**中文图书分类号：TP391**

**UDC： 004**

**学 校 代 码：10005**

****

**硕士专业学位论文**

**PROFESSIONAL MASTER DISSERTATION**

**论 文 题 目：基于生成式数据增强和深度学习的脑肿瘤图像分类研究**

**论 文 作 者：黄钰凝**

**专业类别/领 域：软件工程**

**指 导 教 师：李建强**

**论文提交日期：2026年6月**

UDC：004 学校代码：10005

中文图书分类号：TP391 学 号：S202175122

**北京工业大学硕士专业学位论文**

**（全日制）**

|  |  |  |
| --- | --- | --- |
| **题 目** | ： | 基于生成式数据增强和深度学习的脑肿瘤图像分类研究 |
| **英文题目** | ： | RESEARCH ON BRAIN TUMOR IMAGE CLASSIFICATION BASED ON GENERATIVE DATA AUGMENTATION AND DEEP LEARNING |

|  |  |  |
| --- | --- | --- |
| **论文作者** | ： | 黄钰凝 |
| **专业类别/领域** | ： | 软件工程 |
| **研究方向** | ： | 大数据与信息服务工程 |
| **申请学位** | ： | 电子信息硕士专业学位 |
| **指导教师** | ： | 李建强 |
| **所 在 单 位** | ： | 计算机学院 |
| **答 辩 日 期** | ： | 2026年5月 |
| **授予学位单位** | ： | 北京工业大学 |

**独 创 性 声 明**

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京工业大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签 名：

日 期： 2026 年 6 月 10 日

**关于论文使用授权的说明**

本人完全了解北京工业大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

（保密的论文在解密后应遵守此规定）

签 名： 日 期： 2026 年 6 月 10 日



导师签名： 日 期： 2026 年 6 月 10 日

# 

# 摘 要

脑胶质瘤为中枢神经系统高异质性肿瘤，准确分级与分型直接关系到预后评估与治疗策略；基于磁共振成像（Magnetic Resonance Imaging, MRI）的自动化分类可为临床提供客观、及时的辅助判读，提升诊疗效率与一致性。该任务通过对病灶形态与纹理等关键表征的深入挖掘与精确表达，实现分级与亚型的准确判定。近年来，数据驱动的深度学习方法凭借自动特征提取能力，在脑肿瘤MRI自动化分类进程中发挥关键作用。然而，临床实景数据存在标注数据稀缺致泛化受限、脑肿瘤与正常脑组织在纹理上高度相似且影像对比度较低，难以稳定提取判别性局部信息和单一MRI模态难以反映分子层面信息。为此，本文以提升数据可用性、加强病灶关键特征关注度、增强特征辨识性为主要研究目标，展开以下三方面研究内容：

（1）**针对临床数据稀缺导致模型训练有效性与泛化能力受限的问题，研究提出基于条件去噪扩散概率模型的合成数据增强策略**。临床样本受隐私与伦理约束、标注成本高及跨设备成像差异等因素影响，形成“量少且分布偏移”的双重约束，从源头限制了深度模型的判别能力。为此，方法由三部分构成：首先，条件扩散生成模块在公开数据集进行大规模预训练以学习脑肿瘤MRI的统计分布，并以分级/亚型等标签为条件生成数据；随后，迁移微调模块以少量临床样本对生成模型进行域自适应微调，使合成影像在对比度、噪声与病灶形态上贴近临床数据；最后，质量评估与筛选模块按“仅生成、仅临床、混合”构建三套数据集，在统一设置下以U-Net开展分割推理，以Dice相似系数等客观指标评估合成样本对下游任务的增益，并结合快速人工复核筛选高质量病例，纳入后续分类数据集，从而系统提升样本多样性与类内变异覆盖。具体流程包括体素归一化、在GLIS-RT和GLIS-RT上预训练并以临床数据微调、U-Net统一生成掩膜、三数据集并行验证与优选等环节，形成可复用的数据生成与评估闭环，为分类模型提供规模化且可控的训练样本，实质性缓解小样本与过拟合问题。

（2）**针对****脑肿瘤与正常脑组织纹理相似、影像对比度较低而难以区分的问题，研究提出****基于注意力的空间通道双分支脑肿瘤图像分类方法。**临床MRI中不同分级和亚型往往仅呈现细微差异，传统深度学习模型虽然可以自动提取特征，但在缺乏放射学先验约束的情况下，难以稳定捕捉这些细粒度差异，导致分类结果在低对比度场景下易波动。为此，该方法引入成像先验与病灶轮廓两类领域知识，通过双分支协同和注意力加权机制引导模型聚焦判别性区域。首先，原图分支直接接收原始切片，依据输入的对比度、纹理分布和整体形态学线索，自动估计疑似病灶区域，生成形态—纹理注意图，以突出具有诊断意义的局部结构与内部异质性。其次，Mask分支以病灶轮廓为先验，重点关注不规则边界、浸润范围和边界模糊区域，生成边界注意力图，从轮廓层面显式刻画肿瘤的侵袭性边界特征和空间外扩特征。随后，形态—纹理注意图与边界注意力图分别加权到原图分支的输入和中间特征表示中，并通过通道注意力与空间注意力的残差式门控融合，在不破坏原始信息完整性的同时，抑制背景噪声与无关组织信号，强化模型对真正具有区分度区域的响应。该策略使网络在单模态条件下同时具备对内部结构和外部边界的联合表征能力，缓解了小样本、低对比度、类间差异细微等场景下的不稳定性，并在分级与相近亚型识别任务中提升分类的准确性与鲁棒性。

（3）**针对****单模态 MRI 分类稳定性不足且临床部署阶段难以持续依赖生物标志物的问题，本文提出“多模态融合教师模型—学生蒸馏”的两阶段框架**。上一个研究内容在单模态条件下通过领域知识引导与双分支注意力提升了判别能力，但其有效性仍受真实场景制约：单模态模型对低对比度与细粒度差异的鲁棒性有限，因此本文提出多模态融合分类方法，但临床部署受时效与可得性约束，难以长期依赖多模态输入或大规模模型推理。因此，所提方法模拟临床“先综合、后决策”的流程：首先，构建融合 MRI 与生物标志物的大模型教师模型，得到经不确定性校准的软判别边界与病灶关注图，形成可靠的患者级判据；其次，在教师输出的温度化软标签与中间表征引导下，对仅用原图的轻量学生进行蒸馏训练，并结合患者级证据聚合以稳定切片层面波动。该方法显式利用任务间依赖与证据可靠性，实现从多源强监督到单源快速推理的迁移，不增加部署阶段的数据与计算负担，同时提升分级与分亚型识别的准确性、稳定性与可解释性；并系统评估不同预训练模型（ImageNet、DINOv2）与冻结策略在小样本条件下的迁移效果，明确其适用范围。该方案在不依赖分割与标志物参与推理的条件下，实现对胶质瘤分级与分亚型的快速、稳健且具可解释性的识别，显著提升准确性、一致性与临床可用性。

关键词：脑肿瘤分类；数据生成；双分支模型；多模态融合；迁移学习

# **Abstract**

Glioblastoma is a highly heterogeneous tumor of the central nervous system, where accurate grading and subtyping directly impact prognosis assessment and treatment strategies. MRI-based automated classification provides objective and timely auxiliary interpretation for clinicians, enhancing diagnostic efficiency and consistency. This task achieves precise grading and subtype determination through in-depth exploration and accurate representation of key features such as lesion morphology and texture. In recent years, data-driven deep learning methods have played a pivotal role in automated brain tumor MRI classification due to their automatic feature extraction capabilities. However, clinical real-world data faces challenges: scarce annotated data limits generalization, brain tumors exhibit high textural similarity to normal brain tissue with low image contrast, making it difficult to reliably extract discriminative local information, and a single MRI modality struggles to reflect molecular-level information. Therefore, this paper focuses on enhancing data usability, increasing attention to key lesion features, and improving feature discriminability. The research is conducted in three main areas:

(1) To address the limitations in model training effectiveness and generalization capability caused by scarce clinical data, this study proposes a synthetic data augmentation strategy based on conditional denoising diffusion probabilistic models. Clinical samples face dual constraints of “low volume and distribution skew” due to privacy and ethical restrictions, high annotation costs, and cross-device imaging variations, fundamentally limiting the discriminative power of deep models. To address this, the method comprises three components: First, the conditional diffusion generation module undergoes large-scale pre-training on public datasets to learn the statistical distribution of brain tumor MRIs, generating data conditioned on labels such as grade/subtype. Subsequently, the transfer learning module performs domain adaptation fine-tuning on the generative model using a small number of clinical samples, ensuring synthetic images closely resemble clinical data in terms of contrast, noise, and lesion morphology. Finally, the quality assessment and screening module constructs three datasets—“generation-only,” “clinical-only,” and “hybrid”—to perform segmentation inference using U-Net under unified settings. Objective metrics like Dice similarity evaluate the synthetic samples' contribution to downstream tasks. Combined with rapid manual review, high-quality cases are selected for inclusion in subsequent classification datasets, systematically enhancing sample diversity and intra-class variability coverage. The specific workflow includes voxel normalization, pre-training on GLIS-RT and GLIS-RT with fine-tuning on clinical data, unified mask generation via U-Net, and parallel validation and optimization across three datasets. This establishes a reusable data generation and evaluation loop, providing scalable and controllable training samples for classification models while substantially alleviating small-sample and overfitting issues.

(2) Addressing the challenge of distinguishing brain tumors from normal brain tissue due to similar textures and low imaging contrast, this study proposes a domain-knowledge-guided automated brain tumor classification method. In clinical MRI, tumors of different grades and subtypes often exhibit only subtle differences. While traditional deep learning models can automatically extract features, they struggle to consistently capture these fine-grained distinctions without radiological prior constraints, leading to unstable classification results in low-contrast scenarios. To address this, the method incorporates two types of domain knowledge: imaging priors and lesion contours. It employs dual-branch coordination and attention-weighted mechanisms to guide the model toward discriminative regions. First, the Original Image Branch directly processes raw slices. Based on input contrast, texture distribution, and overall morphological cues, it automatically estimates suspected lesion regions and generates a Morpho-Texture Attention Map to highlight diagnostically relevant local structures and internal heterogeneity. Second, the Mask branch uses lesion contours as prior knowledge, focusing on irregular borders, infiltration ranges, and blurred boundary regions. It generates a boundary attention map that explicitly characterizes the tumor's invasive boundary features and spatial expansion characteristics at the contour level. Subsequently, the morphology-texture attention map and boundary attention map are respectively weighted into the input and intermediate feature representations of the original image branch. Through residual gated fusion of channel attention and spatial attention, background noise and irrelevant tissue signals are suppressed without compromising the integrity of original information, thereby enhancing the model's response to genuinely discriminative regions. This strategy enables the network to jointly represent both internal structures and external boundaries under unimodal conditions, mitigating instability in scenarios with small samples, low contrast, and subtle inter-class differences. It enhances classification accuracy and robustness in hierarchical and sub-type identification tasks.

(3) Addressing the issues of insufficient classification stability in unimodal MRI and the difficulty of continuously relying on biomarkers during clinical deployment, this paper proposes a two-stage framework: “Multimodal Fusion Teacher Model—Student Distillation.” Previous research enhanced discriminative capabilities under single-modality conditions through domain knowledge guidance and dual-branch attention. However, its effectiveness remains constrained by real-world limitations: single-modality models exhibit limited robustness to low contrast and fine-grained differences. Consequently, this paper introduces a multimodal fusion classification approach. Yet, clinical deployment faces constraints of timeliness and availability, making long-term reliance on multimodal inputs or large-scale model inference impractical. Thus, the proposed method mimics the clinical workflow of “first synthesize, then decide”: First, a large teacher model integrating MRI and biomarkers is constructed to generate uncertainty-calibrated soft decision boundaries and lesion attention maps, forming reliable patient-level criteria. Second, guided by the teacher's temperature-calibrated soft labels and intermediate representations, a lightweight student model trained on raw images undergoes distillation training. Patient-level evidence aggregation is integrated to stabilize slice-level fluctuations. This approach explicitly leverages task dependencies and evidence reliability to achieve transfer from multi-source strong supervision to single-source rapid inference. It does not increase data or computational burdens during deployment while enhancing the accuracy, stability, and interpretability of grading and subtype identification. We systematically evaluate the transfer effectiveness of different pre-trained models (ImageNet, DINOv2) and freezing strategies under small-sample conditions, clarifying their applicability. This approach enables rapid, robust, and interpretable identification of glioma grades and subtypes without relying on segmentation or annotations during inference, significantly improving accuracy, consistency, and clinical utility.

Keywords: Brain tumor classification, Data generation, Bifurcated model, Multimodal fusion, Transfer learning

# 目 录

[摘 要 I](#_Toc215426751)

[**Abstract** III](#_Toc215426752)

[目 录 V](#_Toc215426753)

[第1章 绪论 1](#_Toc215426754)

[1.1 研究背景与意义 1](#_Toc215426755)

[1.2 国内外研究现状 2](#_Toc215426756)

[1.2.1 医学图像生成 2](#_Toc215426757)

[1.2.2 脑肿瘤分类及迁移学习 3](#_Toc215426758)

[1.2.3 多模态脑肿瘤数据分类 6](#_Toc215426759)

[1.3 研究问题与内容 7](#_Toc215426760)

[1.4 论文组织结构 8](#_Toc215426761)

[第 2 章 相关概念及技术介绍 10](#_Toc215426762)

[2.1 脑胶质瘤数据相关知识 10](#_Toc215426763)

[2.1.1 脑胶质瘤MRI数据 10](#_Toc215426764)

[2.1.2 脑胶质瘤生物标志物数据 12](#_Toc215426765)

[2.1.3 脑胶质瘤MRI数据特性 13](#_Toc215426766)

[2.2 扩散概率模型图像生成算法 14](#_Toc215426767)

[2.3 经典图像分割算法 16](#_Toc215426768)

[2.3.1扩展的3D U-Net 16](#_Toc215426769)

[2.3.2 联合的2D-3D CNN 16](#_Toc215426770)

[2.4 图像分类相关技术 17](#_Toc215426771)

[2.4.1 经典分类算法 17](#_Toc215426772)

[2.4.2 注意力机制 18](#_Toc215426773)

[2.5 知识蒸馏技术 18](#_Toc215426774)

[2.6 本章小结 19](#_Toc215426775)

[第3章 基于条件去噪扩散概率模型生成脑肿瘤图像研究 20](#_Toc215426776)

[3.1 引言 20](#_Toc215426777)

[3.2 方法描述 21](#_Toc215426778)

[3.2.1 基于条件去噪扩散概率模型的脑胶质瘤MRI数据生成 21](#_Toc215426779)

[3.2.2生成数据有效性评估 24](#_Toc215426780)

[3.3 实验结果及分析 27](#_Toc215426781)

[3.3.1 数据集 27](#_Toc215426782)

[3.3.2 评价指标 28](#_Toc215426783)

[3.3.3 实验设置 29](#_Toc215426784)

[3.3.4 对比实验 30](#_Toc215426785)

[参 考 文 献 33](#_Toc215426786)

# 

# 第1章 绪论

## 1.1 研究背景与意义

脑胶质瘤是最常见的原发性颅内肿瘤，起源于神经胶质干细胞或祖细胞，占恶性脑肿瘤的 81%[1]。在美国，脑肿瘤的年发病率为6/10万，5年死亡率约为5%[2]。胶质瘤的特征是高度异质性，通常会在诊断时表现为侵袭性强、易转移的肿瘤[3]。诊断通常依赖于影像学检查，如磁共振成像MRI和计算机断层扫描（Computed Tomography, CT）扫描，通过这些手段可以发现肿瘤的位置、大小和形态特征[4]。为了进一步确定肿瘤的性质，通常需要通过手术切除肿瘤组织进行病理学分析，进而确定胶质瘤的亚型和分级[5]。根据肿瘤细胞的分化程度和增殖活性，胶质瘤被分为不同的病理亚型，其中最常见的是星形胶质瘤、少突胶质瘤和室管膜瘤[6]。世界卫生组织的 2021 年中枢神经系统肿瘤分类法将脑肿瘤分为 1 至4级，I级通常是良性的，IV级则是最具侵袭性和致命性的类型[7]。胶质瘤的预后与多个因素密切相关，包括肿瘤的分级、分子特征、患者的年龄和健康状况等[8]。高级别胶质瘤预后较差，患者的中位生存期通常在15个月左右，低级别胶质瘤的预后相对较好，但依然存在复发的风险[9]。IDH突变、1p/19q共缺失等分子特征通常与较好的预后相关，这些分子标志物在临床治疗中具有重要的指导意义[10]。

目前脑肿瘤分类领域主要使用CT和MRI影像学诊断。MRI能够提供更高分辨率的软组织成像，尤其对脑部结构的细节表现优于CT[11]；CT通过X射线扫描获得图像，长期或频繁使用可能会对患者带来辐射风险，MRI则采用磁场和无线电波进行成像，避免了辐射的潜在风险[12]；MRI提供了不同的成像模态，如T1加权成像（T1WI）、T2加权成像（T2WI）、扩散加权成像（DWI）和增强扫描（Gd-enhanced MRI），这些模态可以帮助评估胶质瘤的不同特征，如肿瘤的形态、边界、血供和水肿区域等。综合考虑两种影像的不同特点，本研究基于MRI图像实现高效、准确的脑肿瘤分类任务。

基于上述需求，研究者致力于构建面向临床可用的脑肿瘤自动分类方法，以支持快速病情评估与治疗规划。结合本研究的总体摘要，当前领域主要面临三项关键挑战：（1）临床标注数据稀缺且跨设备存在分布偏移，造成模型训练有效性与泛化能力不足；（2）脑肿瘤与正常脑组织在纹理上高度相似、影像对比度较低，细粒度差异易被淹没，难以稳定提取具有判别力的局部信息；（3）单模态MRI虽提供形态与解剖信息，但难以反映分子层面特征；同时临床部署受时效与可得性约束，难以长期依赖多模态输入或大规模模型推理，导致单模态分类稳定性不足。

针对上述问题，本文提出“先生成增强，再优化分类”的技术路线，并显式利用任务间依赖以形成端到端可落地的系统方案。首先，采用条件去噪扩散模型生成与任务标签一致的3D合成样本，并通过统一的质量评估与筛选机制将高质量生成数据与临床数据混合，缓解小样本与分布偏移带来的过拟合风险，提升可用数据的多样性与覆盖度。其次，在单模态条件下提出领域知识引导的双分支注意力模型：以原图分支聚焦形态/纹理表征、以Mask分支强调不规则边界与外扩特征，并通过通道—空间注意的残差式门控与分支间加权融合，稳定突出低对比度场景下的关键判别区域。最后，考虑临床部署的时效与可得性，构建多模态融合教师—学生蒸馏框架：训练期由融合MRI与生物标志物的大模型教师产生经校准的软判别与关注图，学生仅以原图为输入并在温度化软标签与中间表征对齐的引导下学习；结合患者级证据聚合，有效抑制切片层面波动，实现从多源强监督到单源快速推理的知识迁移。

本文聚焦于脑胶质瘤智能诊断中的两项核心任务：MRI自动化分级/分亚型识别与患者级稳健判别与轻量化推理。研究意义在于通过提出的生成式数据增强、领域知识引导的双分支注意力分类模型以及“多模态教师—单模态学生”蒸馏框架，系统提升小样本、低对比度与细粒度差异场景下的准确性、鲁棒性与校准性；同时在不依赖分割与生物标志物参与推理的前提下，实现快速、轻量化的模型推理。该方法可为临床提供及时、可信的影像学证据与患者级风险分层，支撑个体化治疗方案制定与随访管理，具有明确的应用价值与推广意义。

## 1.2 国内外研究现状

### 1.2.1 医学图像生成

在医学研究中，信息通常是以群体而非个人为单位进行分析的。因此，解决医疗数据共享瓶颈的一个潜在方法是生成合成患者数据，这些数据总体上具有与源数据集相似的统计特性，但不会泄露个人的敏感隐私信息。

最近，生成式深度学习方法，例如生成对抗网络（GAN），已展示出生成逼真的高分辨率图像数据集的能力[13]~[15]。Frid-Adar等人[16]使用三个DCGAN为囊肿、转移和血管瘤生成合成样本，发现生成的样本对病变分类任务有益，与真实训练数据相结合时灵敏度和特异性均有所提高，结果增加到 85.7% 的灵敏度和 92.4% 的特异性。Bermudez等人[17]称神经放射学家发现生成的MRI图像质量与真实图像相当，但在解剖准确性方面存在差异。Calimeri等人[18]利用 GAN 生成的脑部 MRI 在定性和定量评估中均获得了高分。Bowles等人[19]在两个脑分割任务中将 GAN 衍生的合成数据引入训练数据集，在不同条件下 Dice相似系数提高了1% - 5%，并验证在其数据集上训练图像堆栈少于十个时效果最明显。Shin等人[20]训练了一个 GAN 来生成合成的脑肿瘤 MRI，并评估了使用生成的数据训练的后续分割网络的性能，报告结果显示使用合成数据及真实数据混合训练的分割网络Dice值达到0.84，优于仅用其中一种图像的结果。

最近，新的生成式学习方法扩散模型（Diffusion）被研究人员广泛研究，在医学图像生成上展示了突出的能力[21]~[23]。Khader等人[24]使用去噪扩散概率模型生成膝关节MRI图像、阿尔茨海默病MRI图像、肺癌MRI图像和乳腺癌MRI图像，并证明了合成图像可用于自监督预训练。Peng等人[25]提出一个新的扩散概率模型cDPM，通过应用cDPM生成新的解剖一致的 3D脑MRI图像。Rombach等人[26]利用潜在扩散模型生成胸部 X 光片图像，同时提出了一种采样策略，以在生成过程中保护敏感生物特征信息的隐私，并在胸部异常分类任务上对图像进行了评估，结果表明所提出的方法优于基于GAN的方法。Jiang等人[27]]提出了潜伏扩散模型CoLa-Diff，采用了一种协同过滤方法来减少内存消耗的架构。此外，他们考虑将大脑区域掩码作为密度分布的先验，以指导扩散过程。并且实施了一种自动权重自适应技术，以有效利用多模态信息。实验表明 CoLa-Diff 作为促进多模态 MRI 合成的有效工具具有巨大的潜力。Lyu等人[28]使用非条件模型迭代应用扩散模型将输入图像编码到潜在空间，然后从潜在空间逆向扩散过程解码健康状态图像，解码过程以健康状态为条件，并通过从生成的反事实中减去输入图像来生成病变热图。

对于这些应用，生成式学习，尤其是扩散模型，在医学图像生成领域展现了显著的能力，能够生成极为逼真的图像。然而，这一优势也导致了一些潜在问题的忽视，目前，大多数研究集中于提升生成图像的质量和多样性，但对这些生成图像与真实医学图像之间的差异、生成图像在临床任务中的实际应用效果等方面的深入研究相对较少。尽管生成的图像在视觉上接近真实图像，但这些生成图像是否能够有效反映疾病特征、辅助诊断，或在临床实践中与真实数据的表现相媲美，这些问题仍然缺乏系统的评估。生成图像的潜在偏差可能在某些情境下引发误诊风险，尤其是在用于辅助诊断或治疗决策时。因此，生成式学习模型产生的医学图像是否能够提供与真实影像同等的临床价值，需要进一步验证和评估。

### 1.2.2 脑肿瘤分类及迁移学习

深度学习技术已广泛应用于医学图像分类任务[29]~[31]。例如，Fu等人[32]和Wang等人[33]开发了利用注意力机制的深度学习模型，专门用于识别脑 CT 扫描中的脑出血，此外，放射学专业的学生在人机交互实验中使用训练有素的模型时表现出了更高的效率。在脑胶质瘤分类和分级任务领域已有大量研究。根据数据来源，研究可分为两类：使用开放数据集（BraTS[34], TCGA[35]）开发的方法和应用于临床环境私有数据集的模型。

一般来说，在公共数据集上进行的研究显示出较高的性能，如在 BraTS 和 TCGA 数据集上，Zhuge等人[36]使用了公开的数据集BraTS 2018和TCGA LGG，其中包含三种模式的核磁共振成像。他们在三维脑肿瘤分割中筛选出肿瘤面积最大的切片，在受试者水平上分割数据集，并微调 Mask R-CNN 模型，将胶质瘤分为 HGG 和 LGG，准确率达到 89.1%。Banerjee等人[37]使用TCGA-GBM和TCGA-LGG数据集（包含四种模式）中的461个受试者对VGG16和ResNet50模型进行了微调。他们使用BraTs数据集对这些模型进行了评估，将肿瘤分为HGG和LGG。ResNet50的准确率为 72%。Ge等人[38]提出了一种多流二维CNN，利用TCGA数据集对胶质瘤进行分类，准确率达到86.53%，并在BraTS数据集上将肿瘤分为HGG和LGG，准确率达到90.70%。Naser等人[39]使用公开数据集TCIA中的三种核磁共振成像模式，微调了基于ImageNet的预训练VGG模型，用于胶质瘤分级任务。他们在患者层面的准确率达到95%。经过训练的UNet在测试集上的Dice分数的中值和均值分别为0.98和0.84，VGG在分级任务中的AUC为0.97。然而，这样一个完美的模型，即分割模型的中值在 1 左右，分类模型的准确率达到完美水平，在其他研究中很难再次实现。

有一些研究对临床数据进行了实验。Matsui 等人[40]使用了217名受试者的临床数据集，其中包括三种模式的 MRIS。脑肿瘤被分为弥漫性星形细胞瘤lDH-野生型、弥漫性星形细胞瘤lDH-突变型和少突胶质细胞瘤，ResNet的准确率达到58.5%。Waghmare等人[41]使用了包含三种模式的375名受试者的临床数据集。他们将每种模式的相同切片堆叠在一起，在受试者水平上从训练队列中获得15个切片，从测试队列中获得6个切片。使用ResNet50模型进行了微调，F1-Score为 48/%。Xiong等人[42]使用了包含三种模式的211例受试者临床数据集。他们选取了轴向肿瘤面积最大的图像切片，并在肿瘤面积最大的相邻两层各增加了间隔 5 毫米的附加切片。数据集在患者层面进行了划分，并使用 ResNet34 对模型进行了微调，准确率达到 50%。Fasihi 等人[43]使用了包含四种核磁共振成像模式的 58 例受试者临床数据集。实验是在图像层进行的，但值得注意的是，他们在图像层将三维数据拆分为二维数据，这可能会导致数据泄漏问题并产生不准确的结果。他们使用一个预训练模型从图像中提取特征，并将这些特征作为输入来训练一个新的机器学习分类模型，以便将脑肿瘤分为 I 至 IV 级。他们报告了极高的性能，F1 分数高达 99.38%。然而，数据泄露问题可能导致了错误的结果。

通过比较这些相关工作，发现一些使用公共数据的研究取得了非常高的性能，准确率或 F1分数约为 90%[36],[38],[39]。只有一项研究使用了临床数据集[43]，并报告了99.38%的 F1 分数。然而，这一结果很可能受到数据泄露问题的影响，而数据泄露可能导致错误的结果。使用公共数据集取得的高性能可能是由于数据质量好且经过标准化预处理。相比之下，使用临床数据进行的研究在这项任务中的性能往往较低[40]~[42]，通常达到的准确率或 F1 分数约为 50%。

这些研究凸显了在临床数据集上开展研究的必要性，因为胶质瘤分级和分类任务的模型开发仍未解决。目前利用临床数据实现高性能所面临的挑战表明，在这一领域还需要做更多的工作。与这些使用临床数据集的相关研究相比，我们的研究通过结合临床特征和使用流行的计算机视觉基础模型 DINOv2 进行实验来解决分类和分级任务。

深度学习领域越来越多地应用可适用于各种任务的基础模型。在计算机视觉领域，一个显著的例子是 Segment Anything model (SAM)[44]，这是一个基础模型，尤其是在医学图像分割任务中引起了极大的兴趣[45],[46]。另一个具有代表性的模型是DINOv2[47]，这是一种利用Vision Transformer架构的无标签、自馏分方法，其性能明显优于其他自监督学习方法和某些监督学习方法。在DINOv2成功的基础上，Oquab等人开发了DINOv2，这是一个融合了iBOT元素的增强版本，并在包含1.42亿张自然图像的LVD-142 M数据集上进行了预训练。DINOv2保持了DINOv2的优点并做了进一步改进，可用于图像和视频格式的分类、分割、深度估计和图像检索等任务。

基于基础模型的这些优势，研究人员正致力于将 DINOv2 应用于医学图像分析。Baharoon 等人[48]将 DINOv2 模型用于放射图像分析，并使用 DINOv2 在不同模式（X 光、CT 和 MRI）上进行了 100 多次实验，任务包括疾病分类和器官分割。研究结果表明，与监督、自监督和弱监督模型相比，DINOv2 具有良好的跨任务泛化能力。不过，研究人员在整个实验过程中都保持了冻结骨干，这一策略在迁移学习中相当重要。由于其重要性，这种方法的基本机制值得进一步研究。

Prokop 等人[49]广泛评估了 DINOv1 和 DINOv2 等模型，在小样本 X 射线图像分类场景中使用各种预训练 CNN 和视觉变换器 (ViT) 作为特征提取器和多个分类器的组合。研究任务包括 COVID-19 识别任务和肺结核识别任务。在这两项任务中，作为特征提取器的 ViT 在几乎所有情况下都优于基于 CNN 的模型。这表明，DINOv2和MAE视觉变换器可能是度量学习模型中特征提取器的良好选择。然而，在COVID-19识别任务中，DINOv2-ViT-B/14的平均AUROC落后于DINOv2-ViT-B/8。

Huix等人[50]在四个医学图像分类数据集上对五个基础模型进行了验证。他们的研究表明，使用DINOv2作为迁移学习的基础模型具有很强的迁移能力，有可能取代ImageNet预训练在医学分类任务中的作用。此外，他们的研究还深入探讨了冻结机制，发现冻结基础模型经常导致多个数据集的性能下降。

### 1.2.3 多模态脑肿瘤数据分类

最新研究结果表明，在神经胶质瘤患者组织和血液中，多种生物标志物可用于神经胶质瘤的辅助诊断和预后评估。Lerousseau 等人[51]将MRI图像特征与WSI图像特征结合，采用多通道网络进行分类，实验结果表明，融合特征的分类准确率分别比单一模态特征提高了4.9%和3.8%。Cho 等人[52]提取了MRI图像的T1、T2、T1C和T2FLAIR四种模态数据进行特征融合后进行分类，准确率达到0.89。Hsu 等人[53]使用两个独立的网络分别提取WSI图像特征和MRI图像特征，然后基于WSI图像特征的结果进行投票，得出了最终分类结果。实验表明，基于标签融合的预测方法优于单一方法。Xiao 等人[54]提取了图像感兴趣区域的不同类型的放射组学特征，并对VGG-16模型进行了微调以提取深度学习特征，然后融合这两类特征进行机器学习分类，结果表明，混合特征的分类准确率高于单一模态特征。Gao 等人[55]结合MRI图像特征和生物标志物（如GFAP、Ki67、S100）的信息，通过逻辑回归和支持向量机等机器学习方法对胶质瘤进行分级。

然而，目前大多数将生物标志物与MRI多模态融合的研究主要集中在预测患者生存时间[56]~[60]，并且这些研究往往局限于少量的生物标志物特征，通常仅包括IDH、Ki67和1p/19q等几种标志物。这些标志物虽然在胶质瘤的临床诊断中具有一定的预测价值，但它们并未全面涵盖胶质瘤的生物学特征，导致现有的研究存在一定的局限性。生物标志物的数量较少，且大多数研究采用的标志物与肿瘤的分子异质性、基因突变等复杂特征的关系尚未深入探讨。尽管现有研究已经在生存预测方面取得了一定进展，但将生物标志物与MRI图像结合进行胶质瘤分类的研究较为稀缺。胶质瘤的分类不仅依赖于影像学特征，还需要充分考虑生物标志物所提供的结构和功能信息。现有的多模态融合方法多集中于生存时间的预测，而对于胶质瘤的具体分类任务，尤其是在亚型识别和肿瘤分级方面的应用，尚未得到充分探索。这一缺乏针对性研究的现状使得基于多模态信息的胶质瘤分类方法尚未达到最优效果，存在进一步提升的空间。本研究计划通过引入更多种类的生物标志物来丰富现有的特征集合，并将其与MRI图像进行深度融合，从而提高分类的准确性和可靠性。

## 1.3 研究问题与内容

本研究针MRI脑肿瘤临床数据稀缺及传统深度学习模型一步到位实现肿瘤分类任务效果不佳问题进行深入探讨研究，具体有三方面关键问题：

（1）**临床数据稀缺**。脑肿瘤 MRI 的获取与共享受多重因素制约：1）患者隐私与伦理合规严格，原始 DICOM 的脱敏与跨院流通门槛高；2）体素级标注成本昂贵、专家间一致性有限，3D 体数据标注尤为耗时；3）跨设备、扫描协议与中心差异导致数据分布碎片化，且类别极不均衡（罕见亚型、边界样本稀少），模型易过拟合并对域外数据失配；4）随访与质控信息缺失，难以沉淀可复用高质量训练集。上述因素共同造成“量小、噪声大、分布偏移”的现实环境，限制了传统深度学习的训练效率与泛化能力。为缓解瓶颈，隐私友好的条件生成与合成数据成为重要补充路径：在不暴露个体信息前提下，学习真实影像统计特性，按分级/亚型与采集条件进行可控增广，扩大样本规模与类内变异覆盖，为小样本建模与跨域泛化提供数据基础。

（2）**脑肿瘤与正常脑组织在纹理上高度相似，且影像对比度较低**。判别性局部线索，如边缘、细纹理、局部结构等，易被背景结构、偏置场与噪声淹没。肿瘤具有显著空间异质性与不规则边界，水肿带与坏死区进一步破坏灰白质界面，使病灶—非病灶统计分布高度重叠。传统的的卷积式建模依赖有限感受野与全局池化，难以在复杂背景中自动定位最具判别力区域，对无关组织敏感且缺乏对边界模糊带的显式刻画；在小样本条件下，注意聚焦易漂移，表现为低对比度病灶漏检、相邻组织误判与跨中心泛化不稳。此外，真实采集中病灶尺寸差异、姿态变化与采集参数不一致，使基于固定阈值或手工 ROI 的方案难以泛化，进一步放大特征混淆与类别边界重叠。核心症结在于缺少同时面向形态与轮廓的协同建模与先验约束，无法稳定聚焦关键区域并抑制无关干扰。因此，如何协同建模局部细节与全局上下文、抑制无关区域干扰并稳定聚焦病灶边界，成为该问题的核心挑战。

（3）**单一MRI图像模态数据难以精确分类。**胶质瘤的分类不仅依赖于影像学特征，还需充分整合生物标志物提供的分子层面信息。现有的多模态融合研究多侧重于生存预测任务，而在具体分类问题，在亚型识别和肿瘤分级方面的应用上，尚未得到深入探索。传统单模态方法，如单一MRI图像或单一生物标志物特征，往往无法全面捕捉肿瘤的异质性，导致分类精度有限，在不同亚型的精确识别和肿瘤等级的划分上存在较大挑战。影像学特征主要呈现空间形态与解剖结构，而分子层面的生物标志物刻画肿瘤的遗传与功能特性。二者在信息粒度与语义层面具有天然互补，弥补单模态数据的局限性。本研究将重点探讨如何在配对稀缺、模态缺失与域移并存的条件下，有效整合影像—分子互补信息、缓解粒度不匹配与不确定性，并保证小样本场景下的稳定判别，以期实现更为精确的胶质瘤肿瘤分型和分级。

针对上述三方面关键问题，本文总体研究框架图如图1-1所示。

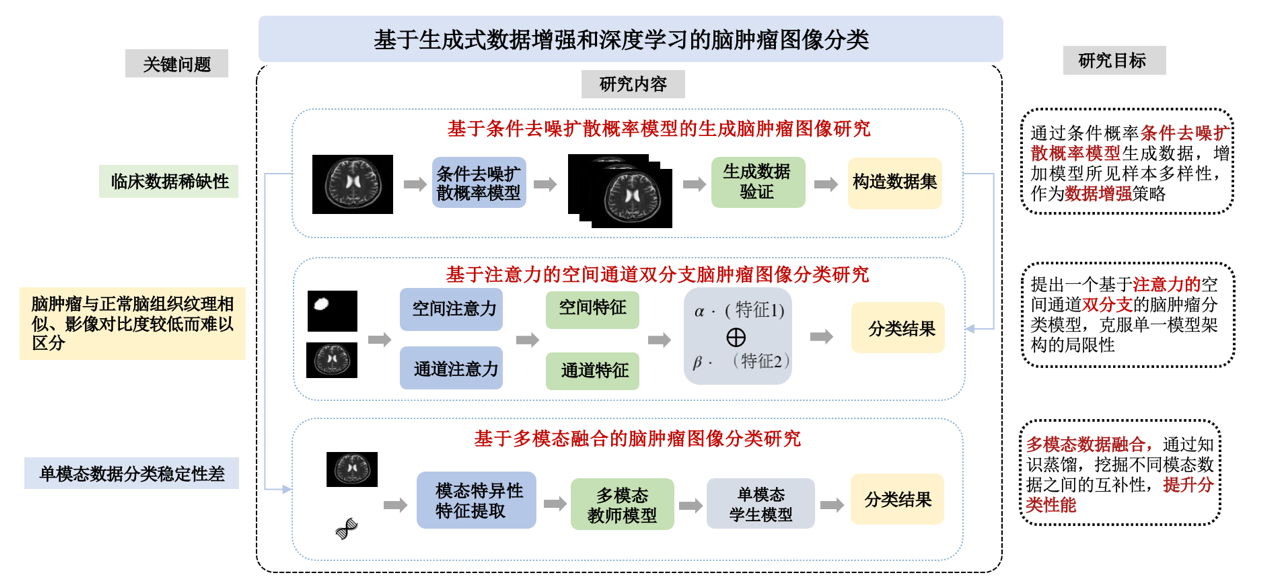


图1-1 本文关键问题、研究内容与研究目标

Fig. 1-1 Key questions, contents and objectives of this paper

## 1.4 论文组织结构

基于生成式数据增强和深度学习的脑肿瘤图像分类研究共包括五个章节，具体组织结构如下：

第一章：绪论。阐明脑肿瘤图像自动分类研究背景与意义，并梳理总结了现有脑肿瘤自动识别领域的国内外研究成果和进展，在此基础上，根据临床中的脑胶质瘤MRI数据特性，提出本文所解决的关键问题、研究内容和研究目标。

第二章：相关概念及技术介绍。对脑胶质瘤MRI图像相关知识进行详细介绍，然后介绍了胶质瘤图像增强、图像分割、图像分类和知识蒸馏的相关技术。

第三章：基于条件去噪扩散扩散概率模型的生成脑肿瘤图像增强研究。针对临床数据稀缺的问题，本章详细介绍了提出的条件去噪扩散扩散概率模型的生成脑肿瘤图像增强方法，并构造三个数据集进行实验分析，以验证本章研究内容的有效性。

第四章：基于注意力的空间通道双分支脑肿瘤图像分类研究。针对脑肿瘤与正常脑组织纹理相似、影像对比度较低而难以区分问题，本章详细介绍了提出的基于注意力的空间通道双分支脑肿瘤图像分类方法，并基于临床脑胶质瘤MRI分类数据集进行实验分析，以验证本章研究内容的有效性。

第五章：基于多模态融合的脑肿瘤图像分类研究。针对单一MRI图像模态数据分类稳定性差的问题，本章详细介绍基于多模态融合的脑肿瘤图像分类研究方法框架，并基于临床脑胶质瘤MRI分类数据集进行实验分析，以验证本章研究内容的有效性。

# 第 2 章 相关概念及技术介绍

胶质瘤MRI中的判别线索涵盖形态结构、细粒度纹理、边界形态及周围组织上下文，是分级与分亚型识别的主要依据。受低对比度、成像噪声和跨设备差异影响，上述线索易被弱化。本章主要介绍胶质瘤MRI的基本理论与表征要点、图像生成、图像分割、图像分类、知识蒸馏等关键技术，为后续方法设计与实验评估提供技术基础。

## 2.1 脑胶质瘤数据相关知识

### 2.1.1 脑胶质瘤MRI数据

MRI 的射频激发与梯度编码按照特定时序组合形成“序列”，其参数设定决定组织对比、空间分辨率、信噪比与伪影特性。不同序列由不同物理敏感性主导，从而获得互补的影像信息。脑部检查中常用序列包括：对比增强 T1 加权序列，依赖血脑屏障通透性变化与造影剂弛豫效应，突出强化病灶与肿瘤实质；T2 加权序列，对游离水含量敏感，强化水肿、坏死与囊性成分的显示；弥散加权成像（DWI），以水分子布朗运动受限为对比基础，反映细胞密度与微观结构改变；灌注加权成像（PWI），通过动态对比敏感技术估算相对脑血容量、脑血流量与平均通过时间等血流动力学指标，表征肿瘤血供与新生血管特征；液体衰减反转恢复（FLAIR），对脑脊液信号实施抑制，以增强皮质下与室周病变的可见性。

不同于普通2D图像，脑胶质瘤影像数据在临床侧以 DICOM 为主、科研侧常转为 NIfTI，临床脑胶质瘤MRI数据收集主要包括三大步骤：即MRI 扫描仪扫描、数据导出、人工标注，如图2-1所示。

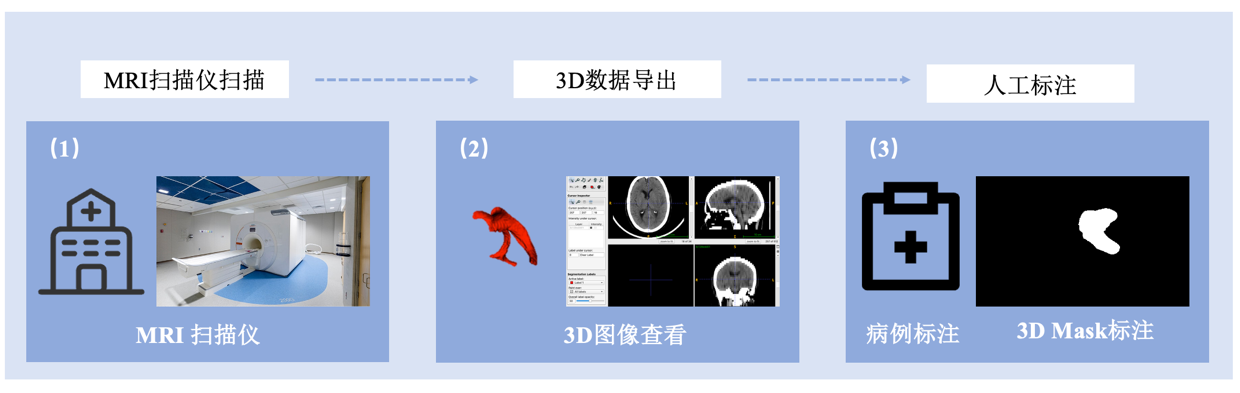


图2-1 临床脑胶质瘤MRI数据收集过程

Fig. 2-1 The process of collecting clinical MRI images of Brain Gliomas

本文三维临床数据的主要采集参数如下：

* T1 加权结构成像：重复时间 TR 2300 ms，回波时间 TE 2.32 ms，翻转角8°；视野 240×240 mm，矩阵 240×240，层厚 0.9 mm，体素 0.9×0.9×0.9 mm；
* T2 加权结构成像：TR 5000 ms，TE 105 ms，翻转角 150°；视野 220×200 mm，矩阵 448×358，层厚 3.0 mm，体素 0.5×0.5×3.0 mm；
* T2 FLAIR 结构成像：TR 5000 ms，TE 387 ms；视野 230×230 mm，矩阵 256×256，层厚 0.9 mm，体素 0.9×0.9×0.9 mm。

由于不同设备与采集协议导致的体素大小与各向异性差异，三维MRI体素数据在空间尺度上不可直接比较：1）几何测量随体素尺寸变化产生系统偏差；2）三维卷积的物理感受野在不同体素间距下不一致，影响跨中心泛化；3）跨序列/跨次检查的三维配准需插值到统一网格，若各向异性显著，易引入层间平滑与台阶伪影并造成掩膜与影像错位；4）厚层方向的部分容积效应加剧边界模糊。为降低上述影响，建模前应统一空间分辨率与方向：选定参考序列T1的 0.9×0.9×0.9 mm 体素作为目标间距，对全部序列做三维重采样至各向同性；影像使用三线性插值，标签使用最近邻插值，保持方向矩阵、起始点与视野一致。

本文聚焦于从三维数据中提取的二维轴向切片，特别是那些由二值掩膜标志的肿瘤感兴趣区域（ROI）尺寸超过30像素的切片。这些切片构成了数据集的基础，满足模型对二维输入的要求。在实验过程中，本文考虑以分辨率214×214像素为参数裁剪脑胶质瘤MRI 2D图像，图像示例如图2-2左侧所示。在筛选完剪裁好的图像中存在扫描运动伪影、化学位移伪影等情况，如图2-2右侧所示。同时存在数据量不足，训练模型欠拟合从而导致准确率低低问题。

本文基于三维 MRI 体数据构建二维轴向切片数据集。首先，以二值化肿瘤掩膜为依据筛选病灶面积大于 30 像素的切片，确保样本包含充分的病灶信息，并据此形成训练与评估的基础样本集合。然后，对候选切片按固定视野进行中心裁剪至 214×214 像素，以统一输入尺度并降低无关背景干扰，如图2-2左。原始脑部切片同时包含肿瘤、正常脑实质、脑室、颅骨及成像背景等多种结构，直接用于建模易引入与任务无关的强边缘与高对比区域，造成特征混淆与决策边界漂移。本研究以肿瘤区域为主要观测对象，如图2-2右。

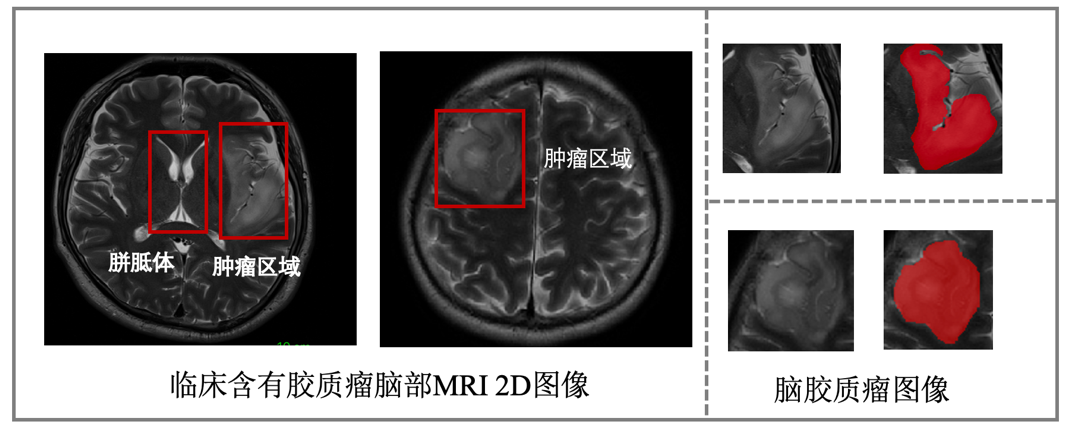


图2-2 脑胶质瘤MRI图像剪裁结果

Fig. 2-3 Cutting results of MRI images of Brain Gliomas

本文使用的数据集包括共九种类别，分别为：WHO II级、WHO III级、WHO IV级；少突胶质细胞瘤、星形细胞瘤、胶质母细胞瘤、弥漫性中线胶质瘤。九类脑胶质瘤图像示例如图2-3所示。

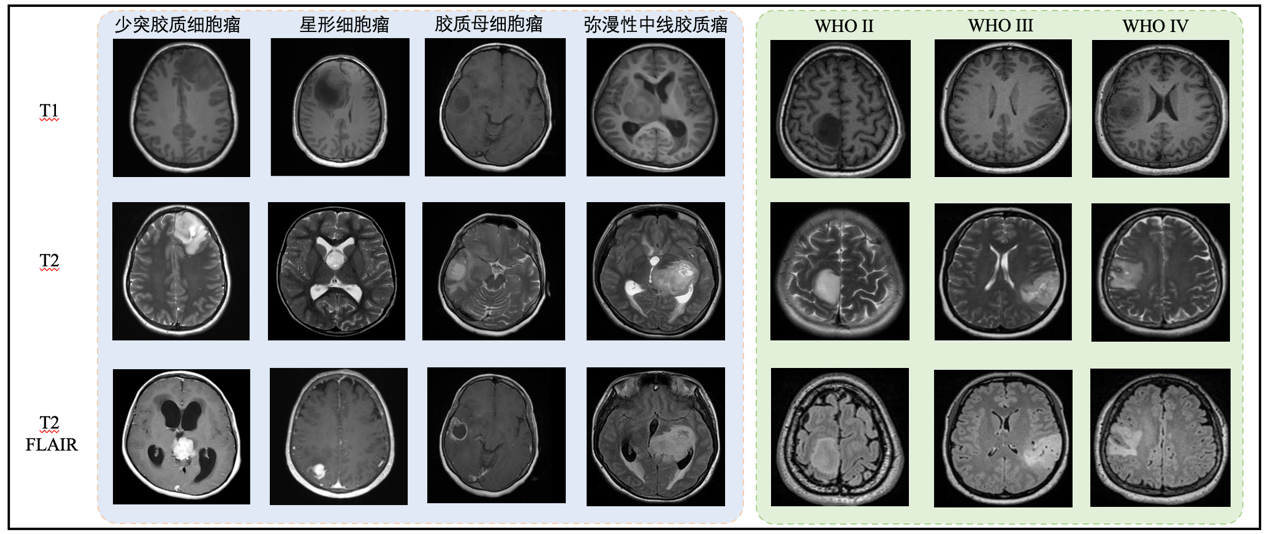


图2-3 九类脑胶质瘤图像

Fig. 2-4 Example of nine Brain Gliomas images

### 2.1.2 脑胶质瘤生物标志物数据

本文采用12项临床生物标志物数值特征，覆盖人口学、分子分型、增殖活性与染色体改变四类信息：性别与年龄作为基线协变量反映总体风险与流行病学差异；IDH与TERT突变刻画弥漫性胶质瘤的关键分层信号，二者的双突变表征特定分子谱系的联合效应；Ki-67指数量化肿瘤细胞增殖活性；MGMT与烷化剂化疗反应及预后关联；1p/19q共缺失为少突胶质细胞瘤的标志性拷贝数改变；ATRX缺失与TP53突变多见于星形细胞系；FUBP1与CIC突变常见于少突系，指示特定遗传背景与生物学行为。上述特征共同构成影像学之外的分子层面表征，用于补充MRI对空间结构与形态学的刻画，在分级与分亚型判别、预后分层与治疗策略选择中具有重要参考价值。

### 2.1.3 脑胶质瘤MRI数据特性

本节将从脑胶质瘤MRI图像特性，以及鉴别肿瘤关键特征两方面展开介绍。

临床脑胶质瘤MRI图像特性主要体现在以下2个方面：

（1）低质量图像占比大。临床采集到的脑胶质瘤MRI数据存在伪影和数量少的情况。在质控环节，观察到部分切片存在扫描运动伪影与化学位移伪影等成像缺陷，如图2-4，可能影响边界判定与纹理表征。受限于合格样本规模，训练过程中出现欠拟合倾向，表现为训练与验证精度均偏低。

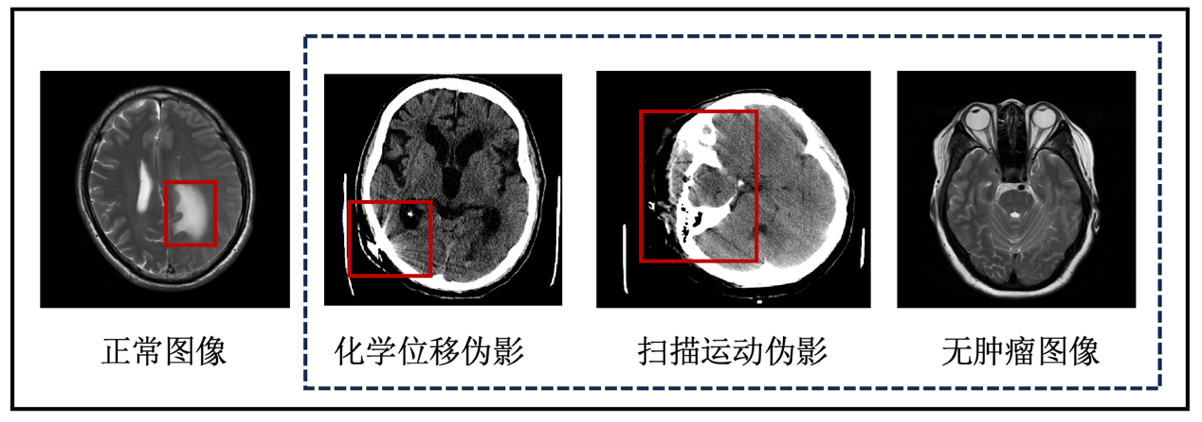


图2-4低质量图像示例

Fig. 2-4 Examples of low-quality images

（2）脑肿瘤与正常脑组织在纹理上高度相似，且影像对比度较低，如图2-6所示。这种2D切片除肿瘤本身外常同时包含皮髓界、脑室旁白质、血管影与其他解剖结构的纹理信息，尤其是胼胝体，呈弧形带状的高髓鞘白质，在T2和T2-FLAIR模态上与邻近水肿带可形成梯度过渡与局部纹理相似，叠加偏置场、部分容积效应及体素各向异性等因素，易引发边界模糊与类间特征重叠，进而导致假阳性或错误聚焦。

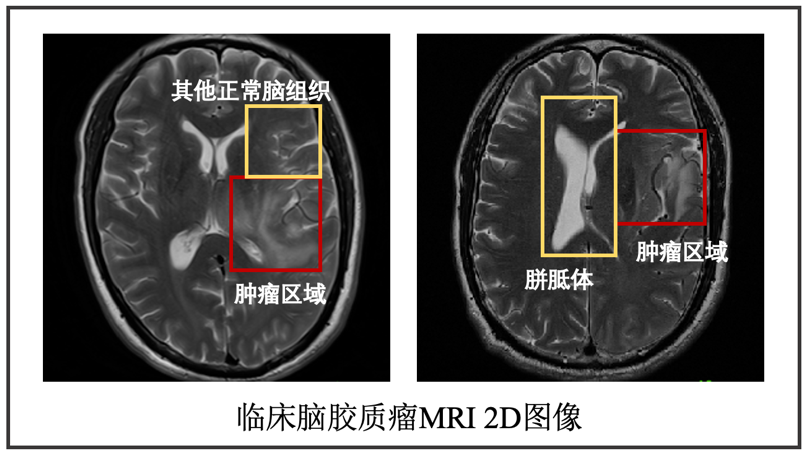


图2-6无关组织干扰

Fig. 2-6 Unrelated tissue interference

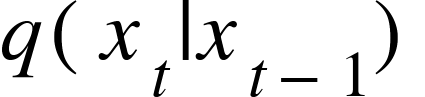
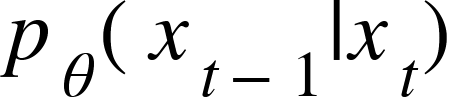
鉴别肿瘤关键特征时，主要关注脑肿瘤区域的信号、轮廓、纹理特征。

（1）信号特征：肿瘤区域常表现为相对于邻近脑组织的信号异常升高或降低，直观上在灰度图像中呈现更亮或更暗的区域。该差异带有显著的序列依赖性：肿瘤相关水肿与浸润带多见高信号，坏死或囊变亦可偏高信号；血脑屏障受损与新生血管导致的对比剂外渗使肿瘤实质或环形边缘出现强化，从而与周围实质形成清晰的信号对比。

（2）轮廓特征：肿瘤常呈不规则/分叶、可见环形强化与坏死中心；长径/短径、圆度、紧致度、Fractal维度等指标区分团块样病灶与规则白质结构；边界梯度强度、曲率、边界“模糊带”宽度、是否浸润越界等。病变常呈不规则状外形，边界参差且存在浸润性外扩，伴随占位效应、邻近沟回受压与中线偏移。可出现典型的环形强化边缘与中心坏死区，提示肿瘤内部的血流动力学与组织学异质性； “边界模糊带”反映浸润与水肿的过渡区域。边缘的连续性、光滑度与曲率变化有助于区分团块样肿瘤与规则的白质条带结构；边界梯度的陡峭与否亦影响模型对分割与分类的稳定性。

（3）纹理特征：肿瘤常呈“斑片—颗粒—粗糙”等多尺度混合纹理，内部可能共存坏死、囊变、出血与高细胞密度区，导致信号在短距离内频繁波动，显示为不均匀与不规则的细微结构。相对地，正常白质在宏观上趋于均匀且结构方向性明确，胼胝体等高髓鞘纤维束表现出更稳定与各向一致的纹理模式。

## 2.2 扩散概率模型图像生成算法

2015年，Jascha Sohl-Dickstein等人受非平衡统计物理启发提出扩散概率模型（Diffusion Probabilistic Model，简称扩散模型）[61]，为此类生成建模奠定了方法学基础。从变分建模视角，扩散模型可视为将VAE（变分自编码器）的潜变量由单步扩展为随时间展开的马尔可夫层级（即 MHVAE）：将潜变量按序组织成逐步加噪的前向过程，并学习与之对应的反向生成。在此框架下，通过多步噪声调度定义从弱到强的扰动序列，配合对目标函数的重写，使整体训练由最大化分层变分下界（ELBO）转化为由加权重构项与KL项组成的可优化目标，进一步等价为去噪分数匹配形式，从而在每一时间步学习稳定的去噪映射或分数场。相较MHVAE，扩散模型在噪声设计、参数共享与目标重构上的处理更具结构化与数值稳定性，便于大规模训练并提升样本保真度。图2-7[62]展示了二者在图模型与优化目标上的对应关系与等价连接。

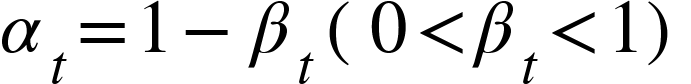
图形用户界面, 应用程序

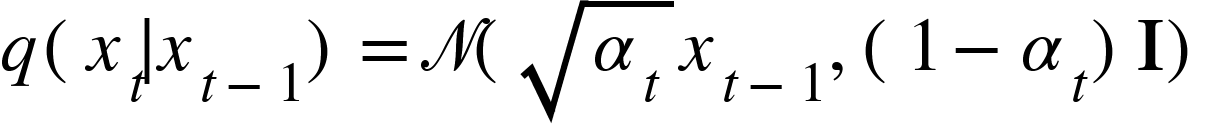
AI 生成的内容可能不正确。

图2-7[62]扩散概率模型图

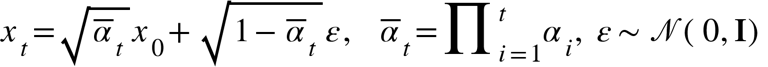
Fig. 2-6 Diffusion probabilistic model image

从整体来看，扩散模型包括前向扩散过程和后向去噪过程，在后向去噪过程中每一次预测的是噪声，通过逐层递减来去噪，进而生成清晰图像。

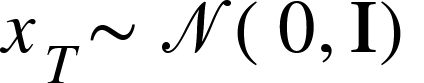
前向扩散过程。从真实样本出发，沿马尔可夫链逐步加噪，使变量趋于各向同性高斯。设,则每一步的编码分布为：

 （2-1）

并可写成闭式：

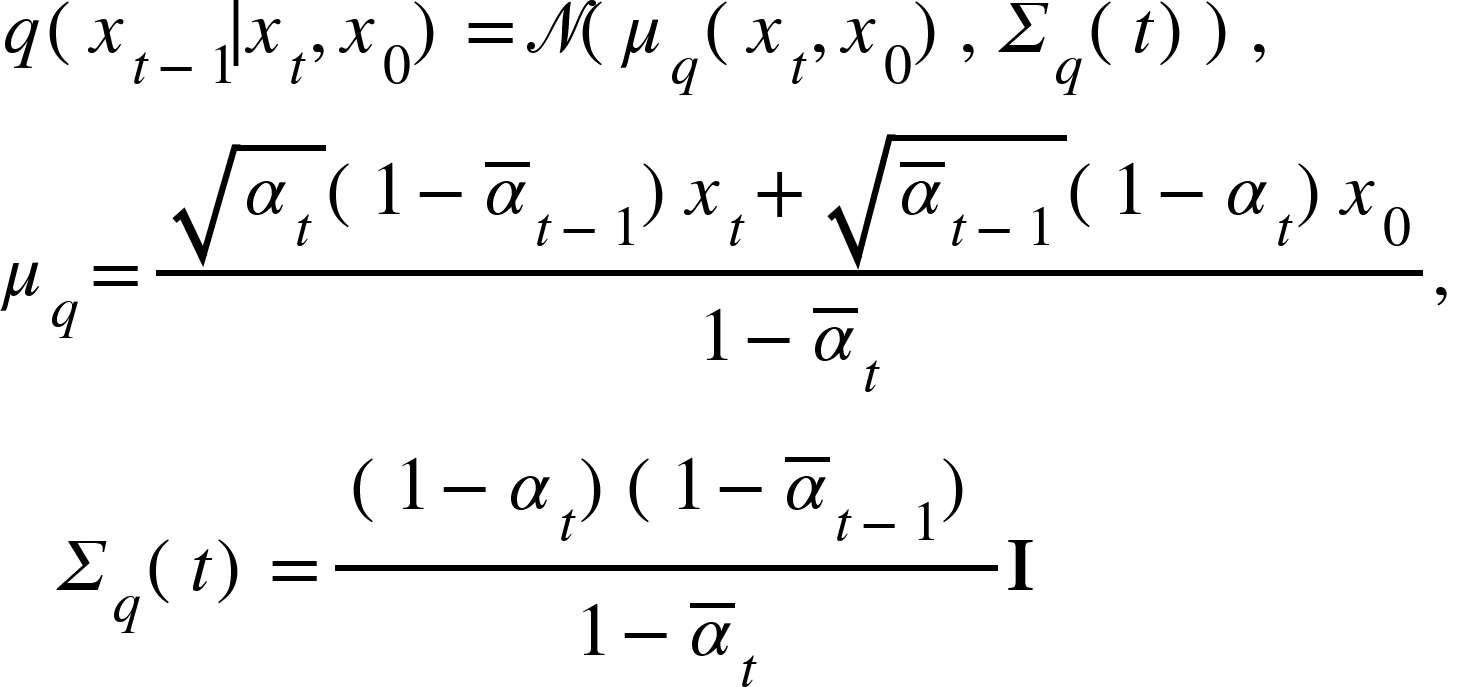
 （2-2）

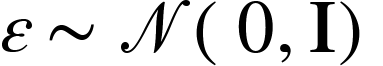
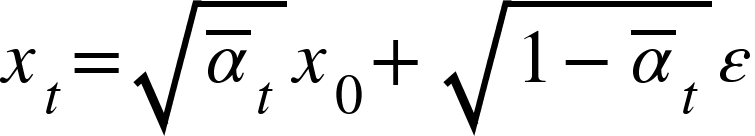
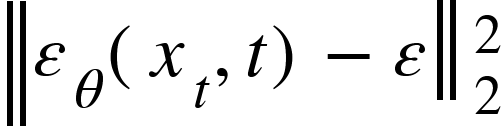
该设置保证当{"mathml":"<math style=\"font-family:stix;font-size:12px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mi>T</mi><mo>&#x2192;</mo><mo>&#x221E;</mo></math>","origin":"MathType Legacy","version":"v3.20.0"}时，近似{"mathml":"<math style=\"font-family:stix;font-size:14px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mi mathvariant=\"script\">N</mi><mo>(</mo><mn>0</mn><mo>,</mo><mi mathvariant=\"bold\">I</mi><mo>)</mo></math>","origin":"MathType Legacy","version":"v3.20.0"}。

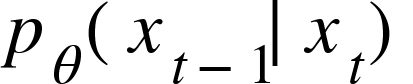
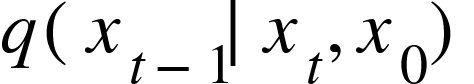
后向去噪过程。生成从逐步去噪：

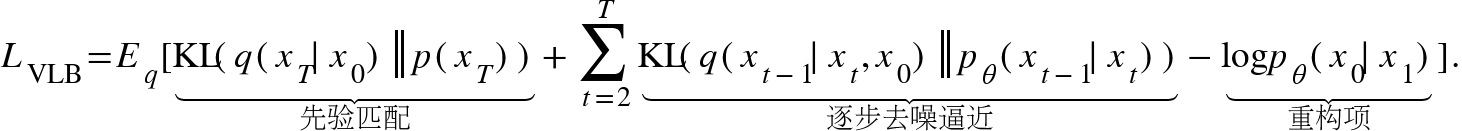
 （2-3）

其真实后验为：

 （2-4）

训练时用“噪声预测”目标：采样，令,最小化，从而使逼近，获得高质量采样。

最大化ELBO。训练的本质是让可学习的反向链逼近真实后验。对应的负对数似然可由变分下界上界化，得到逐步可分解的目标：

（2-5）

## 2.3 经典图像分割算法

### 2.3.1扩展的3D U-Net

3D U-Net[63]将2D U-Net的编码—解码与跳连全面拓展到体素域，以端到端的 3D卷积和反卷积实现从稀疏体素标注学习致密分割，成为医学三维分割的基础范式；其关键在于体数据块训练、弹性形变增广与跨层特征拼接以兼顾细节与上下文。在此基础上，V-Net[64]以体素FCN方式引入Dice目标函数，直接缓解前景或背景极度不均衡问题，并借助数据增强与体素级端到端学习提升3D MRI分割的鲁棒性与效率。随后，DenseVoxNet[65]通过密集连接与辅助分支加强梯度流与特征复用，用更少参数获得更强表示能力与稳定训练，体现“轻量—高效”的三维网络取向。面向保持高分辨率细节与大感受野的矛盾，HighRes3DNet[66]采用空洞（扩张）卷积叠加的残差块在 3D 语境下构建紧凑高分辨率主干，兼顾长程上下文与细粒度边界信息。

工程层面，nnU-Net[67]将三维 U-Net 的“隐性经验”系统化：自动化配置预处理（体素重采样、强度归一化）、网络尺度与深度、训练/后处理与3D-cascade 策略，使3D、2D级联各变体在跨数据集上具备“开箱即用”的强基线表现，并显著降低任务间迁移门槛。综上，扩展的3D U-Net 家族围绕三类核心矛盾持续演进：1）类别不均衡与小目标（Dice 类损失、深监督）；2）计算/显存受限与分辨率保持（块状训练、空洞卷积、级联多尺度）；3）异质成像带来的域移（归一化、重采样、规则化后处理）。它既提供了统一、稳健的体素级分割基线，也为后续注意力、Transformer/混合编码器及自监督预训练等路线提供了即插即用的三维承载框架。

### 2.3.2 联合的2D-3D CNN

联合2D–3D的思想旨在融合“面内细节敏感的 2D 卷积”与“捕获体上下文的 3D 卷积”，在显存、各向异性与临床分辨率受限的约束下取得更稳健的分割。H-DenseUNet[68]作为代表性框架，以2D DenseUNet高效抽取切片内纹理与边界，再以轻量 3D 分支聚合层间上下文，并在融合层端到端优化，从而兼顾 2D 网络忽略体上下文与 3D 网络计算开销大的双重短板；该范式在肝/肿瘤等任务上验证了混合建模的有效性。

从系统视角，2D–3D 联合常与级联协同[69]：先用2D（或2.5D）网络进行粗定位与器官分割，再以3D子网在ROI内细化边界与小体积病灶；或并行训练2D 与3D模型并在特征或决策层融合，以降低单一路线对数据各向异性与对比度波动的敏感性。近年的工作[70]还将混合范式与轻量化、注意力、密集连接、深监督等模块化手段组合，以进一步抑制误检、提升小目标与长程依赖建模能力，同时保持可部署的计算成本。总体而言，2D–3D 联合为临床常见的低对比、各向异性与资源受限场景提供了工程上可落地的权衡方案：2D 侧重精细纹理与清晰边界，3D 负责全局一致性与上下文约束，二者在融合/级联中形成互补，推动了医学影像分割走向更稳健与普适的系统设计。

## 2.4 图像分类相关技术

### 2.4.1 经典分类算法

VGG[71]网络标志着深度学习在图像分类领域取得重大突破。其中VGG16凭借其16层架构（包含13个卷积层和3个全连接层）实现了计算机视觉领域的重大飞跃。这种深度结构结合小尺寸（3x3）卷积滤波器，显著增强了模型提取复杂特征的能力。VGG16的设计使识别复杂抽象的视觉模式成为可能，为图像处理中的深度网络架构树立了新标杆。该模型通过迁移学习在医学影像领域的应用展现出显著成效。

ResNet[72]显著推动了图像分类领域的发展。其关键创新在于残差学习方法，该方法有效解决了深度神经网络中常见的梯度消失问题。ResNet通过整合“跳跃连接”实现这一目标，使部分输入能够绕过一个或多个层，随后与后续层的输出进行合并。该架构特性确保网络能在必要时学习恒等函数，从而维持训练稳定性并提升学习效率。更重要的是，该设计使更深层网络的训练成为可能，显著提升了复杂图像分类任务的性能。在医学影像领域，ResNet50通过迁移学习成功应用于疾病分类研究。

DenseNet[73]在卷积网络架构领域实现了突破性进展，其核心在于采用密集连接的设计。与传统架构不同，DenseNet的每一层都能接收所有前置层的输入，从而实现全面的特征融合。这种结构显著增强了网络中信息与梯度的流动性，不仅提升了训练效率，还减少了信息损耗。DenseNet的突出优势在于卓越的特征复用效率，可在不牺牲性能的前提下减少参数数量。在众多变体中，DenseNet121因在各类图像分类任务中表现强劲而尤为突出，尤其擅长处理具有复杂纹理与图案的图像。其在医学图像分析领域的应用，特别是疾病诊断与肿瘤检测已获得广泛认可与采用。

### 2.4.2 注意力机制

通道注意力机制[74]（Channel Attention, CAM）的核心是把卷积特征的每个通道视作不同的语义子空间，通过学习一组通道权重对其“重标定”，从而放大判别性通道、抑制冗余或噪声通道；典型实现以全局聚合——跨通道映射——逐通道缩放为主线，其中SE块以全局池化提取通道描述子，经轻量MLP生成门控向量并与残差路径结合，显著提升网络表征而代价很小，如图2-8上所示。

空间注意力机制[75]（Spatial Attention, SAM）旨在在空间维度上为特征图的每个位置分配显式权重，使网络在复杂背景中聚焦于与任务相关的区域与边界。其通用流程为：先在通道维进行信息压缩以获得空间提示图，再通过小卷积或相关性建模生成单通道注意力图，并以 Sigmoid 归一化后对原特征做位置逐点加权。与通道注意力强调“选通道”不同，SAM 强调“选位置”，两者串联或并联可形成互补：前者筛选语义子空间，后者突出关键区域与轮廓。在医学影像中，SAM 有助于抑制颅骨、血管等无关结构并强化肿瘤边缘与浸润带；在解码器或跳连处插入可显著提升细粒度分割/分类的判别性，如图2-8下所示。

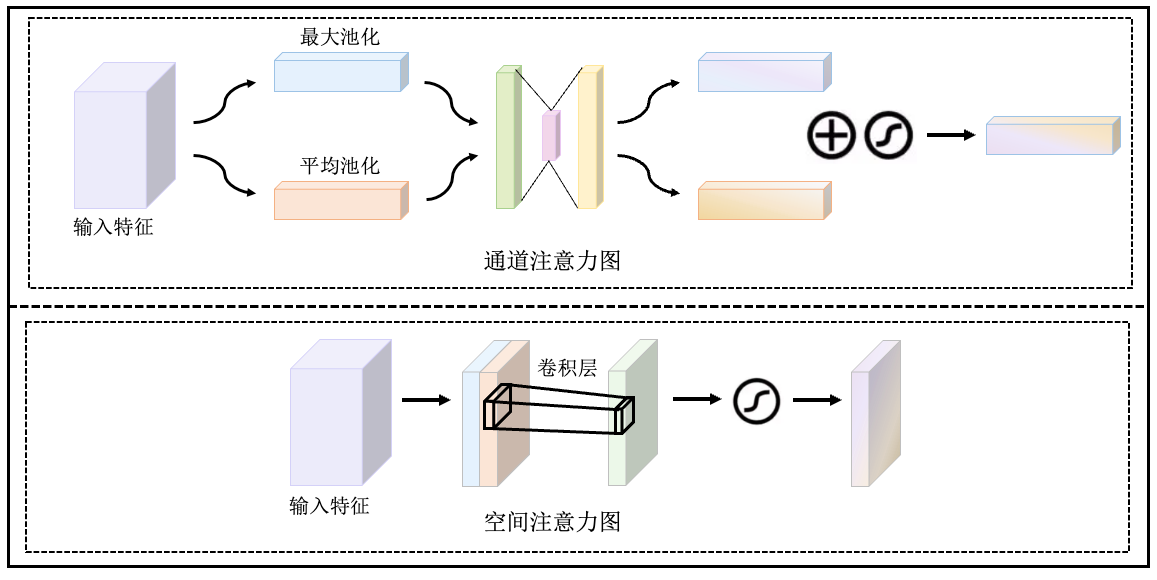


图2-8通道注意力和空间注意力机制模型图

Fig. 2-8 Schematic diagram of the channel attention and spatial attention mechanism model

## 2.5 知识蒸馏技术

知识蒸馏（Knowledge Distillation, KD）是一类用“教师–学生”范式将强模型的判别能力迁移到更小、更快模型的技术，如图2-9所示。经典做法以温度化softmax产生的软标签作为教学信号，学生通过最小化与教师输出分布的差异来学习，从而在显著压缩计算与存储成本的同时尽量保持精度；这一思路最早由Hinton[76]等系统化提出并验证了对大型语音与视觉模型的有效性。KD 常以监督交叉熵 + 蒸馏损失的加权组合训练学生：前者对齐真标签，后者对齐教师的软输出；温度T用于平滑教师分布、显式暴露非目标类的相对次序信息，从而提供更丰富的梯度信号并改善概率校准。除离线KD外，还出现了自蒸馏/在线蒸馏等范式：例如Born-Again Networks[76]反复“以生为师”，在相同结构下多代训练能使学生反超教师，显示KD不仅是压缩手段，也是一种有效的优化策略。

图示

AI 生成的内容可能不正确。

图2-9知识蒸馏模型图

Fig. 2-9 Knowledge distillation model diagram

## 2.6 本章小结

本章梳理了基于生成式数据增强和深度学习的脑肿瘤图像分类研究的相关概念和技术，详细介绍了脑胶质瘤MRI和生物标志物数据、脑胶质瘤MRI数据特性知识，以及详细介绍了图像生成、3D图像分割、图像分类和知识蒸馏相关技术，为本文提供一定的技术支撑。

# 第3章 基于条件去噪扩散概率模型生成脑肿瘤图像研究

基于第2章节中介绍的脑肿瘤MRI相关知识和图像生成技术。本章针对临床数据稀少难以利用的问题，开展了基于条件去噪扩散概率模型生成脑肿瘤图像研究，旨在改善临床数据质量和数量问题，从而为第4章和第5章图像分类提供高质量的数据基础。

## 3.1 引言

脑肿瘤MRI自动分类依赖于高质量数据以支撑稳健的特征学习与泛化；样本的数量、标注精度与跨中心一致性直接决定模型对分类的判别能力。然而，真实临床环境下的数据获取与共享存在显著制约：1）合规与共享受限。隐私与伦理合规要求严格，原始 DICOM 的脱敏与跨机构流转成本高、流程复杂，且成像协议异构，难以在可控合规前提下构建规模化、结构均衡的数据资源库。2）质控不足与有效样本稀释。数据中混入运动伪影、化学位移伪影以及大量无肿瘤非诊断性切片，在未充分质控与剔除的情况下显著降低有效样本占比并引入噪声，削弱模型学习的稳定性与上限。3）域间差异导致性能落差。公开基准与真实临床场景在扫描协议、体素各向异性、人口学与病种构成等方面存在显著域移，致使公开数据上取得的分类性能难以在临床数据上复现，表现为明显的泛化性能衰减。如图2-4所示。上述因素共同构成数据瓶颈，限制了传统深度学习在临床场景的训练效率与泛化能力。

为缓解该问题，本章提出隐私友好的合成数据增强方案：基于条件去噪扩散概率模型学习真实影像的统计与结构先验，在不暴露个体信息的前提下按分类标签生成可控样本；结合跨域一致化与弱监督质控，对合成与临床样本进行筛选，实现类内变异扩展与类别再平衡；最终将“仅临床”“仅合成”与“混合”数据构成的训练集纳入统一流程，支撑小样本建模与跨中心泛化，为后续分类模型提供规模化、可控且合规的数据基础。

本章提出的基于条件去噪扩散概率模型生成脑肿瘤图像研究，主要贡献包括两方面：（1）本章构建了基于条件去噪扩散概率模型的图像生成模块。采用“两阶段”训练范式，即公开数据预训练，少量临床样本迁移微调。以分类标签为条件实现可控合成，为后续研究构建优质数据集。（2）本章构建了“生成—分割—验证”自动化评测流程。搭建从数据生成到数据评估的一体化流水线：按规则构建三套数据集，即仅生成、仅临床、混合，在统一预处理后调用固定权重的 U-Net [77]进行分割推理，并以Dice 为主指标，评估生成图像数据的有效性。

## 3.2 方法描述

基于条件去噪扩散概率模型生成脑肿瘤图像研究包括量个模块：图像生成模块和图像数据有效性评估模块，研究框架图如图3-1所示。首先本章通过图像生成模块生成脑胶质瘤MRI 3D数据。其次，用临床数据集和生成数据集构建三个数据集：仅生成、仅临床、混合数据集，用以传递给图像数据有效性评估模块；最后，将构建好的三个数据集用同样的方法进行数据有效性检验。图像生成和图像数据有效性评估两个模块以串行方式运行，实现“生成—分割—验证”自动化评测流程，以构建后续研究所需数据集。

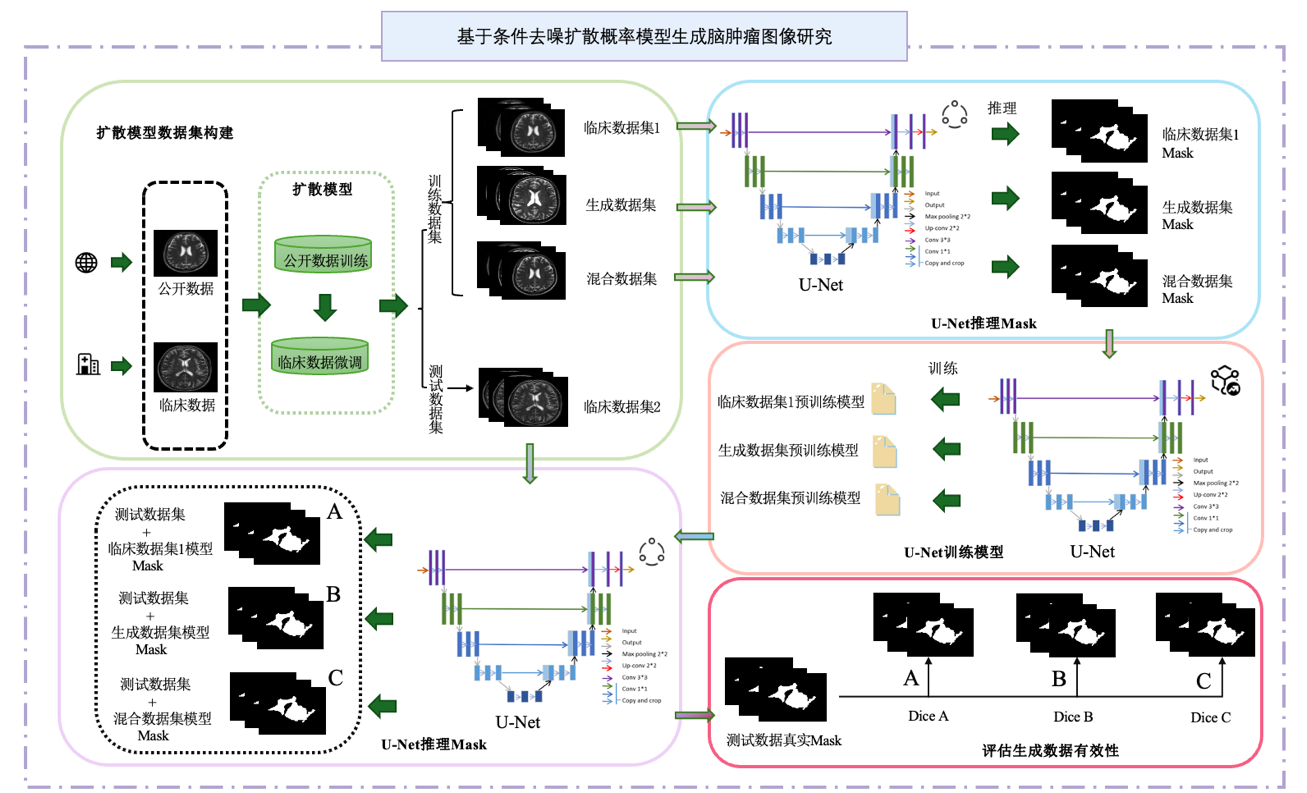


图3-1 脑胶质瘤图像生成研究内容框架图

Fig. 3-1 Framework diagram of glioblastoma image generation research content

### 3.2.1 基于条件去噪扩散概率模型的脑胶质瘤MRI数据生成

为了解决临床数据稀缺、存在图像数据质量不高和无肿瘤数据占比大的情况，本章设计了一个图像生成模块。相较于传统的几何或光照扰动仅在像素层面做“表面变化”，生成模型能从数据分布层面学习脑部解剖、序列对比度与肿瘤形态的联合先验，在此基础上按分类条件“可控地”合成解剖合理、对比度可信的3D数据，从而在不泄露个体信息的前提下扩大样本规模与类内多样性、定向补齐稀有类别与边界样本，缓解长尾不均衡与过拟合。

为了实现生成数据的目标，研究者们广泛探索了多种生成模型，其中最为常见的是GAN和Diffusion。虽然GAN在医学图像生成中表现出了较强的能力，但其训练过程常常不稳定，容易出现模式崩溃问题，即生成的图像缺乏多样性。与GAN相比，Diffusion通过逐步引入噪声并在后续步骤中去噪重构数据，能够以逐步逼近真实数据分布的方式生成高质量的样本，具有更为稳定的训练过程，和较强的表达能力，能够捕捉图像中的细节信息。综上考虑，本研究最终选择了Diffusion作为合成患者数据的生成方法。

具体的操作步骤如下：

（1）条件去噪扩散模型训练。通过在公开的医学影像数据集GLIS-RT[78]和GLIS-RT[79]上进行大规模训练，模型学习到该数据集的统计特征。

（2）微调。考虑到生成的合成数据可能与临床数据之间存在一定的差异，本研究采用微调的策略，在生成模型完成预训练后，通过少量的临床数据对模型进行微调，以进一步提升生成图像的质量和临床适用性。

为了验证生成数据的有效性，本研究构建了三种数据集：1）仅生成数据集：该数据集仅包含通过去噪扩散模型生成的合成数据，用于评估在缺乏真实数据支持的情况下，生成数据是否能够提供足够的有效信息，以支撑模型的训练与推理。2）仅临床数据集：该数据集完全由临床的医学影像数据构成，旨在评估仅依赖公开数据训练的模型在分割任务中的表现，作为基准标准。3）混合数据集：该数据集通过按一定比例混合生成数据与临床数据集，旨在探讨生成数据与真实数据结合后，是否能够改善模型性能并提升分割效果。数据集构建框架流程如图3-2所示。

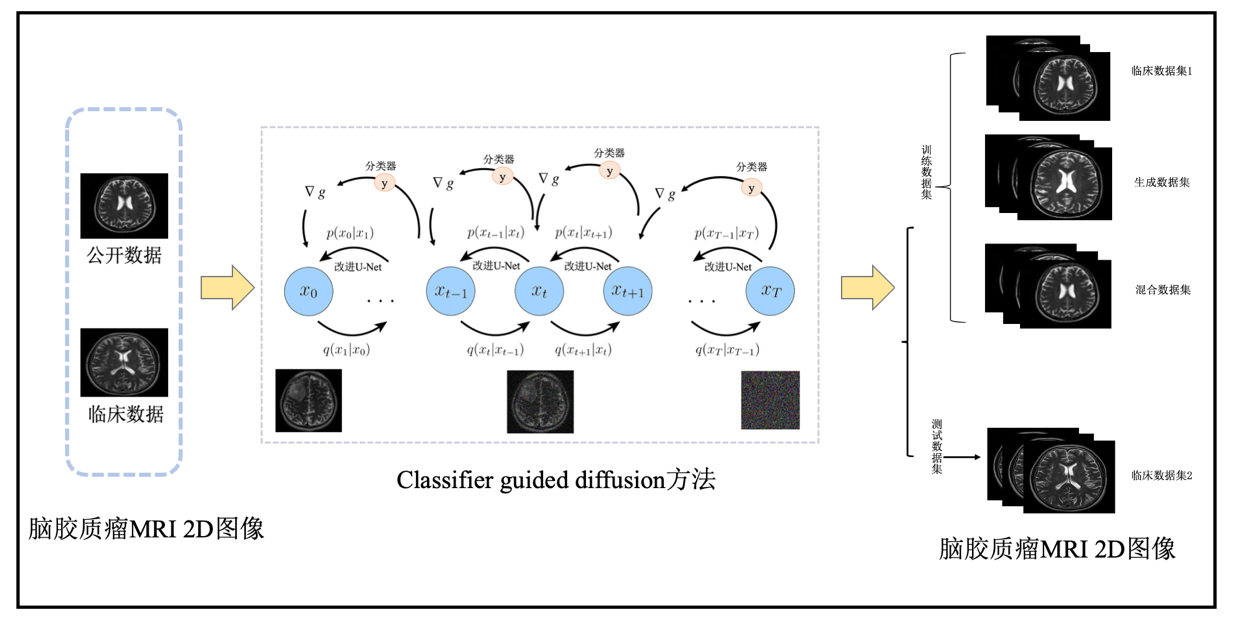
****

图3-2 数据集构建流程图

Fig. 3-2 Data sets construction flowchart

其中，本章采用Classifier guided diffusion方法[80]作为本章图像生成模块的生成模型，接收原始脑胶质瘤MRI 2D图像作为输入，输出为新的脑胶质瘤MRI 2D图像。该模型结构示例如图3-3所示。

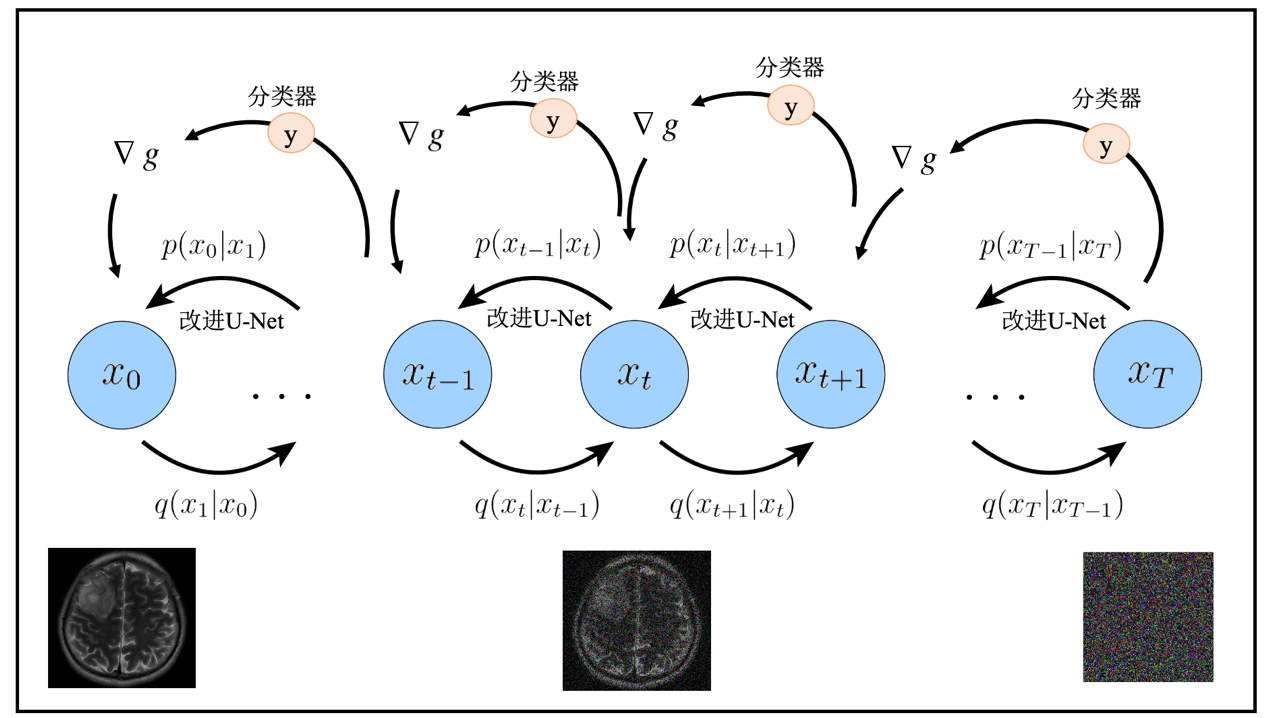
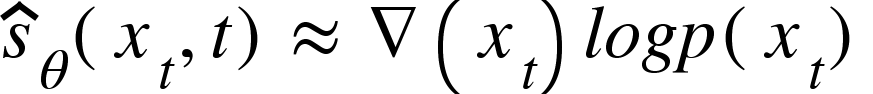
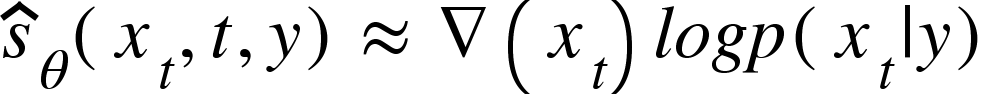
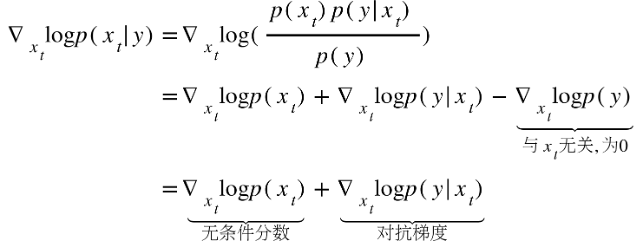


图3-3 Classifier guided diffusion模型架构

Fig. 3-3 Classifier guided diffusion model architecture

该模型为通过类别指导扩散模型生成的代表模型，利用图片类别标签指导反向去噪过程，使生成结果朝向目标类别收敛。为提升条件表达与去噪能力，模型对原始噪声预测U-Net进行了结构性改造，在保持多尺度残差块连接的基础上，引入自适应组归一化，并在每个残差块的归一化操作之后注入，将步数和类别的嵌入向量加入每个残差连接块中。

具体来说，在扩散模型中参数化神经网络预测的是的对抗梯度，无条件下，模型近似，引入类别条件信息之后，目标从边缘分布的梯度转为条件分布的梯度。其中条件分布梯度为：

 （3-1）

展开后的第一项即为无条件分数，第二项为对抗梯度，为分类器，是分类器的输入，是分类器预测的标签，如图3-3所示，先独立训练判别器、再以梯度引导无条件扩散采样。

本章具体训练步骤如下：

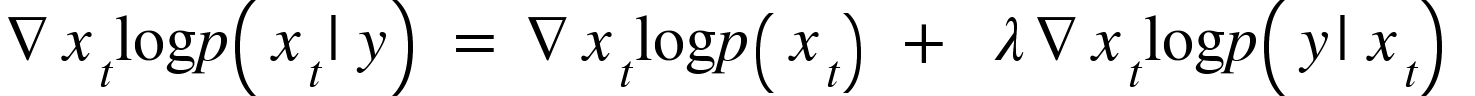
（1）预训练噪声分类器。构建在带噪样本上工作的判别器。骨干采用U-Net的encoder下采样主干，并在8×8空间分辨率处加入注意力池化层以汇聚全局判据。训练数据由前向扩散得到的构造，目标为标准交叉熵。训练完成后冻结分类器参数。

（2）在非条件扩散模型每一个时刻的采样过程中：

1）逐步推断时的无条件预测。对当前带噪样本计算无条件输出;

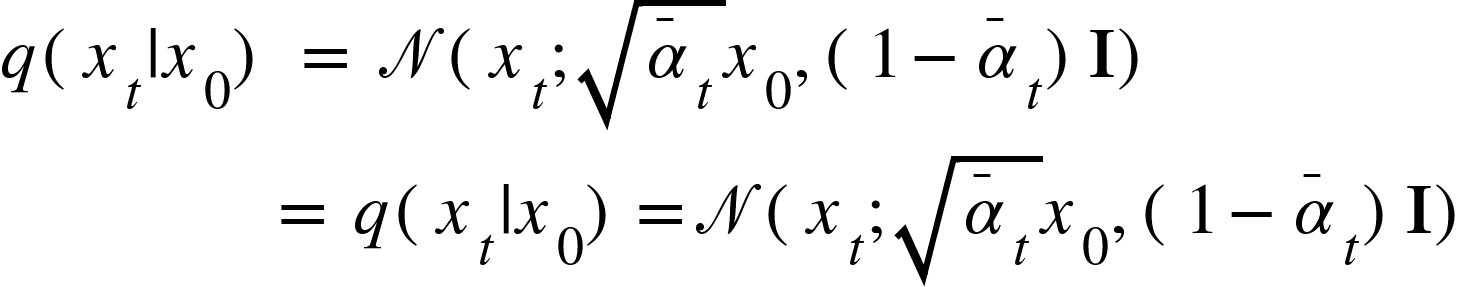
2）分类器引导梯度。将同一输入到已经冻结的噪声分类器，并计算出梯度；

3）条件得分合成与步进。为了灵活控制分类器引导强度，设定超参数权重，将无条件得分与引导梯度线性合成，得到面向类别的条件得分：

 （3-2）

在本章中分别为“WHO II级”、“WHO III级”、“WHO IV级”、“少突胶质细胞瘤”、“星形细胞瘤”、“胶质母细胞瘤”、“弥漫性中线胶质瘤”，为10。

在2.2章节中公式（2-5）提到，可通过优化以下变分下界来实现最大化ELBO，为了使学习到真实数据分布，该方法通过优化变分下界来实现最大化ELBO：

 （3-3）

{"mathml":"<math style=\"font-family:stix;font-size:12px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><msub><mi>L</mi><mrow><mi>v</mi><mi>l</mi><mi>b</mi></mrow></msub><mo>:=</mo><msub><mi>E</mi><mrow><mi>t</mi><mo>~</mo><mo>[</mo><mn>1</mn><mo>,</mo><mi>T</mi><mo>]</mo><mo>,</mo><msub><mi>x</mi><mn>0</mn></msub><mo>~</mo><mi>q</mi><mo>(</mo><msub><mi>x</mi><mn>0</mn></msub><mo>)</mo><mo>,</mo><mi>&#x3F5;</mi><mo>~</mo><mi mathvariant=\"script\">N</mi><mo>(</mo><mn>0</mn><mo>,</mo><mi mathvariant=\"bold\">I</mi><mo>)</mo></mrow></msub><mo>[</mo><mo>|</mo><mo>|</mo><mi>&#x3F5;</mi><mo>&#x2212;</mo><msub><mi>&#x3F5;</mi><mi>&#x3B8;</mi></msub><mo>(</mo><msub><mi>x</mi><mi>t</mi></msub><mo>,</mo><mi>t</mi><mo>)</mo><mo>|</mo><msup><mo>|</mo><mn>2</mn></msup><mo>]</mo></math>","origin":"MathType Legacy","version":"v3.20.0"} （3-4）

其中，，，为前向扩散过程在第步加入的高斯噪声方差的调度参数，决定注入噪声的强度，值越大，加入的噪声越多、信噪比下降越快。

本章首先在公开数据集上训练去条件去噪扩散模型，以学习该数据集的分布并生成类似的合成数据。随后，采用迁移学习策略，将在公开数据集上训练得到的模型迁移到临床数据上进行生成，从而生成具有较高临床相关性的合成数据。

### 3.2.2生成数据有效性评估

为评估生成式数据对下游任务的有效性并避免标注信息泄漏，本章将临床数据按是否具备专家标注的掩膜（Mask）划分为两批：无标注批与有标注批。首先仅基于无标注批训练条件扩散模型以合成与临床分布一致的影像样本；随后以有标注批完全留作测试集，用于独立检验生成数据的有效性与泛化表现。由此构建如下数据方案：

（1）训练集 A（仅无Mask临床数据集）：仅包含无标注临床影像，用于建立基线模型。

（2）训练集 B（仅生成数据集）：仅包含扩散模型生成影像，用于评估纯生成数据对模型可用性的支撑程度。

（3）训练集 C（混合数据集）：由无标注临床影像与生成影像组成的混合集，用于检验生成样本对小样本场景的补益作用。

（4）测试集（有Mask临床数据集）：仅包含有专家掩膜的临床影像，不参与任何训练或调参与采样，用于独立评测与统计检验。

为检验生成数据对分割建模的增益，采用两阶段“伪标注—再训练”流程并在独立测试集上评估。具体为：首先，使用预训练U-Net作为伪标注器，对三套训练数据（训练集A、B、C）分别进行推理，得到对应的伪掩膜。随后，按照伪掩膜上大于30像素的阈值筛选出原图，再将“原始图像 + 伪掩膜”作为监督对U-Net进行再训练，在相同超参数与训练日程下分别得到三组模型：、、。测评阶段，在仅用于测试的有标注临床数据集上对三种模型分别推理分割结果，与专家标注进行对齐比对，并以 Dice系数作为主要评价指标，报告平均值±95%置信区间。框架流程如图3-4所示。

图示

AI 生成的内容可能不正确。

图3-4数据有效性验证框架流程图

Fig. 3-4 Data Validity Assessment Framework flowchart

U-Net是一类用于像素级分割的编码器–解码器网络，核心思想是“下采样提取上下文 + 对称上采样恢复分辨率”，并通过多层级跳跃连接把编码端的高分辨率细节直接注入解码端，从而同时兼顾语义与边界信息，这一结构最早在生物医学2D图像上验证。受益于该网络对特征提取的能力，使用U-Net模型作为推理Mask的模型，结构图如图3-5所示。

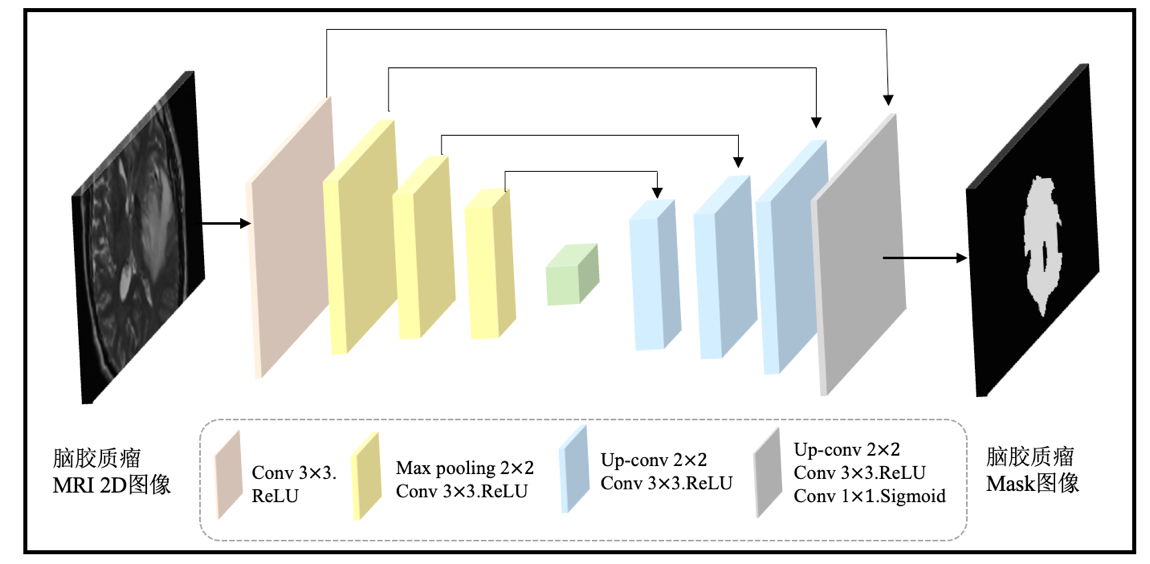
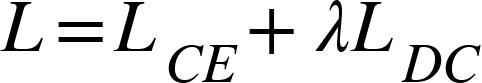


图3-5 脑胶质瘤Mask推理网络

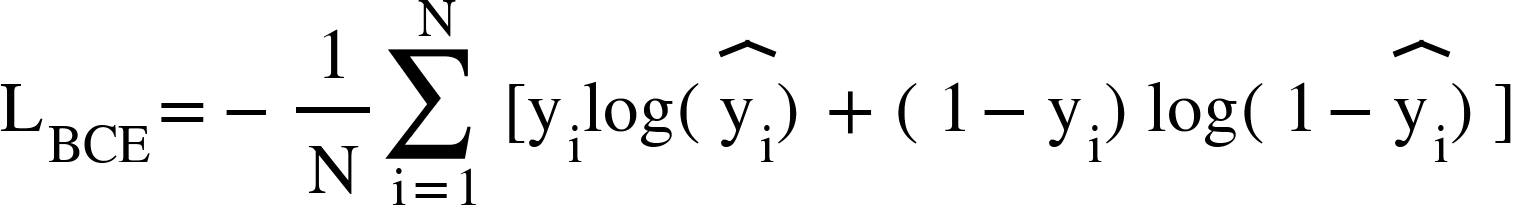
Fig. 3-5 Glioblastoma mask inference network

在本章中，U-Net 输出像素级概率图。单独使用交叉熵虽能稳定地最小化像素级负对数似然，但在“前景稀少、背景占优”的不均衡场景下易偏向背景。Dice损失直接度量预测与真值的区域重叠，可显著提升小体积病灶与模糊边界的召回，故采用混合损失，见式（3-5）。

 （3-5）

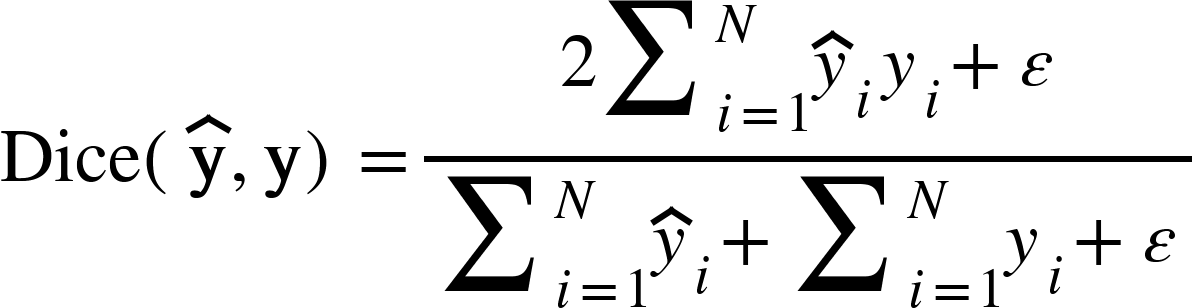
其中控制两者权重；Dice 取可微的soft Dice，并加入平滑以避免分母接近零。分别计算CE与Dice后求和或加权；CE可配合类权重或标签平滑以进一步缓解不均衡并改善概率校准。该混合目标兼顾像素级判别，即CE的稳定梯度与概率学习，与区域级一致性，即Dice的重叠优化，在不改变网络结构的前提下提升整体Dice、边界精度与小病灶检出率。

在脑胶质瘤分割中，主要关注病灶区域和背景信息，属于二分类问题，本章构建BCEWithLogitsLoss函数，为激活函数sigmoid和二值交叉熵损失函数的复合函数，即式中，见式（3-6）。

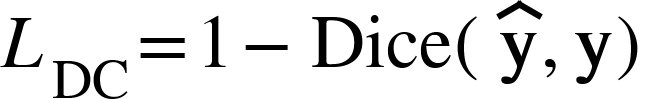
 （3-6）

其中，表示像标签，取值为0或1。表示像素预测值的置信度，取值范围为[0,1]。

DiceLoss基于Dice系数，Dice系数源于衡量医学图像分割领域的两个集合的相似性，见式（3-7）。

 （3-7）

其中，N为脑胶质瘤图像的总像素数， > 0为平滑项，避免分母接近0并稳定梯度，对应的Dice损失为，见式（3-8）。

 （3-8）

## 3.3 实验结果及分析

本节首先介绍用于研究的临床和公开数据集，然后介绍本章使用的评估指标，接着展开实验细节描述，最后进行对比实验结果展示。

### 3.3.1 数据集

本章使用的数据集总体分为两种：（1）公开数据集GLIS-RT和BraTS2017，用于3.2.1的基于条件去噪扩散概率模型的脑胶质瘤MRI数据生成训练；（2）无标注临床数据集，用于3.2.1的基于条件去噪扩散概率模型的脑胶质瘤MRI数据生成，以及3.2.2的数据有效性评估的数据集构建；（3）有标注临床数据集，用于3.2.2的数据有效性评估的测试。

公开数据集GLIS-RT。该数据集包含230例在麻省总医院接受手术和辅助放疗的胶质母细胞瘤和低级别胶质瘤患者。包含198例胶质母细胞瘤、23例间变性星形细胞瘤、5例星形细胞瘤、2例间变性少突胶质细胞瘤和2例少突胶质细胞瘤。

公开数据集BraTS2017。该数据集包含210例高分级胶质瘤患者数据和75 例低分级胶质瘤患者数据。

无标注临床数据集。该数据集包含152例在北京天坛医院接受手术的高级别胶质瘤和低级别胶质瘤患者。包含48例II级胶质瘤患者、44例III级胶质瘤患者和60例IV级胶质瘤患者。

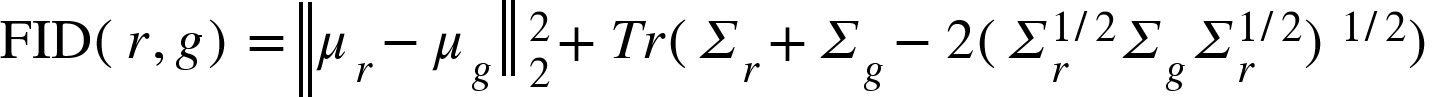
有标注临床数据集。该数据集包含102例在北京天坛医院接受手术的高级别胶质瘤和低级别胶质瘤患者。包含38例II级胶质瘤患者、34例III级胶质瘤患者和30例IV级胶质瘤患者。以及一位拥有8年经验的放射科医师为每张图像提供了体素级别的肿瘤标注。

由于实验是在二维切片层面进行的，需要筛选出包含病灶（病灶>30像素）的切片。为防止数据泄露，按患者级别将数据集划分为训练集和测试集，在训练数据集中，进一步采用8:2的比例划分为训练集和验证集。

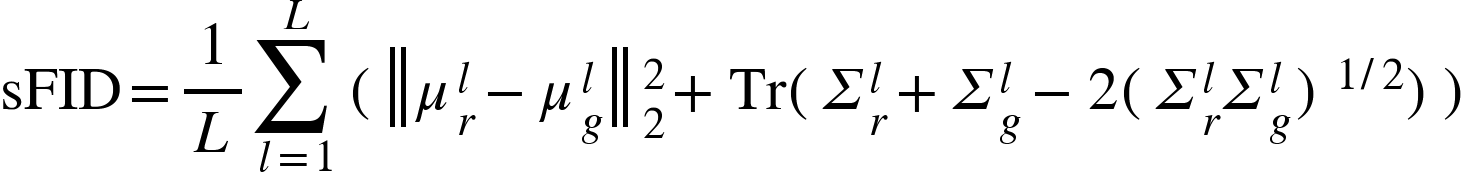
### 3.3.2 评价指标

对于生成数据部分，本章采用FID（Fréchet Inception Distance）[81]、sFID（sliced Fréchet Inception Distance）[82]、R-precision（R精确度）[83]指标进行评估，计算公式见（3-9）-（3-11）。

FID是衡量生成样本与真实样本在特征空间距离的指标，值越小越好。在Inception-V3 特征空间上进行计算，令真实与生成特征分别近似为高斯和，计算公式为：

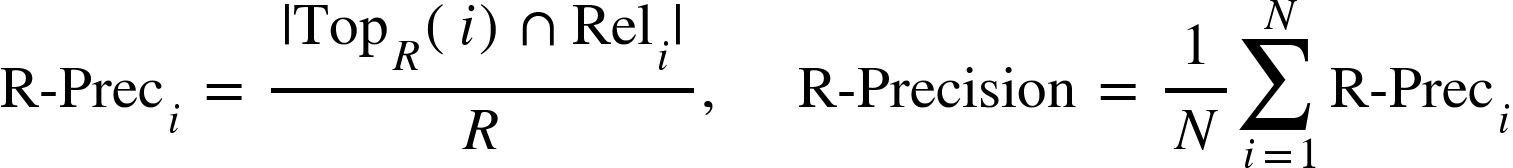
 （3-9）

sFID是对FID的投影版评估指标，值越小越好。随机采样个投影，在每个切片上计算FID并取平均值，计算公式为：

 （3-10）

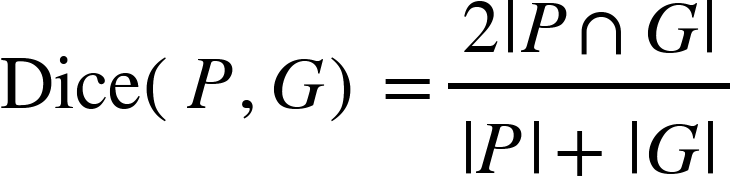
其中，，。

R-precision是评估文本生成图像中，生成图像与文本描述相符度的指标，取值区间为[0,1]，值越大越好。给定第张图像的候选文本集合，其中真实相关文本集合为。用同一嵌入器计算相似度并排序，取前个为，计算公式为：

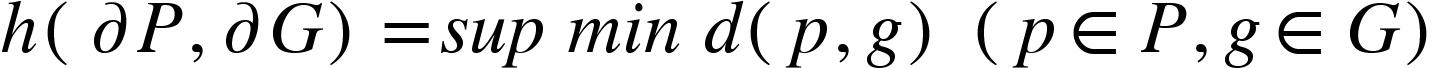
 （3-11）

对数据有效性评估部分，本章采用Dice系数（Sørensen–Dice coefficient）和HD95（95% Hausdorff距离）进行评估，并且输出95%置信区间，计算公式见（3-12）-（3-15）。

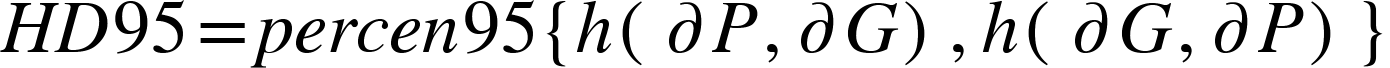
Dice系数是衡量两个集合重叠程度的指标，取值区间为[0,1]，值越大越好。对预测掩膜和真实标记计算：

 （3-12）

HD95是衡量两个轮廓之间的最坏边界误差的指标，值越小越好。设，分别为预测掩膜和真实标记的边界点集，欧式距离为，有向Hausdorff距离为:

 （3-13）

双向Hausdorff距离为，HD95使用双向边界点到对方边界的集合的第95百分位数代替，单位为mm，反映边界对齐与最大可容忍误差，与Dice形成互补，当Dice相近时，HD95更能区分边界是否精确，计算公式为：

 （3-13）

### 3.3.3 实验设置

本章涵盖两部分：2D脑胶质瘤MRI合成数据生成与生成数据有效性评估。训练阶段，首先基于条件去噪扩散概率模型在公开数据上预训练并在无标注临床数据上域自适应，用以生成与临床分布一致的2D切片；随后构建三套训练集（仅临床、仅生成、临床+生成），分别训练对应的U-Net分割模型。测试阶段，采用“公开集预训到临床域生成”的流程获得临床域合成切片；再以预训练U-Net对三套训练数据推理得到伪掩膜，并按病灶面积大于等于30像素的阈值筛选有效切片以统一数据质量与任务难度。最终，在有标注的临床留出集上，使用三个已训练模型分别进行推理评估，从而量化“仅临床”“仅生成”与“混合训练”对分割性能的影响。下文将给出具体的实验设置与实现细节。

基于条件去噪扩散概率模型的脑胶质瘤MRI数据生成实验设置。本章使用Classifier guided diffusion模型在ImageNet数据集上的预训练参数为基础，在GLIS-RT和BraTS2017数据集上进行微调训练，并使用无标注临床数据集进行下采样。

训练核心参数设置如下：使用Adam优化器[84]，其中，；输入切片尺寸为；类别数为7；时序步数为1000；噪声日程采用cosine；基础通道数160，attention\_resolutions为，attention heads为16，每分辨率2个残差块，BigGAN [85]式上下采样，归一化采用组归一化；学习率为2e-4，eps=1e-8，权重衰减为0；混合精度为FP16；EMA衰减为0.9999；训练输入批大小为256；引导强度为1.0。

下采样核心参数设置如下：采样步数为250；噪声日程采用linear；基础通道160，attention\_resolutions为，attention heads为16，每分辨率2个残差块，BigGAN式上下采样；混合精度为FP16。

生成数据有效性评估实验设置。本章使用U-Net作为分割网络，先推理出三个数据集的伪掩膜，然后用原图像和伪掩膜训练U-Net，再用有标注临床数据集在训练好的三个模型上推理。

伪掩膜推理核心参数设置如下：编码、解码深度4，基础通道64，Conv（3×3）–BN–ReLU×2，双线性上采样，3×3卷积；阈值化取0.5，执行形态学闭运算（kernel=3）与连通域过滤；批大小32；开启AMP混合精度。

基于伪掩膜的U-Net训练核心参数设置如下：以结构一致的U-Net进行端到端学习；式（3-5）混合损失的为1.5；优化器采用Adam，其中，；eps=1e-8，权重衰减为=1e-5；学习率为1e-4；学习率策略为cosine；epoch为300；验证集Dice触发早停，patience为15；批大小32；梯度裁剪1.0；AMP开启。

测试推理与评估推理核心参数设置如下：以结构一致的U-Net进行推理，推理流程与伪掩膜阶段一致，全流程在受试者层面去重、预处理一致，确保客观与可复现。

实验环境：本章模型基于Pytorch框架，在 Ubuntu系统上使用NVIDIA RTX 4090Ti显卡实现。

### 3.3.4 对比实验

本节对比实验包括数据生成部分、生成数据有效性评估的性能验证。

对于数据生成部分，本章选择经典生成方法BigGAN以及是否修改框架中的U-Net做对比实验。通过在对比实验和消融实验上测试，以确定使用的框架结构。首先，针对基准模型对比实验，表3-1中展示了BigGAN、StyleGAN2[86]和Classifier guided diffusion的性能。主要依据FID和R-precision分数来选择基准模型，即确定Classifier guided diffusion为基准模型。

表3-1 生成基准模型性能对比展示

Tab. 3-1 Performance comparison of generated benchmark models

|  |  |  |  |
| --- | --- | --- | --- |
| **模型** | **FID** | **sFID** | **R-precision** |
| BigGAN | 8.26 | 9.18 | 0.79 |
| StyleGAN2 | 9.03 | 8.69 | 0.71 |
| Classifier guided diffusion | **6.84** | **5.57** | **0.83** |

其次，针对Classifier guided diffusion中的U-Net改进进行了消融实验，表3-2中展示了包括减少通道数、增加attention heads数、使用attention和BigGAN的残差模块的性能，用FID指标来评估。

表3-2 Classifier guided diffusion的U-Net消融实验性能展示

Tab. 3-2 Performance demonstration of U-Net ablation experiments in Classifier guided diffusion

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **通道数** | **attention heads** | **attention\_resolutions** | **BigGAN up/downsamples** | **FID** |
| 128 | 1 | 16 |  | 15.33 |
| 128 | 4 | 32,16,8 |  | 12.89 |
| 128 | 4 | 32.16,8 |  | 8.72 |
| 160 | 1 | 16 |  | 14.51 |
| 160 | 4 | 32,16,8 |  | 10.25 |
| 160 | 4 | 32,16,8 |  | **6.84** |

在基于条件去噪扩散概率模型的脑胶质瘤MRI数据生成模块中，本文对 BigGAN与Classifier guided diffusion的生成结果进行了并列展示，覆盖 II级、III级、IV级、少突胶质细胞瘤、星形细胞瘤、胶质母细胞瘤、弥漫性中线胶质瘤等七类病理类型的T2模态合成图像，见图3-6。从可视化对比看，Classifier guided diffusion在边界锐度、组织纹理还原与结构连续性方面表现更优，整体细节保真度与视觉清晰度均高于BigGAN，且颗粒噪声相对更少，体现出更好的解剖一致性与病灶形态可辨性。

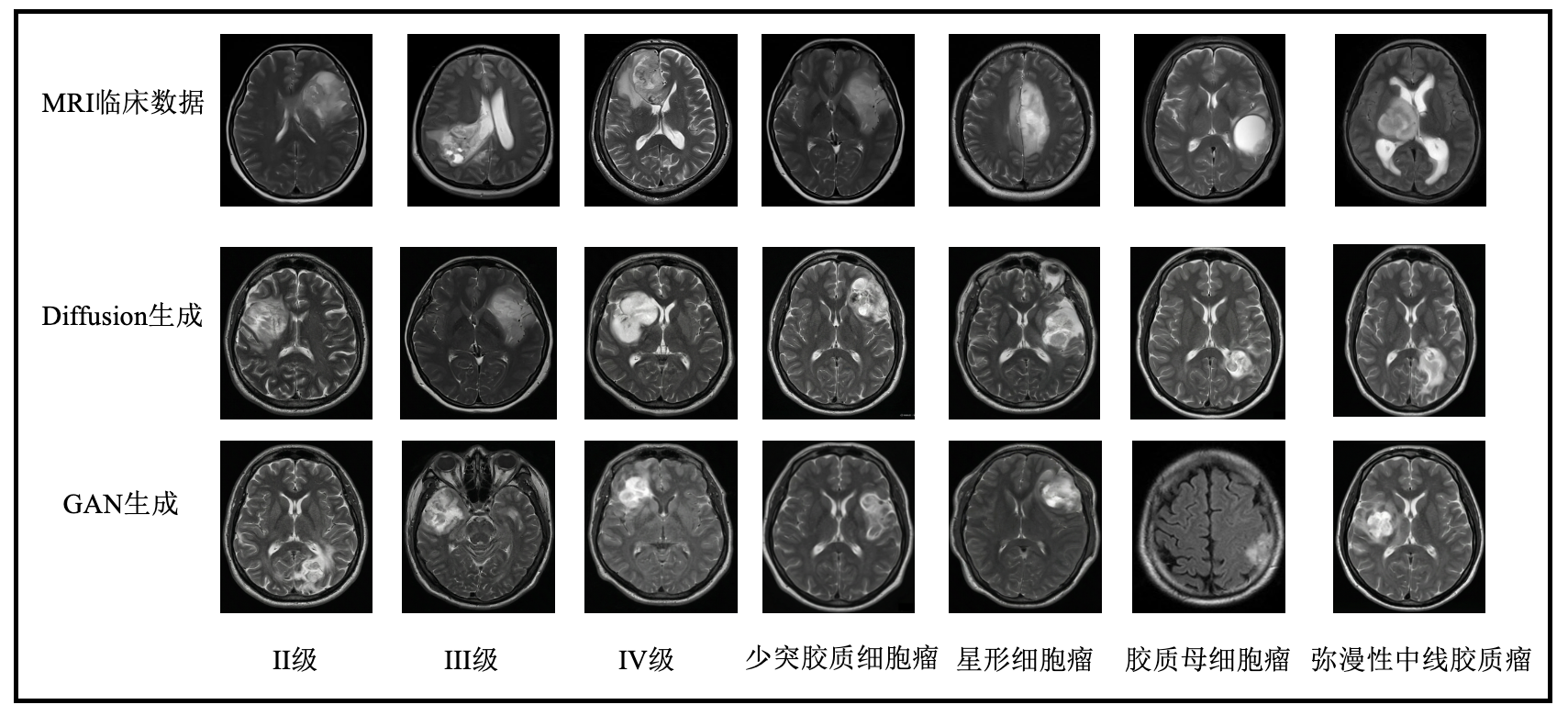


图3-6生成图像展示

Fig. 3-6 Generated images display

鉴于图3-6为大尺度原图，缩略显示对细节解析不足，难以直观比较两类方法的生成差异。为增强可辨性，本文选取IV级T2模态的生成图像代表性样本，对Classifier guided diffusion与BigGAN的生成结果进行局部放大对比，如图3-7所示。放大视图显示：Classifier-guided Diffusion 在边界与轮廓方面呈现更高的连续性与锐利度，边缘锯齿与过度平滑现象较少；肿瘤形态学特征方面能更稳定地保持团块形态与内部异质结构，异常空洞或形变失真减轻；脑组织纹理方面对皮层沟回与白质纹理的细节保留更充分，背景伪影与颗粒噪声较低。综合上述细节指标的视觉比较，Classifier guided diffusion在边界轮廓、肿瘤形状与脑组织纹理的保真度上整体优于BigGAN。

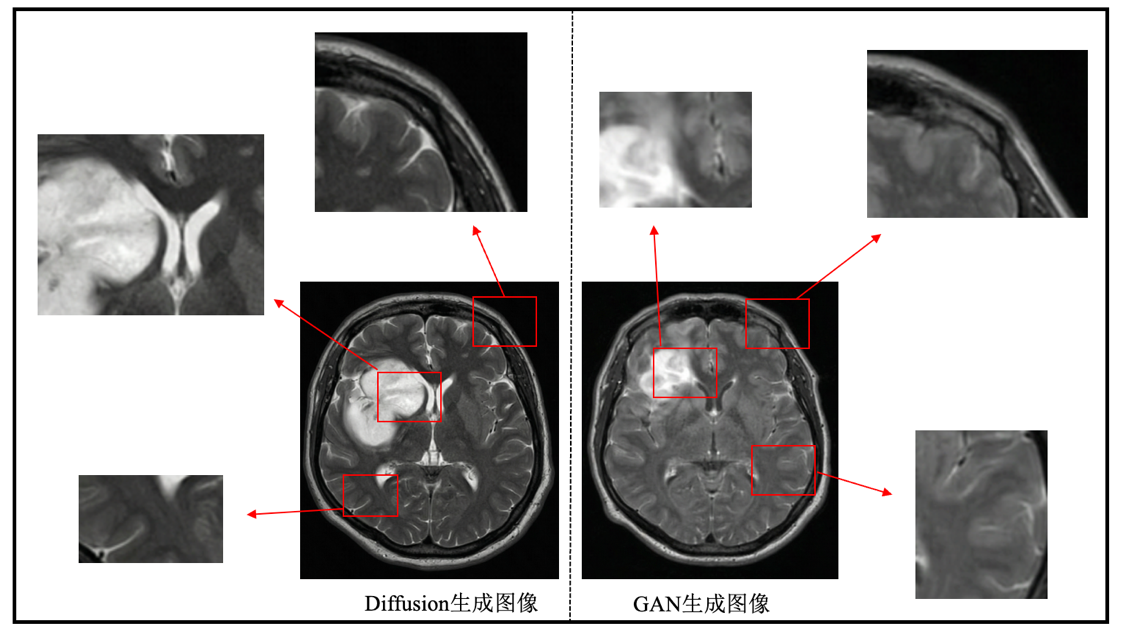


图3-7生成图像细节展示

Fig. 3-7 Generated images detail display

# 参 考 文 献

1. Ostrom Q T, Bauchet L, Davis F G, et al. The epidemiology of glioma in adults: a “state of the science” review[J]. Neuro-oncology, 2014, 16(7): 896-913.
2. Quinn T Ostrom, David J Cote, Mustafa Ascha, Carol Kruchko, and Jill S Barnholtz-Sloan. Adult glioma incidence and survival by race or ethnicity in the united states from 2000 to 2014. JAMA oncology, 4(9):1254–1262, 2018.
3. Weller M, Wen P Y, Chang S M, et al. Glioma[J]. Nature Reviews Disease Primers, 2024, 10(1): 33.
4. Seyhan A A. Circulating liquid biopsy biomarkers in glioblastoma: Advances and challenges[J]. International Journal of Molecular Sciences, 2024, 25(14): 7974.
5. Wang X, Zhao J, Marostica E, et al. A pathology foundation model for cancer diagnosis and prognosis prediction[J]. Nature, 2024, 634(8035): 970-978.
6. Weller M, Wen P Y, Chang S M, et al. Glioma[J]. Nature Reviews Disease Primers, 2024, 10(1): 33.
7. Takashi Komori. Grading of adult diffuse gliomas according to the 2021 who classification of tumors of the central nervous system. Laboratory Investigation, 102(2):126–133, 2022.
8. Hayashi N, Fukai J, Nakatogawa H, et al. Neuroradiological, genetic and clinical characteristics of histone H3 K27-mutant diffuse midline gliomas in the Kansai Molecular Diagnosis Network for CNS Tumors (Kansai Network): multicenter retrospective cohort[J]. Acta Neuropathologica Communications, 2024, 12(1): 120.
9. Ledford A, Rodriguez A, Lipinski L, et al. Functional prediction of response to therapy prior to therapeutic intervention is associated with improved survival in patients with high-grade glioma[J]. Scientific Reports, 2024, 14(1): 19474.
10. Huang Y R, Fan H Q, Kuang Y Y, et al. The Relationship Between the Molecular Phenotypes of Brain Gliomas and the Imaging Features and Sensitivity of Radiotherapy and Chemotherapy[J]. Clinical Oncology, 2024, 36(9): 541-551.
11. Taghipour Zahir S, Dehghani F. Evaluation of diagnostic value of CT scan and MRI in brain tumors and comparison with biopsy[J]. Iranian Journal of Pediatric Hematology and Oncology, 2011, 1(4): 121-125.
12. Bernstein M, Al-Anazi A R, Kucharczyk W, et al. Brain tumor surgery with the Toronto open magnetic resonance imaging system: preliminary results for 36 patients and analysis of advantages, disadvantages, and future prospects[J]. Neurosurgery, 2000, 46(4): 900-909.
13. Karras, T., Aila, T., Laine, S. & Lehtinen, J. Progressive growing of GANs for improved quality, stability, and variation. In Proceedings of the 6th International Conference on Learning Representation (ICLR, 2018).
14. Karras, T. et al. Analyzing and improving the image quality of StyleGAN. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 8107–8116 (IEEE, 2020).
15. Brock, A., Donahue, J. & Simonyan, K. Large scale GAN training for high fidelity natural image synthesis. In Proceedings of the 7th International Conference on Learning Representation (ICLR, 2019).
16. Frid-Adar M, Diamant I, Klang E, et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification[J]. Neurocomputing, 2018, 321: 321-331.
17. Bermudez C, Plassard A J, Davis L T, et al. Learning implicit brain MRI manifolds with deep learning[C]//Medical Imaging 2018: Image Processing. SPIE, 2018, 10574: 408-414.
18. Calimeri F, Marzullo A, Stamile C, et al. Biomedical data augmentation using generative adversarial neural networks[C]//International conference on artificial neural networks. Cham: Springer International Publishing, 2017: 626-634.
19. Bowles C, Chen L, Guerrero R, et al. Gan augmentation: Augmenting training data using generative adversarial networks[J]. arXiv preprint arXiv:1810.10863, 2018.
20. Khader F, Müller-Franzes G, Tayebi Arasteh S, et al. Denoising diffusion probabilistic models for 3D medical image generation[J]. Scientific Reports, 2023, 13(1): 7303.
21. Lyu Q, Wang G. Conversion between CT and MRI images using diffusion and score-matching models[J]. arXiv preprint arXiv:2209.12104, 2022.
22. Saharia C, Ho J, Chan W, et al. Image super-resolution via iterative refinement[J]. IEEE transactions on pattern analysis and machine intelligence, 2022, 45(4): 4713-4726.
23. Song Y, Sohl-Dickstein J, Kingma D P, et al. Score-based generative modeling through stochastic differential equations[J]. arXiv preprint arXiv:2011.13456, 2020.
24. Khader F, Müller-Franzes G, Tayebi Arasteh S, et al. Denoising diffusion probabilistic models for 3D medical image generation[J]. Scientific Reports, 2023, 13(1): 7303.
25. Peng W, Adeli E, Bosschieter T, et al. Generating realistic brain mris via a conditional diffusion probabilistic model[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 14-24.
26. Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 10684-10695.
27. Jiang L, Mao Y, Wang X, et al. Cola-diff: Conditional latent diffusion model for multi-modal mri synthesis[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 398-408.
28. Lyu Q, Wang G. Conversion between CT and MRI images using diffusion and score-matching models[J]. arXiv preprint arXiv:2209.12104, 2022.
29. Litjens G, Kooi T, Bejnordi B E, et al. A survey on deep learning in medical image analysis[J]. Medical image analysis, 2017, 42: 60-88.
30. Xu X, Zhang L, Li J, et al. A hybrid global-local representation CNN model for automatic cataract grading[J]. IEEE journal of biomedical and health informatics, 2019, 24(2): 556-567.
31. Jianqiang Li, Guanghui Fu, Yueda Chen, Pengzhi Li, Bo Liu, Yan Pei, and Hui Feng. A multi-label classification model for full slice brain computerised tomography image. BMC bioinformatics, 21:1–18, 2020.
32. Guanghui Fu, Jianqiang Li, Ruiqian Wang, Yue Ma, and Yueda Chen. Attention-based full slice brain ct image diagnosis with explanations. Neurocomputing, 452:263–274, 2021.
33. Ruiqian Wang, Guanghui Fu, Jianqiang Li, and Yan Pei. Diagnosis after zooming in: A multilabel classification model by imitating doctor reading habits to diagnose brain diseases. Medical physics, 49(11):7054–7070, 2022.
34. Bjoern Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Niklaus Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats), 2018. URL https://www.med.upenn.edu/cbica/brats2018/data.html. Accessed: 2024-09-17.
35. Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Matthew Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Segmentation labels for the pre-operative scans of the tcga-gbm collection [data set], 2017. Accessed: 2024-09-17.
36. Ying Zhuge, Holly Ning, Peter Mathen, Jason Y Cheng, Andra V Krauze, Kevin Camphausen, and Robert W Miller. Automated glioma grading on conventional mri images using deep convolutional neural networks. Medical physics, 47(7):3044–3053, 2020.
37. Subhashis Banerjee, Sushmita Mitra, Francesco Masulli, and Stefano Rovetta. Glioma classification using deep radiomics. SN Computer Science, 1(4):209, 2020.
38. Chenjie Ge, Irene Yu-Hua Gu, Asgeir Store Jakola, and Jie Yang. Deep semi-supervised learning for brain tumor classification. BMC Medical Imaging, 20:1–11, 2020.
39. Mohamed A Naser and M Jamal Deen. Brain tumor segmentation and grading of lower-grade glioma using deep learning in mri images. Computers in biology and medicine, 121:103758, 2020.
40. Yutaka Matsui, Takashi Maruyama, Masayuki Nitta, Taiichi Saito, Shunsuke Tsuzuki, Manabu Tamura, Kaori Kusuda, Yasukazu Fukuya, Hidetsugu Asano, Takakazu Kawamata, et al. Prediction of lower-grade glioma molecular subtypes using deep learning. Journal of neuro-oncology, 146:321–327, 2020.
41. Priyanka Tupe-Waghmare, Piyush Malpure, Ketan Kotecha, Manish Beniwal, Vani Santosh, Jitender Saini, and Madhura Ingalhalikar.
42. Diaohan Xiong, Xinying Ren, Weiting Huang, Rui Wang, Laiyang Ma, Tiejun Gan, Kai Ai, Tao Wen, Yujing Li, Pengfei Wang, et al.
43. Oktay Fasihi Shirehjini, Farshid Babapour Mofrad, Mohammadreza Shahmohammadi, and Fatemeh Karami. Grading of gliomas using transfer learning on mri images. Magnetic Resonance Materials in Physics, Biology and Medicine, 36(1):43–53, 2023.
44. Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In Proc. ICCV 2023, pages 4015–4026, 2023.
45. Yuhao Huang, Xin Yang, Lian Liu, Han Zhou, Ao Chang, Xinrui Zhou, Rusi Chen, Junxuan Yu, Jiongquan Chen, Chaoyu Chen, et al. Segmentanything model for medical images? Medical Image Analysis, page 103061, 2023.
46. Maciej A Mazurowski, Haoyu Dong, Hanxue Gu, Jichen Yang, Nicholas Konz, and Yixin Zhang. Segment anything model for medical image analysis: an experimental study. Medical Image Analysis, 89:102918, 2023.
47. Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In Proceedings of the IEEE/CVF international conference on computer vision, pages 9650–9660, 2021.
48. Mohammed Baharoon, Waseem Qureshi, Jiahong Ouyang, Yanwu Xu, Kilian Phol, Abdulrhman Aljouie, and Wei Peng. Towards general purpose vision foundation models for medical image analysis: An experimental study of dinov2 on radiology benchmarks. arXiv preprint arXiv:2312.02366, 2023.
49. Jakub Prokop, Javier Montalt Tordera, Joanna Jaworek-Korjakowska, and Sadegh Mohammadi. Deep metric learning for few-shot x-ray image classification. medRxiv, pages 2023–08, 2023.
50. Joana Palés Huix, Adithya Raju Ganeshan, Johan Fredin Haslum, Magnus Söderberg, Christos Matsoukas, and Kevin Smith. Are natural domain foundation models useful for medical image classification? In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 7634–7643, 2024.
51. Lerousseau M, Deutsch E, Paragios N. Multimodal brain tumor classification[C]//Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6. Springer International Publishing, 2021: 475-486.
52. Cho H H, Park H. Classification of low-grade and high-grade glioma using multi-modal image radiomics features[C]//2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017: 3081-3084.
53. Hsu W W, Guo J M, Pei L, et al. A weakly supervised deep learning-based method for glioma subtype classification using WSI and mpMRIs[J]. Scientific Reports, 2022, 12(1): 6111.
54. Xiao T, Hua W, Li C, et al. Glioma grading prediction by exploring radiomics and deep learning features[C]//Proceedings of the Third International Symposium on Image Computing and Digital Medicine. 2019: 208-213.
55. Isensee F, Jaeger P F, Kohl S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature methods, 2021, 18(2): 203-211.
56. Islam, M.; Wijethilake, N.; Ren, H. Glioblastoma multiforme prognosis: MRI missing modality generation, segmentation and radiogenomic survival prediction. Comput. Med. Imaging Graph. 2021, 2021, 91.
57. Sun, L.; He, Y. Survival prediction of LGG patients based on deep learning of MRI data and genomic data integration. In Proceedings of the 2021 China Automation Congress (CAC), Hangzhou, China, 22–24 October 2021; pp. 6473–6477.
58. Wijethilake, N.; Islam, M.; Ren, H. Radiogenomics model for overall survival prediction of glioblastoma. Med. Biol. Eng. Comput. 2020, 58, 1767–1777.
59. Jiang, S.; Zanazzi, G.J.; Hassanpour, S. Predicting prognosis and IDH mutation status for patients with lower-grade gliomas using whole slide images. Sci. Rep. 2021, 11, 16849.
60. Braman, N.; Gordon, J.W.H.; Goossens, E.T.; Willis, C.; Stumpe, M.C.; Venkataraman, J. Deep Orthogonal Fusion: Multimodal Prognostic Biomarker Discovery Integrating Radiology, Pathology, Genomic, and Clinical Data. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 27 September 2022–1 October 2022; Springer: Berlin/Heidelberg, Germany, 2021; Volume 12905, pp. 667–677.
61. Sohl-Dickstein J, Weiss E, Maheswaranathan N, et al. Deep unsupervised learning using nonequilibrium thermodynamics[C]//International conference on machine learning. pmlr, 2015: 2256-2265.
62. Luo C. Understanding diffusion models: A unified perspective[J]. arXiv preprint arXiv:2208.11970, 2022.
63. Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//International conference on medical image computing and computer-assisted intervention. Cham: Springer International Publishing, 2016: 424-432.
64. Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 fourth international conference on 3D vision (3DV). Ieee, 2016: 565-571.
65. Yu L, Cheng J Z, Dou Q, et al. Automatic 3D cardiovascular MR segmentation with densely-connected volumetric convnets[C]//International conference on medical image computing and computer-assisted intervention. Cham: Springer International Publishing, 2017: 287-295.
66. Li W, Wang G, Fidon L, et al. On the compactness, efficiency, and representation of 3D convolutional networks: brain parcellation as a pretext task[C]//International conference on information processing in medical imaging. Cham: Springer International Publishing, 2017: 348-360.
67. Isensee F, Jaeger P F, Kohl S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature methods, 2021, 18(2): 203-211.
68. Li X, Chen H, Qi X, et al. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes[J]. IEEE transactions on medical imaging, 2018, 37(12): 2663-2674.
69. Zheng H, Zhang Y, Yang L, et al. A new ensemble learning framework for 3D biomedical image segmentation[C]//Proceedings of the AAAI Conference on Artificial.
70. Fu Y, Lei Y, Wang T, et al. A review of deep learning based methods for medical image multi-organ segmentation[J]. Physica Medica, 2021, 85: 107-122.
71. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
72. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. CoRR, abs/1512.03385, 2015. <URL:http://arxiv.org/abs/1512.03385>.
73. Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. CoRR, abs/1608.06993, 2016. <URL:http://arxiv.org/abs/1608.06993>.
74. HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 7132-7141.
75. JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks[C]//Advances in Neural Information Processing Systems. Montreal Convention Center, Montreal, Canada: Curran Associates, 2015.
76. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[J]. arXiv preprint arXiv:1503.02531, 2015.
77. 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. Cham: Springer international publishing, 2015: 234-241.
78. Shusharina, N., & Bortfeld, T. (2021). Glioma Image Segmentation for Radiotherapy: RT targets, barriers to cancer spread, and organs at risk (GLIS-RT) [Data set]. The Cancer Imaging Archive. <https://doi.org/10.7937/TCIA.T905-ZQ20>
79. Menze B H, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS) IEEE Trans Med Imaging, 2015, 34(10): 1993-2024.
80. Dhariwal P, Nichol A. Diffusion models beat gans on image synthesis[J]. Advances in neural information processing systems, 2021, 34: 8780-8794.
81. Heusel M, Ramsauer H, Unterthiner T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[J]. Advances in neural information processing systems, 2017, 30.
82. Nash C, Menick J, Dieleman S, et al. Generating images with sparse representations[J]. arXiv preprint arXiv:2103.03841, 2021.
83. Kynkäänniemi T, Karras T, Laine S, et al. Improved precision and recall metric for assessing generative models[J]. Advances in neural information processing systems, 2019, 32.
84. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv:1412.6980, 2014.
85. Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis[J]. arXiv preprint arXiv:1809.11096, 2018.
86. Karras T, Laine S, Aittala M, et al. Analyzing and improving the image quality of stylegan[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 8110-8119.
87. A. Krizhevsky, I. Sutskever, and G. E. J. A. i. n. i. p. s. Hinton,“Imagenet classification with deep convolutional neural networks,” vol.25, 2012.
88. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
89. Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
90. Touvron H, Cord M, Douze M, et al. Training data-efficient image transformers & distillation through attention[C]//International conference on machine learning. PMLR, 2021: 10347-10357.
91. K. He, X. Zhang, S. Ren, and J. Sun, ”Deep residual learning for image recognition.” pp. 770-778.
92. K. Simonyan, and A. J. a. p. a. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014.
93. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger,” Densely connected convolutional networks.” pp. 4700-4708.
94. Bjoern Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Niklaus Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats), 2018. URL https://www.med.upenn.edu/cbica/brats2018/data.html. Accessed: 2024-09-17.