0.909
11	$\times 80\%$	$\times 20\%$	0	$\mathcal{M}$	0.028	0.914
12	$\times 1$	$\times 20\%$	0	$\mathcal{M}$	0.025	0.921
13	$\times 1$	$\times 50\%$	0	$\mathcal{M}$	<b>0.020</b>	<b>0.939</b>
14	$\times 1$	$\times 80\%$	0	$\mathcal{M}$	0.021	0.937
15	$\times 1$	$\times 1$	0	$\mathcal{M}$	0.022	0.934
16	$\times 1$	$\times 2$	0	$\mathcal{M}$	0.023	0.931
17	$\times 1$	$\times 3$	0	$\mathcal{M}$	0.023	0.930
18	$\times 1$	0	$\times 20\%$	$\mathcal{M}$	0.027	0.917
19	$\times 1$	0	$\times 50\%$	$\mathcal{M}$	0.025	0.920
20	$\times 1$	0	$\times 80\%$	$\mathcal{M}$	0.026	0.920
21	$\times 1$	0	$\times 1$	$\mathcal{M}$	0.026	0.919
22	$\times 1$	0	$\times 2$	$\mathcal{M}$	0.026	0.919
23	$\times 1$	0	$\times 3$	$\mathcal{M}$	0.027	0.918
24	$\times 1$	$\times 1$	0	$\mathcal{R}$	0.161	0.795
25	$\times 1$	$\times 1$	0	$\mathcal{S}$	<b>0.020</b>	<b>0.940</b>

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

### 5.6.2 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的二十分之一，因此尽管在结构特征图粗标签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 5.6中SkrGAN\*所示，当采用与本文方法相同的数据集对SkrGAN进行近似复现后再对比时，本文的方法在各个指标上都更优。

表 5.6: 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	<b>0.8464</b>	0.9513	<b>0.9762</b>
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	<b>0.9518</b>	0.9749

\* 表示复现的SkrGAN的评估结果。由于SkrGAN [73]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。

### 5.6.3 合成胸部X光线数据用于肺炎分类任务

表 5.7: 胸部X光线合成数据可用性验证实验

Train Data	Test Data	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	<b>0.818</b>

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。然后本研究使用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了一个胸部X-ray肺炎分类模型，再在Kaggle Chest X-ray测试集上对模型进行分类测试，并将该模型和只采用Kaggle Chest X-ray数据集训练的模型的测试结果进行对比。结果如表5.7所示，可见在该数据集上，本文方法的合成数据在分类任务中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。

### 5.6.4 合成肺部CT数据用于肺结节检测任务

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结节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结节检测模型，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。结果如表5.8，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。

表 5.8: 肺部CT合成数据可用性验证实验

Train Data	Test Data	AP
TC Lung CT训练集	TC Lung CT测试集	0.574
TC Lung CT训练集+20000张 合成图像	TC Lung CT测试集	<b>0.592</b>

## 5.7 合成影像的视觉评估

### 5.7.1 合成结构特征图多样性视觉评估

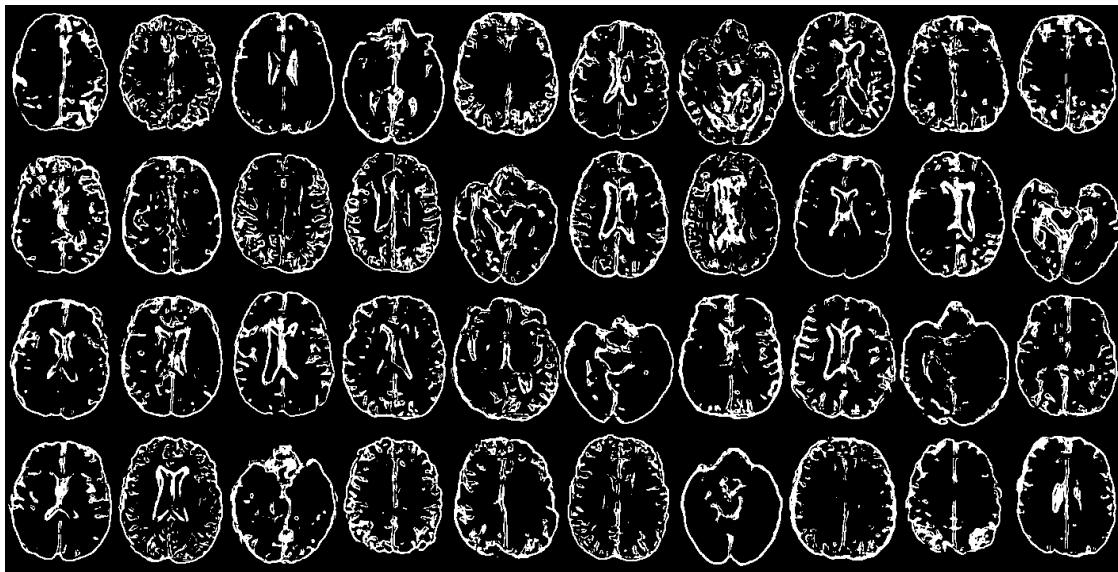


图 5.6: BRATS2015数据集上随机合成的结构特征图示例。

从正态分布随机采样得到如图 5.6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千姿百态、丰富多彩，表现出了非常丰富的多样性。

### 5.7.2 结构特征图正态分布视觉评估



图 5.7: 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。

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 5.7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。

### 5.7.3 合成医学影像视觉评估

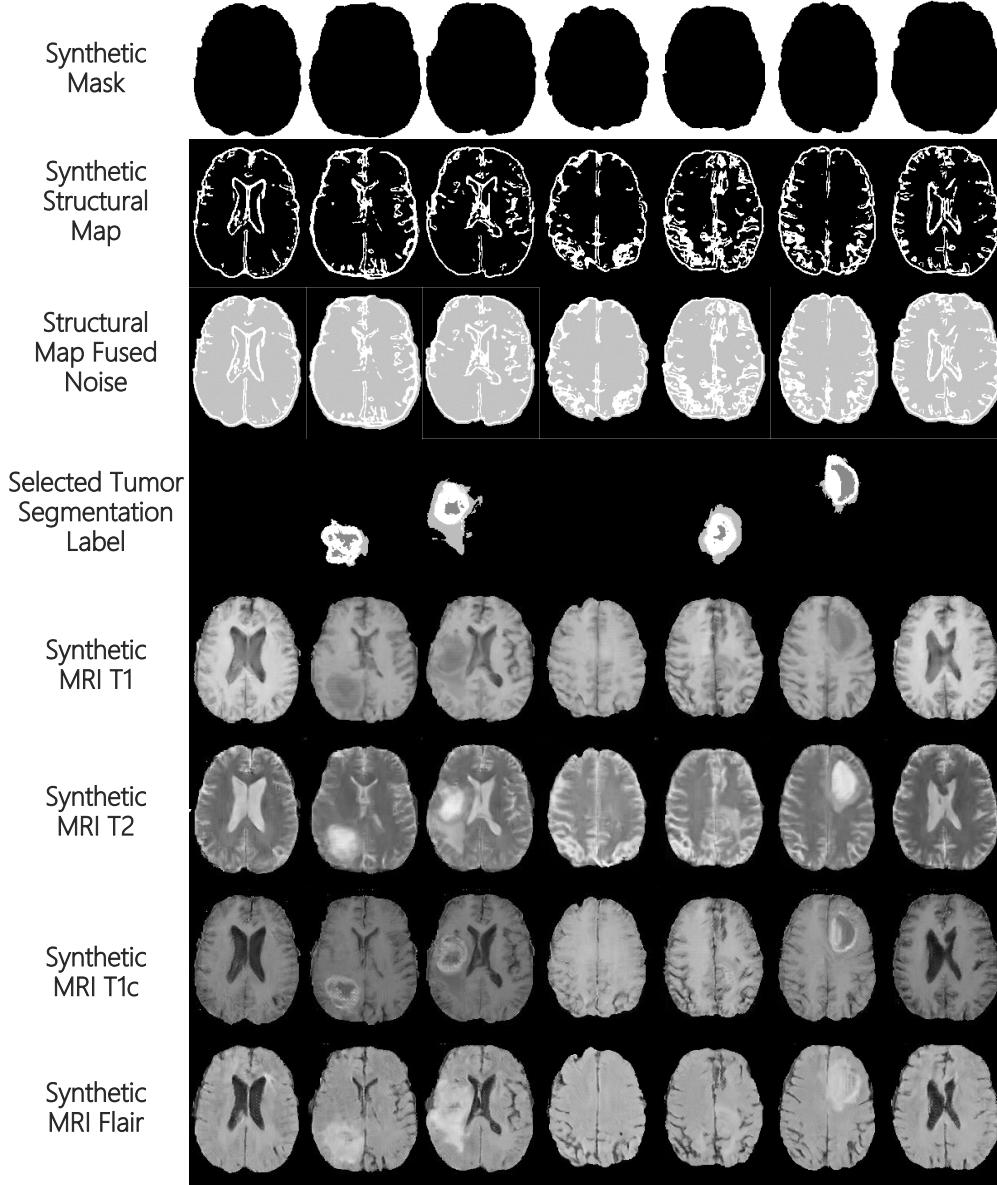


图 5.8: BRATS2015 上合成的结构特征图和多模态MRI。

图 5.8 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影像相互之间具备极高的配准度，同时还根据输入的病灶标签合成了对应的病灶信息。

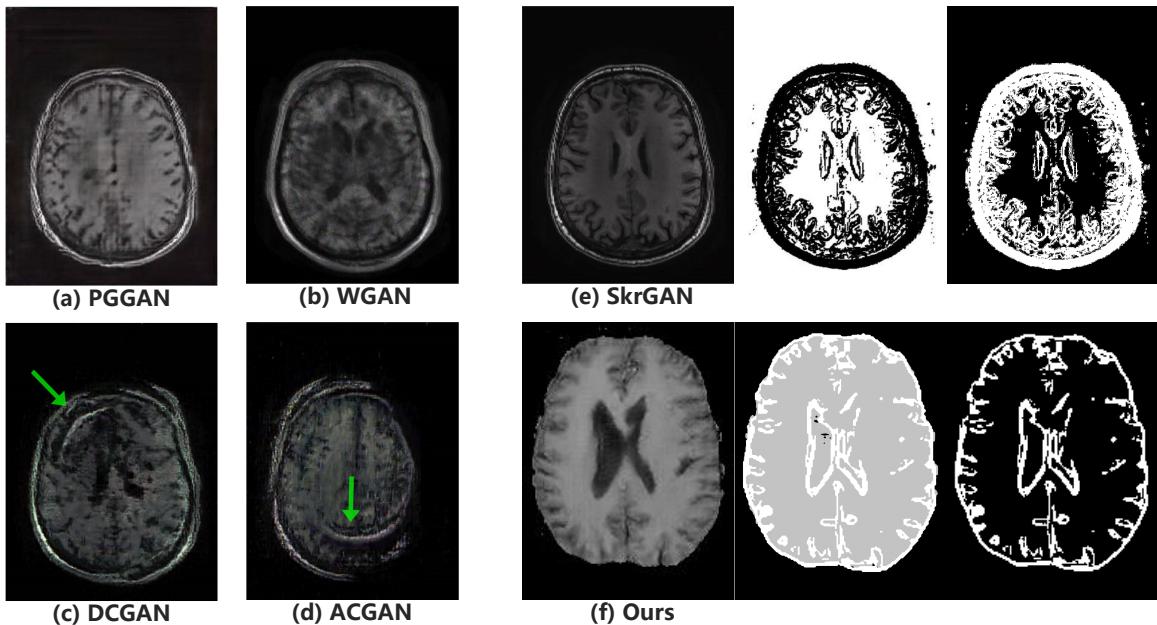


图 5.9: 脑部MRI合成效果对比。(a) PGGAN [73, 94]的合成图像。(b) WGAN [73, 96]的合成图像。(c) DCGAN [50, 73]的合成图像。(d) ACGAN [15, 73]的合成图像，(e) SkrGAN [73]的合成图像、输入的结构特征图和二值反色图。(f) 本文方法的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。(a) - (e) 为在Neuromorphometrics脑MRI数据集上满足一些未知条件筛选出的数据上进行训练的合成结果，(f) 为在类似的BRATS2015上进行训练的合成结果。

表 5.3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 5.9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<sup>9</sup>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 5.9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

<sup>9</sup><http://www.neuromorphometrics.com/>

晰真实、干净整洁，服从输入的结构特征图的合成指导的同时可以有更多样的合成空间。

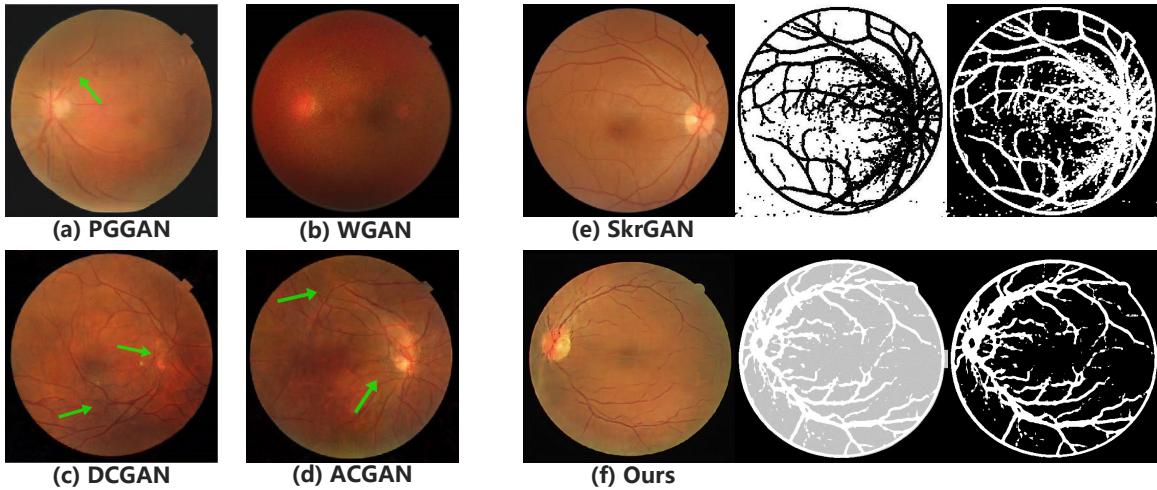


图 5.10: 视网膜合成效果对比。(a) PGGAN [73, 94]的合成图像。(b) WGAN [73, 96]的合成图像。(c) DCGAN [50, 73]的合成图像。(d) ACGAN [15, 73]的合成图像。(e) SkrGAN [73]的合成图像、输入的结构特征图和二值反色图。(f) 本文方法的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。a) - (e) 为在SkrGAN自建数据集上进行训练的合成结果，(f) 为在类似的DRIVE+FIRE数据集上进行训练的合成结果。

表 5.3 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 5.10 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 5.11 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

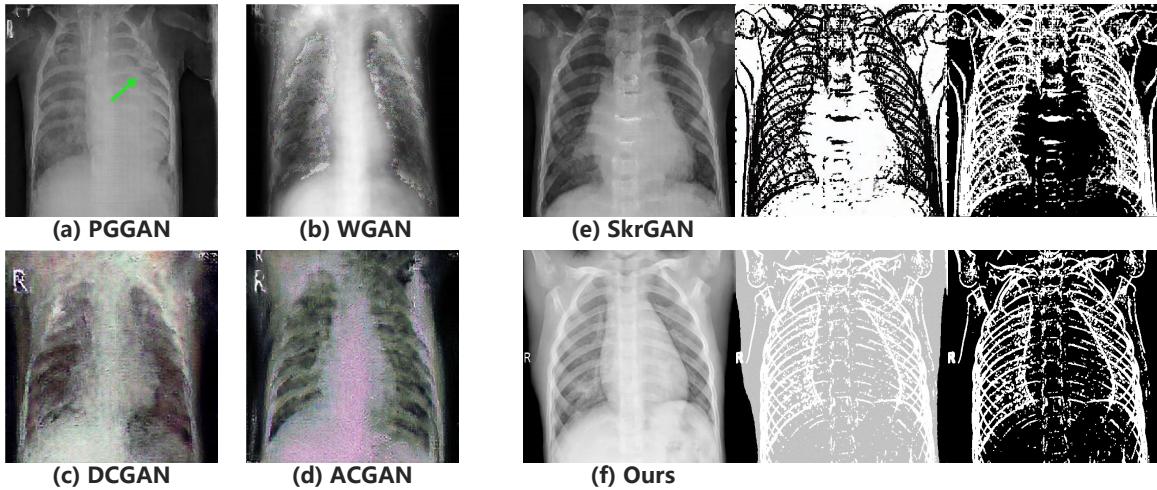


图 5.11: X-ray合成效果对比。(a) PGGAN [73,94]的合成图像。(b) WGAN [73,96]的合成图像。(c) DCGAN [50,73]的合成图像。(d) ACGAN [15,73]的合成图像。(e) SkrGAN [73]的合成图像、输入的结构特征图和二值反色图。(f) 本文方法的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。

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 5.12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明的同时又充满了噪声而显得整体杂乱无章。从合成的CT来看，SkrGAN的合成的肺中缺乏合理的生理结构内容，而本文方法的合成的肺中具有符合生理结构的气管血管等结构。图 5.12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，肺中合成的结构信息丰富详实，整体效果非常逼真。

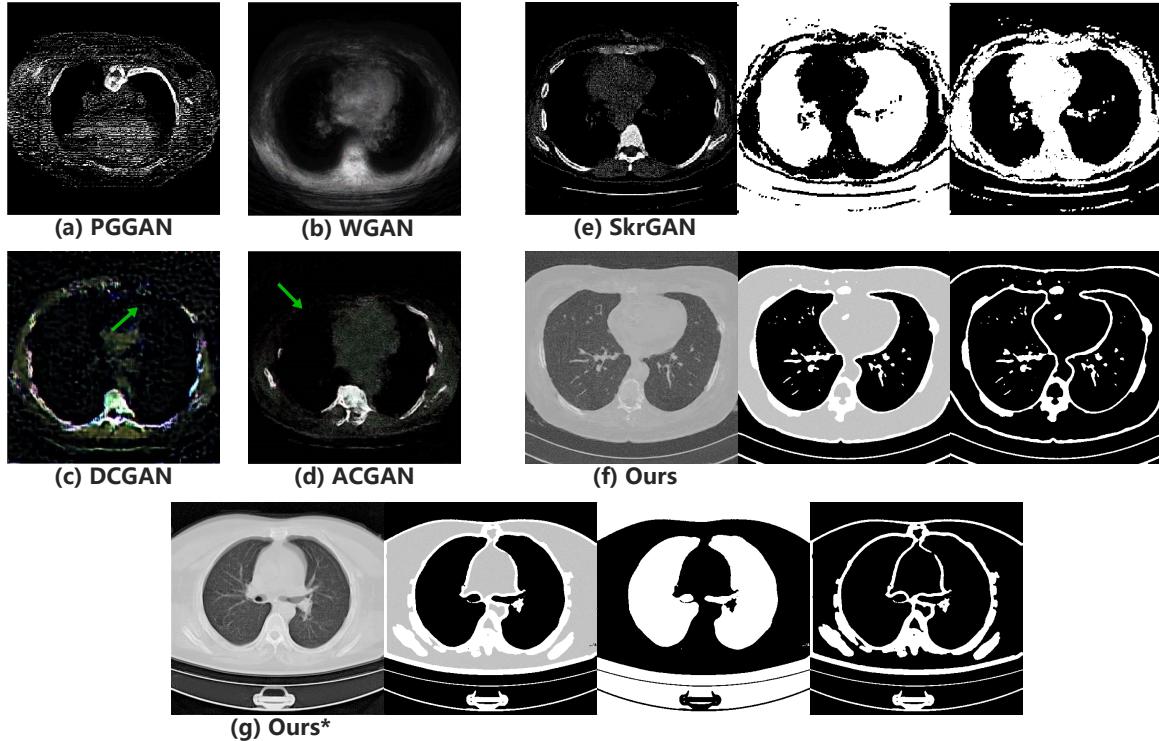


图 5.12: 肺部CT合成效果对比。(a) PGGAN [73, 94]的合成图像。(b) WGAN [73, 96]的合成图像。(c) DCGAN [50, 73]的合成图像。(d) ACGAN [15, 73]的合成图像。(e) SkrGAN [73]的合成图像、输入的结构特征图和二值反色图。(f) 在Kaggle Lung CT数据集上，本文方法的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。(g) 在TC Lung CT数据集上，本文方法的合成图像、输入的融合了噪声的结构特征图和原始结构特征图。(f) - (g) 输出的合成图为标准化后的图像（像素值在0-1之间），视觉上与 (a) - (e) 中未处理的合成结果有一些差异。

## 第六章 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、无法合成精确配准的多模态影像、无法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无法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，提出了一套完整的解决现存问题的解决方案，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，本研究中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，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，采用输入的病灶标签图提供合成病灶的指导信息的同时用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，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，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，再采用基于ACGAN结构的条件生成器合成多模态影像，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。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，以此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，可直接从医学图像中提取出清晰干净的解剖结构信息，无需训练或附加标签数据。

- 本文结合VAE和GAN的特点提出了一种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，可以控制病灶的合成和实现多模态的高度配准。
- 本文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评估合成病灶的有效程度和合成影像的可用程度。

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] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [2] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*, 2014.
- [3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [4] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.
- [5] June-Goo Lee, Sanghoon Jun, Young-Won Cho, Hyunna 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.
- [6] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.
- [7] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.
- [8] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

- [9] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
- [10] Samaneh Azadi, Matthew Fisher, Vladimir G Kim, Zhaowen Wang, Eli Shechtman, and Trevor Darrell. Multi-content gan for few-shot font style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7564–7573, 2018.
- [11] Bo Zhao, Bo Chang, Zequn Jie, and Leonid Sigal. Modular generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 150–165, 2018.
- [12] Xiaodan Liang, Hao Zhang, Liang Lin, and Eric Xing. Generative semantic manipulation with mask-contrasting gan. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 558–573, 2018.
- [13] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797, 2018.
- [14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- [15] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2642–2651. JMLR.org, 2017.
- [16] Irwin Sobel. Camera models and machine perception. *Ph.D.dissertation*, 1970.
- [17] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

- [18] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
- [19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [20] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.
- [21] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.
- [22] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
- [23] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.
- [24] Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 11–19, 2017.

- [25] Adam Paszke, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint arXiv:1606.02147*, 2016.
- [26] Hang Zhang, Kristin Dana, Jianping Shi, Zhongyue Zhang, Xiaogang Wang, Ambrish Tyagi, and Amit Agrawal. Context encoding for semantic segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7151–7160, 2018.
- [27] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [28] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.
- [29] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1925–1934, 2017.
- [30] Md Amirul Islam, Mrigank Rochan, Neil DB Bruce, and Yang Wang. Gated feedback refinement network for dense image labeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3751–3759, 2017.
- [31] Seunghoon Hong, Hyeonwoo Noh, and Bohyung Han. Decoupled deep neural network for semi-supervised semantic segmentation. In *Advances in neural information processing systems*, pages 1495–1503, 2015.
- [32] Maoke Yang, Kun Yu, Chi Zhang, Zhiwei Li, and Kuiyuan Yang. Denseaspp for semantic segmentation in street scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3684–3692, 2018.

- [33] Panqu Wang, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, and Garrison Cottrell. Understanding convolution for semantic segmentation. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 1451–1460. IEEE, 2018.
- [34] Hanchao Li, Pengfei Xiong, Haoqiang Fan, and Jian Sun. Dfanet: Deep feature aggregation for real-time semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9522–9531, 2019.
- [35] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3146–3154, 2019.
- [36] Chenxi Liu, Liang-Chieh Chen, Florian Schroff, Hartwig Adam, Wei Hua, Alan L Yuille, and Li Fei-Fei. Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 82–92, 2019.
- [37] Junjun He, Zhongying Deng, Lei Zhou, Yali Wang, and Yu Qiao. Adaptive pyramid context network for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7519–7528, 2019.
- [38] Chi Zhang, Guosheng Lin, Fayao Liu, Rui Yao, and Chunhua Shen. Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5217–5226, 2019.
- [39] Shun Miao, Sebastien Piat, Peter Fischer, Ahmet Tuysuzoglu, Philip Mewes, Tommaso Mansi, and Rui Liao. Dilated fcn for multi-agent 2d/3d medical image registration. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [40] Hien Van Nguyen, Kevin Zhou, and Raviteja Vemulapalli. Cross-domain synthesis of medical images using efficient location-sensitive deep network. In *International*

- Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 677–684. Springer, 2015.
- [41] Hoo-Chang 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. In *International workshop on simulation and synthesis in medical imaging*, pages 1–11. Springer, 2018.
- [42] 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. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1217–1220. IEEE, 2018.
- [43] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. Learning from simulated and unsupervised images through adversarial training. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2107–2116, 2017.
- [44] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *Advances in neural information processing systems*, pages 700–708, 2017.
- [45] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1857–1865. JMLR.org, 2017.
- [46] Yizhe Zhu, Mohamed Elhoseiny, Bingchen Liu, Xi Peng, and Ahmed Elgammal. A generative adversarial approach for zero-shot learning from noisy texts. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1004–1013, 2018.

- [47] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2018.
- [48] Yunye Gong, Srikrishna Karanam, Ziyan Wu, Kuan-Chuan Peng, Jan Ernst, and Peter C Doerschuk. Learning compositional visual concepts with mutual consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8659–8668, 2018.
- [49] Wenqi Xian, Patsorn Sangkloy, Varun Agrawal, Amit Raj, Jingwan Lu, Chen Fang, Fisher Yu, and James Hays. Texturegan: Controlling deep image synthesis with texture patches. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8456–8465, 2018.
- [50] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [51] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [52] Guim Perarnau, Joost Van De Weijer, Bogdan Raducanu, and Jose M Álvarez. Invertible conditional gans for image editing. *arXiv preprint arXiv:1611.06355*, 2016.
- [53] Asha Anoosheh, Eirikur Agustsson, Radu Timofte, and Luc Van Gool. Combogan: Unrestrained scalability for image domain translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 783–790, 2018.
- [54] Amélie Royer, Konstantinos Bousmalis, Stephan Gouws, Fred Bertsch, Inbar Mosseri, Forrester Cole, and Kevin Murphy. Xgan: Unsupervised image-to-image translation for many-to-many mappings. In *Domain Adaptation for Visual Understanding*, pages 33–49. Springer, 2020.

- [55] Ninon Burgos, M Jorge Cardoso, Filipa Guerreiro, Catarina Veiga, Marc Modat, Jamie McClelland, Antje-Christin Knopf, Shonit Punwani, David Atkinson, Simon R Arridge, et al. Robust ct synthesis for radiotherapy planning: application to the head and neck region. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 476–484. Springer, 2015.
- [56] 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. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6070–6079, 2017.
- [57] Raviteja Vemulapalli, Hien Van Nguyen, and Shaohua Kevin Zhou. Unsupervised cross-modal synthesis of subject-specific scans. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 630–638, 2015.
- [58] Konstantinos Kamnitsas, Christian Baumgartner, Christian Ledig, Virginia Newcombe, Joanna Simpson, Andrew Kane, David Menon, Aditya Nori, Antonio Criminisi, Daniel Rueckert, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In *International conference on information processing in medical imaging*, pages 597–609. Springer, 2017.
- [59] Zhiwen Fan, Liyan Sun, Xinghao Ding, Yue Huang, Congbo Cai, and John Paisley. A segmentation-aware deep fusion network for compressed sensing mri. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 55–70, 2018.
- [60] Rushil Anirudh, Hyojin Kim, Jayaraman J Thiagarajan, K Aditya Mohan, Kyle Champley, and Timo Bremer. Lose the views: Limited angle ct reconstruction via implicit sinogram completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6343–6352, 2018.

- [61] Pedro Costa, Adrian Galdran, Maria Ines Meyer, Meindert Niemeijer, Michael Abràmoff, Ana Maria Mendonça, and Aurélio Campilho. End-to-end adversarial retinal image synthesis. *IEEE transactions on medical imaging*, 37(3):781–791, 2017.
- [62] Juan Eugenio Iglesias, Ender Konukoglu, Darko Zikic, Ben Glocker, Koen Van Leemput, and Bruce Fischl. Is synthesizing mri contrast useful for inter-modality analysis? In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 631–638. Springer, 2013.
- [63] Zizhao Zhang, Lin Yang, and Yefeng Zheng. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern Recognition*, pages 9242–9251, 2018.
- [64] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 417–425. Springer, 2017.
- [65] Anton Osokin, Anatole Chessel, Rafael E Carazo Salas, and Federico Vaggi. Gans for biological image synthesis. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2233–2242, 2017.
- [66] Xin Yi and Paul Babyn. Sharpness-aware low-dose ct denoising using conditional generative adversarial network. *Journal of digital imaging*, 31(5):655–669, 2018.
- [67] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37(6):1348–1357, 2018.

- [68] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. Generative adversarial networks for noise reduction in low-dose ct. *IEEE transactions on medical imaging*, 36(12):2536–2545, 2017.
- [69] Chenyu You, Guang Li, Yi Zhang, Xiaoliu Zhang, Hongming Shan, Mengzhou Li, Shenghong Ju, Zhen Zhao, Zhuiyang Zhang, Wenxiang Cong, et al. Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *IEEE Transactions on Medical Imaging*, 39(1):188–203, 2019.
- [70] Qing Lyu, Chenyu You, Hongming Shan, and Ge Wang. Super-resolution mri through deep learning. *arXiv preprint arXiv:1810.06776*, 2018.
- [71] 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, 2017.
- [72] Thomas Joyce, Agisilaos Chartsias, and Sotirios A Tsaftaris. Robust multi-modal mr image synthesis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 347–355. Springer, 2017.
- [73] 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. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 777–785. Springer, 2019.
- [74] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *international conference on learning representations*, 2014.
- [75] 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.

- [76] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [77] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.
- [78] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [79] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [80] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [81] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [82] Guillaume Alain, Yoshua Bengio, Li Yao, Jason Yosinski, Eric Thibodeau-Laufer, Saizheng Zhang, and Pascal Vincent. Gsns: generative stochastic networks. *Information and Inference: A Journal of the IMA*, 5(2):210–249, 2016.
- [83] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [84] Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.

- [85] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.
- [86] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
- [87] Ross Girshick. Fast r-cnn. *international conference on computer vision*, pages 1440–1448, 2015.
- [88] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.
- [89] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
- [90] Roberts Lawrence G. Machine perception of three-dimensional solids. *Optical and Electro-Optical Infomation Processing*, 1965.
- [91] Judith MS Prewitt. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1):15–19, 1970.
- [92] 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.

- [93] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [94] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.
- [95] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [96] Martin Arjovsky, Soumith Chintala, and Leon Bottou. Wasserstein generative adversarial networks. *international conference on machine learning*, pages 214–223, 2017.

## 附录一 学术成果

### 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. (第一作者, EI)

[2]. Yili Qu, Chufu Deng, Wanqi Su, Ying Wang, Yutong Lu and Zhiguang Chen, Nong Xiao. Multimodal Brain MRI Translation Focused on Lesions[C]//2020 12th International Conference on Machine Learning and Computing (ICMLC). ACM, 2020. (第一作者, 已录用, EI)

[3]. Yili Qu, Wanqi Su and Chufu Deng, Ying Wang, Yutong Lu, Zhiguang Chen, Nong Xiao. Synthesis of Registered Brain Multimodal MRI with Lesions[C]//29th International Conference on Artificial Neural Networks (ICANN). Springer, 2020. (第一作者, 审稿中, CCF C)

[4]. Yu Wei, Qu Yili, Lu Yutong. A Hybrid MapReduce Implementation of PCA on Tianhe-2[C]//Journal of Physics: Conference Series. IOP Publishing, 2019, 1168(5): 052013. (第二作者, EI)

[5]. Wanqi Su, Yili Qu, Chufu Deng, Ying Wang, Fudan Zheng, Zhiguang Chen. Enhance Generative Adversarial Network By Discrete Wavelet Transform to Denoise Low-Dose CT Images[C]//27th IEEE International Conference on Image Processing (ICIP). IEEE, 2020. (第二作者, 审稿中, CCF C)

### 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, 2019-12-06. (第一作者)

[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, 2019-12-06. (第一作者)

[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,2020-01-14.(第一作者)

[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,2019-02-01.(除导师外第一作者)

[5]. 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,2019-04-05. 卢宇彤,瞿毅力,陈志广.(除导师外第一作者)

[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,2019-12-13.(第二作者)

[7]. 苏琬棋,陈志广,瞿毅力,邓楚富,卢宇彤,肖侬,王莹.一种基于生成对抗网络的多域图像转换方法与系统,已公开发明专利,专利公开号CN110084863A,2019-08-02.(除导师外第二作者)

[8]. 邓楚富,肖侬,卢宇彤,陈志广,瞿毅力,苏琬棋,王莹.基于条件生成对抗网络的多域图像转换方法、系统及介质,已公开发明专利,专利公开号CN110675316A,2020-01-10.(除导师外第二作者)

### 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]. 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.(其他研究人员)

[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.依托项目:面向大数据的高性能计算支撑系统研究,依托项目负责人:肖侬.项目负责人:徐风扬.承担单位:中山大学.(项目组主要成员)

[4].基于GAN的带病灶标签的多模态配准医学影像的合成.中山大学优秀研究生创新发展项目.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)

## 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、柯耀斌同学和余玮师姐的诸多帮助和陪伴！感谢实验室所有老师和同学给我的帮助！感谢家人的理解和支持！

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！

瞿毅力

二零二零年四月