

面向医学图像分割的深度学习模型架构与性能评估方法综述

李辉¹, 刘佳煜¹, 徐雅萍²

(1. 北京化工大学信息科学与技术学院, 北京 100029; 2. 中日友好医院骨科, 北京 100029)

摘要: 医学图像分割在多模态成像数据中实现病灶或结构的像素级定位是支撑辅助诊断与临床决策的关键任务。针对医学图像分割网络架构快速演化与评价指标存在的语义歧义、统计不稳等局限, 旨在系统梳理网络结构、任务特征和评价指标三者间的适配关系, 揭示方法发展路径与性能边界, 构建面向实际应用需求的结构-指标匹配机制。基于 2020—2025 年 Web of Science 核心数据库的代表性文献, 首先梳理 Transformer、图神经网络 (GNN)、扩散模型等主干架构的设计机制与演化路径; 然后总结轻量化、混合结构及提示引导范式的关键特征; 接着结合公开数据集实证研究, 对不同网络结构在器官、肿瘤与脑组织等典型任务中的分割性能进行定量对比, 涵盖 95% 豪斯多夫距离 (HD95)、Dice 相似系数 (DSC)、交并比 (IoU) 等常用指标, 并识别出 HD95 在边界复杂任务中波动较大、DSC 对小目标敏感性不足、IoU 在结构区分方面存在局限等问题; 最后进一步揭示了指标误用与任务特征不匹配的统计根源, 构建了任务结构-指标推荐映射, 提出基于任务粒度的指标选择策略, 并探讨动态网络、自监督学习、跨模态建模等方向对模型泛化能力的潜在促进作用。

关键词: 医学图像分割; 深度学习; 网络架构; 评价指标体系; 任务适配

中图分类号: TP391.41

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0253035

Review on Deep Learning Model Architectures and Performance
Evaluation Methods for Medical Image SegmentationLI Hui¹, LIU Jiayu¹, XU Yaping²

(1. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China;

2. Department of Orthopaedics, China-Japan Friendship Hospital, Beijing 100029, China)

【Abstract】 Medical image segmentation enables pixel-level localization of lesions or anatomical structures in multimodal imaging data and serves as a key foundation for computer-aided diagnosis and clinical decision-making. This study addresses the rapid evolution of medical image segmentation network architectures and the inherent limitations (semantic ambiguity and statistical instability) of existing evaluation metrics. This study aims to systematically examine and delineate the alignment among network structure, task characteristics, and evaluation metrics; reveal the method development path and performance boundaries; and establish a structure-metric matching mechanism tailored to practical clinical needs. Based on representative literature from the Web of Science Core Collection between 2020 and 2025, this study first reviews the design mechanisms and evolutionary pathways of core architectures, such as Transformers, Graph Neural Networks (GNNs), and Diffusion Models (DMs), and then summarizes the essential characteristics of lightweight, hybrid, and prompt-guided paradigms. Subsequently, by integrating empirical studies on public datasets, a quantitative comparison is conducted across different architectures in typical segmentation tasks involving organs, tumors, and brain tissues, covering common metrics such as the Dice Similarity Coefficient (DSC), 95% Hausdorff Distance (HD95), and Intersection over Union (IoU). The results indicate that HD95 exhibits high variability in boundary-complex tasks, DSC shows limited sensitivity to small targets, and IoU presents insufficient structural discrimination capability. Furthermore, this study reveals the statistical causes underlying metric misapplication and task-metric mismatch; constructs a task-structure-to-metric recommendation mapping; proposes a task-granularity-based metric selection strategy; and explores how dynamic networks, self-supervised learning, and cross-modal modeling contribute to the enhancement of model generalization.

【Key words】 medical image segmentation; deep learning; network architecture; evaluation metric system; task adaptation

基金项目: 中央高水平医院临床科研业务费专项资金(2025NHLHCRFH04)。

作者简介: 李辉 (CCF 高级会员), 男, 副教授、博士, 主研方向为计算机图形学; 刘佳煜, 硕士研究生; 徐雅萍 (通信作者), 副主任护师、硕士。

收稿日期: 2025-09-18

修回日期: 2025-11-18

E-mail: yaping0535@163.com

0 引言

医学图像分割作为医学图像处理领域的核心任务,旨在将影像中的器官、组织或病灶区域进行像素级精确划分,为辅助诊断提供关键支撑。国内外大量研究推动了该领域方法不断创新,同时促成了高质量公开数据集建设和标准化评价体系形成。

随着深度学习的发展,该领域融合 Transformer、图神经网络(GNN)、扩散模型等多种建模机制。近年来,模型设计呈现显著趋势:一方面 Transformer 等基础结构不断优化,在二维/三维(2D/3D)图像处理、边界建模、小样本学习等方面进展突出;另一方面混合结构、轻量化网络及 Prompt 范式兴起,促使模型向“高性能-低复杂度-交互适应性”发展。在多模态、多任务并行背景下,医学图像分割技术已逐步从“模型驱动”转向“任务-数据驱动”融合阶段,亟需系统总结各类模型在不同任务中的结构优势与适用边界。

与此同时,目前广泛采用的重叠率[Dice 相似系数(DSC)、交并比(IoU)]、边界距离[95%豪斯多夫距离(HD95)、平均对称表面距离(ASSD)]等指标在不同任务下可能存在语义歧义、统计不稳定、可比性不足等问题,易导致性能误判。特别是在多类别分割、拓扑结构建模和小目标检测等任务中,现有指标难以全面反映模型实际能力,缺乏系统性指标选择方案。尽管已有综述从模型结构或任务分类角度进行总结,但针对“结构-任务-指标”三者之间适配关系的系统梳理仍相对欠缺。

在此背景下,结合近 5 年代表性文献,本文从研究趋势、结构分类与评价体系三方面,揭示医学图像分割方法的发展路径、性能表现及指标使用局限性,构建面向实际任务需求的结构-指标匹配机制,推动分割方法在多样化任务场景下的可解释性与高效演化。研究目标主要包括以下 4 个方面:

1)结合近 5 年代表性文献,基于发文趋势、关键词共现、技术领域演化等维度,分析当前医学图像分割领域的研究热点与任务转向。

2)围绕当前主流医学图像分割网络架构,归纳 Transformer、GNN、扩散模型等基础结构,进一步总结混合结构、轻量化网络、Prompt 引导等新兴范式,分析其特点与任务适配能力,明确主流方法的演化逻辑与数据驱动趋势。

3)系统整理当前常用的重叠率、边界距离和体积误差指标,探讨其任务适用性,比较不同类型网络结构在典型任务中的量化性能,识别指标驱动下的

性能趋势与结构优势。

4)面向器官、肿瘤等不同类型分割任务,探讨常用指标任务适用性,识别其在泛化能力评估方面的潜在误区,推动建立更合理、可解释的指标使用体系。

1 文献范围

本文基于 Web of Science 核心数据库,系统检索 2020 年 1 月至 2025 年 2 月期间医学影像人工智能相关文献,采用布尔逻辑组合关键词:("medical imaging" OR "radiology") AND ("artificial intelligence" OR "deep learning" OR "radiomics" OR "multimodal fusion"),排除非英文文献及非医学影像类研究,最终纳入有效文献 1 691 篇。

1.1 文章数量

如图 1(a)所示(彩色效果见《计算机工程》官网 HTML 版,下同),2021 年至 2024 年间该领域发文量呈稳定上升趋势,从 2021 年 253 篇增至 2023 年的高峰 429 篇,随后在 2024 年略微下降至 417 篇。2025 年由于统计时间尚未结束,仅收录 80 篇,数量显著偏低,预计后续将有所上升。

国家层面分析如图 1(b)所示,中国和印度发文最多,沙特阿拉伯和美国也表现活跃,持续有较高文献产出,英国、韩国等国家发文量波动较大且贡献相对有限。整体来看,相关研究呈现出全球化发展趋势,亚洲国家表现突出。

1.2 关键词共现

图 1(c)进一步展示了关键词数量及其占比随时间的变化趋势,以揭示技术热点与发展脉络。

首先,“deep learning”、“convolutional neural networks”和“image segmentation”构成核心关键词集群,体现深度学习的主流地位及卷积神经网络的广泛应用。“classification”和“feature extraction”等反映出对模型性能优化的不断研究。其次,“transfer learning”和“artificial intelligence”占比提升,预示了技术融合趋势,“covid-19”等特定疾病关键词反映出对公共卫生事件的快速响应,“machine learning”显示传统方法与新兴深度学习技术形成互补研究生态。此外,“diagnosis”和“computer-aided diagnosis”占比稳定,表明以临床诊断为导向,向实际诊断辅助工具转化趋势明确。“segmentation”、“medical imaging”和“computed tomography”指向临床影像分割需求,展现技术落地路径。

1.3 研究领域

如图 1(d)所示,Engineering Biomedical、Engineering Electrical Electronic 占据主导,反映了工

程学科与生物医学、信息技术的交叉融合是当前热点。计算机科学下属子领域 Information Systems、AI、Interdisciplinary Applications 表现突出,体现了数

据管理、智能算法、学科融合等方向的活跃性。总体来看,工程与计科驱动主导,医学基础学科作为辅助背景,呈现工程-计科-医学三者交汇的特征。

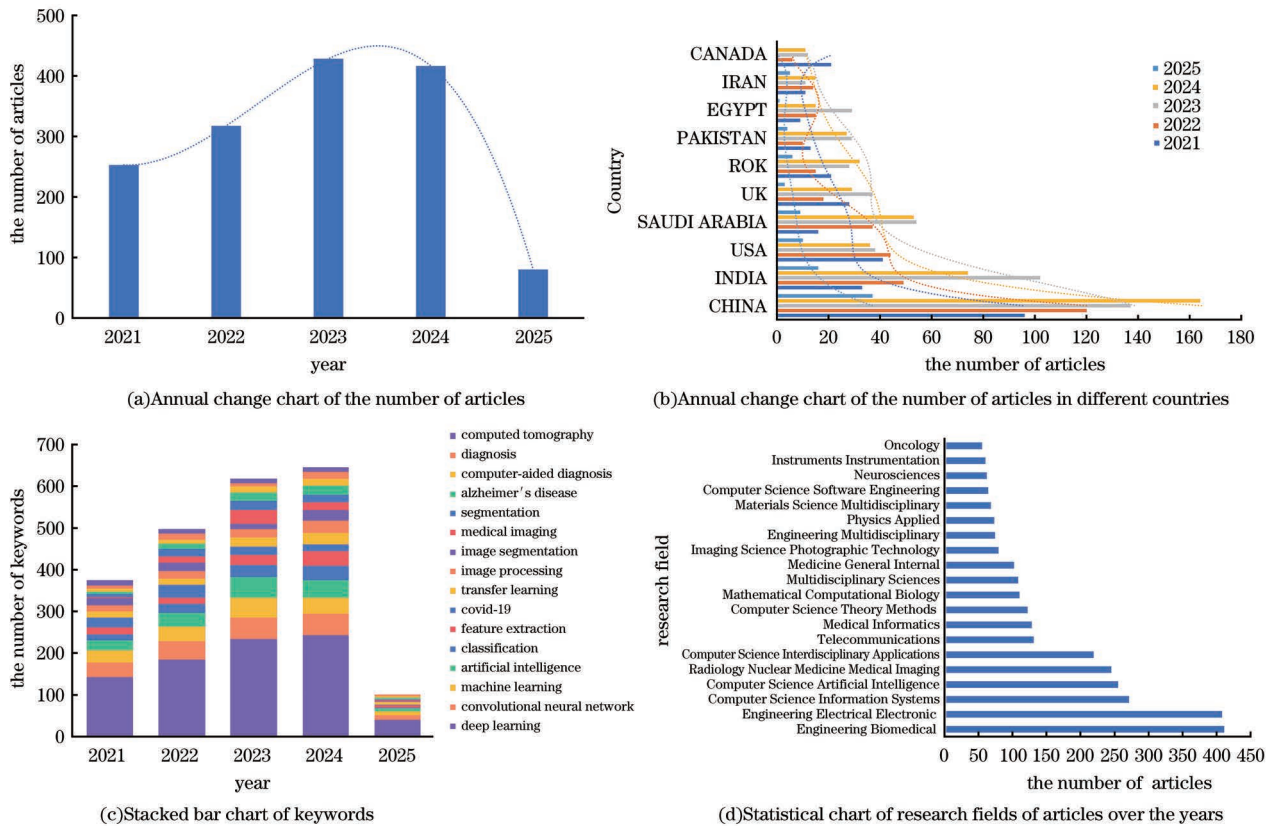


图 1 近 5 年文献统计图

Fig.1 Statistical chart of literatures in the past 5 years

1.4 数据集统计分析

当前公开数据集^[1]主要集中于多器官分割、肿瘤与病灶识别、脑组织分割、血管结构提取等四大方向。在多器官任务中,BTCV、AMOS 和 Synapse 等计算机断层扫描(CT)数据集因覆盖脏器丰富、标注一致性高而被广泛采用;BraTS 系列数据集是脑肿瘤分割的事实标准,提供多模态磁共振成像影像(MRI),支持跨机构评测;LiTS 与 KiTS 分别聚焦于肝脏与肾脏病灶识别,适用于评估小体积目标分割性能。血管与视网膜任务常用数据集包括 DRIVE、CHASE_DB1 与 STARE,强调细长结构与边界连续性。在脑组织结构建模方面,MRBrainS18 和 IBSR 提供高分辨率体素级标注。

2 医学图像分割网络架构

2.1 发展历程及现状

医学图像分割任务网络结构经历了 Transformer、扩散模型、GNN 等架构的演化,并逐渐融合混合建模、轻量优化及 Prompt 范式等技术,呈现出多样化发展趋势。本节按照上述脉络,对主

流分割网络进行系统分类和剖析。

2.1.1 单一主导范式

Transformer 在医学图像分割任务中^[2]的主要优势在于建模长程依赖和捕捉全局上下文。UNETR^[3]最早采用视觉 Transformer (ViT) 作为主编码器并结合卷积解码器,实现 3D 脑 MRI 图像的端到端建模。TransBTS^[4]引入 Transformer-Bottleneck 模块,增强语义建模。MedT^[5]通过门控轴向注意力机制设计双分支网络,分别捕捉局部和全局信息,增强对复杂结构的表达。UNETR++^[6]引入体素聚焦注意力,构建具有线性复杂度的高效 Transformer,缓解了 3D 图像处理的高计算负担。总体来看,Transformer 已逐步发展为高效独立建模体系,展现出良好的适应性。

扩散模型近年在医学图像分割任务中表现突出,主要适用于边界模糊、标签稀缺场景。FEU-Diff^[7]引入动态不确定性融合机制,通过逐步去噪对模糊边界区域具有良好鲁棒性。DDPM-UNet^[8]融合去噪扩散概率模型与 U-Net 架构用于 3D MRI 病灶区域建模,显著提升了小体积病变识别

性能。此外, Semantic Diffusion^[9] 在弱监督场景下通过语义引导生成伪标签, 增强结构一致性, 缓解高质量标签依赖问题。总体来看, 扩散模型正从生成建模扩展至多任务融合与弱监督优化, 展现出巨大潜力。

GNN 因能够建模非欧几里得空间结构, 逐渐被用于医学图像分割任务。早期 GCN-UNet^[10] 通过构建像素或区域邻接图以增强空间关系建模, 适用于结构复杂区域的分割。随后 GAUNet^[11] 引入图注意力机制, 动态建模节点权重, 在细粒度组织分割中表现良好。近期研究进一步拓展了其特征表达和

结构建模能力, AdaHGNN^[12] 利用 k 近邻构建超边, 捕捉高阶区域关系。文献[13]提出了跨模态协作框架, 保持了多模态图像分割一致性。TransGraphNet^[14] 融合了 Transformer 与 GCN, 实现了全局和局部特征的有效结合, 显著改善了复杂结构的分割精度。GNN 正从静态图卷积向动态邻域演化, 在复杂解剖与模态异质任务中展现出更强的结构建模能力。

综上, 如图 2 所示, 医学图像分割技术正呈现出多元并行的发展格局。这些不同路径相互交叉、彼此赋能, 共同驱动着整个领域的技术边界持续拓展。

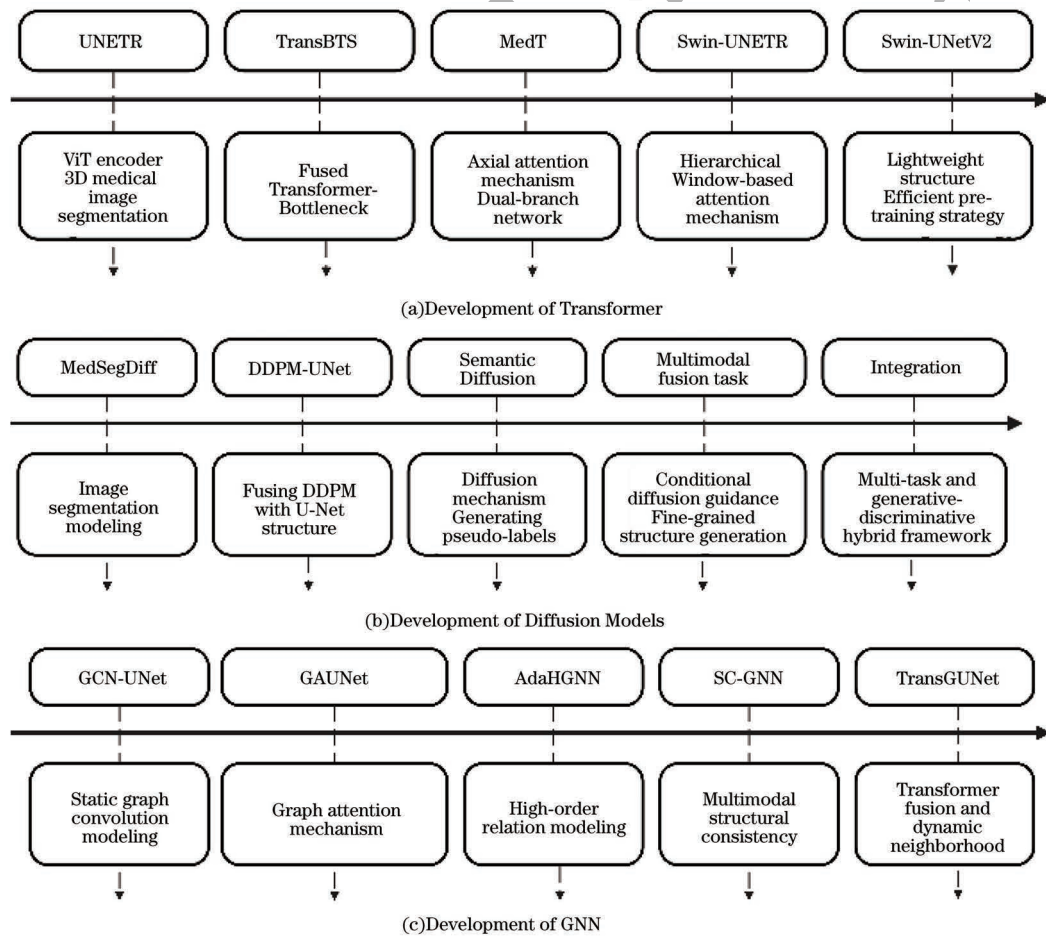


图 2 单一主导范式发展历程

Fig. 2 Evolution of single-dominant paradigms

2.1.2 混合架构

混合结构网络融合不同模型优势, 展现出较强的适应性与鲁棒性。典型代表包括卷积神经网络 (CNN) 与 Transformer 融合、扩散模型与判别模型结合等, 在性能和泛化能力上表现出显著优势。

首先, 以 TransUNet^[15] 为代表的 CNN-Transformer 混合架构有效增强了 U-Net 这类经典模型的长程依赖建模能力, 同时保留高分辨率定位能力。进一步升级的 TransFuse^[16] 采用“并行分

支”设计, CNN 与 Transformer 分支同时提取特征, 再通过 BiFusion 模块融合两者特征, 实现更高效的上下文融合和推理速度。其次, 扩散模型与判别模型的融合也呈现出了新的发展方向。HiDiff^[17] 提出在传统分割器之上加入扩散 refiner, 利用 BBDM (Brownian Bridge Diffusion Model) 对分割结果进行建模并细化。该混合式设计在多个模态数据集上展现了对小目标分割和跨数据集迁移场景的显著优势。再者, MultiResUNet 与 Transformer 的融合架

构^[18]将卷积提取的多尺度特征与 Transformer 的全局注意力结合,并引入深监督机制,通过 SAM (Spatial Attention Module) 优化训练过程,在边界模糊与弱对比度场景下,拥有更优的分割精度与鲁棒性。

近年来,生成式与序列建模范式在混合架构中表现出显著价值。相关工作表明:将扩散 refiner 与 U-Net/Transformer 编码器耦合^[19]能显著提高跨数据集泛化与弱监督场景的鲁棒性。与此同时,生成对抗网络(GAN)在跨模态映射、纹理与边界细化方面仍具优势,常用于生成伪样本或提升细节保真度^[20]。此外,基于状态空间模型(如 Mamba)的架构^[21]通过线性复杂度的长程依赖建模,为高分辨率医学影像的全局上下文捕获提供了高效的替代方案。

2.1.3 轻量化及交互范式

随着边缘部署和实时诊断需求的增长,轻量化医学图像分割网络成为重要研究方向。近年来,众多前沿模型通过 nmODE^[22]技术实现医学图像分割模型网络结构优化与参数量压缩。文献[23]结合 nmODE 与多尺度特征融合,实现高分辨率特征的低开销解码。文献[24]利用基于神经记忆 ODE 的轻量 U-like 网络,实现解码器瘦身与精度保持。相关研究采用离散化 nmODE^[25]以强化训练稳定性, NM-SSM^[26]在多器官分割中显著降低参数量并提升推理效率。这些成果验证了 nmODE 在单任务分割下,有效平衡了模型复杂度与分割精度,为低资源环境临床部署提供了新方向。

Prompt 范式通过引入交互提示与语义引导,为小样本与跨模态医学图像分割带来了突破。SAM-Med^[27]采用视觉嵌入与交互提示实现多器官即插即用分割。Medical SAM Adapter^[28]在主干中嵌入轻量化 Adapter 模块,仅调整少量参数即可适配多模态任务。LViT-Med^[29]将语言提示融入 Transformer 结构,引导目标区域识别与跨模态分割。

近期研究显示:不同参数高效微调(PEFT)机制对性能影响显著。文献[30]通过冻结主干,仅优化提示层与少量附加模块,即可接近全量微调精度。H-SAM^[31]引入分层 LoRA(Low-Rank Adaptation)模块,以低秩更新实现显存节约与快速收敛。Adapter-based^[28,32]展现出更强的跨任务迁移能力。总体来看,Tuning 具备最高精度、LoRA 资源效率最优、Adapter 在跨域场景下表现稳定。Prompt 范式正朝“提示驱动+PEFT 融合”方向演化,为高效通用

的分割模型奠定基础。

2.2 模型复杂度与计算可部署性分析

近年来,医学图像分割模型在精度提升的同时也面临计算复杂度与临床可部署性的挑战。大型 Transformer 和扩散模型虽在特征建模与生成一致性上表现优异,但其参数量($>10^8$)与计算需求使得在资源受限的影像工作站或移动终端部署困难。

为平衡精度与效率,研究者从结构压缩与自适应计算展开探索。一方面,轻量化网络通过通道剪枝、深度可分离卷积和知识蒸馏等方式,显著减少冗余计算;另一方面,硬件感知的动态推理策略可根据输入复杂度自适应分配计算资源。此外,混合精度训练与低比特量化技术进一步降低了算力占用,为实时手术导航和边缘设备部署提供了可行路径。

2.3 当前挑战

医学图像的专业标注依赖领域专家,成本高昂,导致高质量标注样本长期稀缺。为缓解监督信号不足对模型训练的限制,产生了弱监督和半监督学习方法。图像弱监督分割具有典型的三阶段策略,包括伪标签生成、图辅助学习及模型训练过程。然而,不同网络架构在弱标签环境下的鲁棒性仍差异明显,尤其在小样本^[33]、噪声标签条件下,现有 CNN 与 Transformer 混合模型依然难以保持稳定高精度的分割表现,表现出一定的过拟合和泛化能力不足。

临床诊断常依赖多模态影像,不同模态间数据异构且互补性强,有效融合多源信息^[34]以提升准确性是关键难题。当前多模态融合方法(如图 3 所示)多数在 CNN、Transformer 和 GNN 框架内引入跨模态注意力机制、多层次特征融合策略,实现信息交互和语义补充,但在保证临床实时性和鲁棒性的前提下,设计轻量高效且具强泛化能力的融合模型仍具挑战。

在临床应用中,模型的可解释性直接关系医生的信任度和决策采纳率。当前多数深度学习模型仍为“黑盒”^[35],难以明确展示预测依据。虽然已有基于注意力图、梯度可视化和图网络节点权重解释的方法在 Transformer 和 GNN 模型中有所探索,但仍缺乏系统化、标准化解释框架^[36],使得解释结果难以直接被临床医生理解和应用,影响实际临床推广。

综上,医学图像分割网络面临弱标签与少样本泛化挑战、多模态数据融合复杂、模型可解释性不足等多方面挑战。未来研究应结合稳健的弱监督学习、多模态协同融合、系统的可解释框架,向更精准、可靠及临床友好的方向发展。

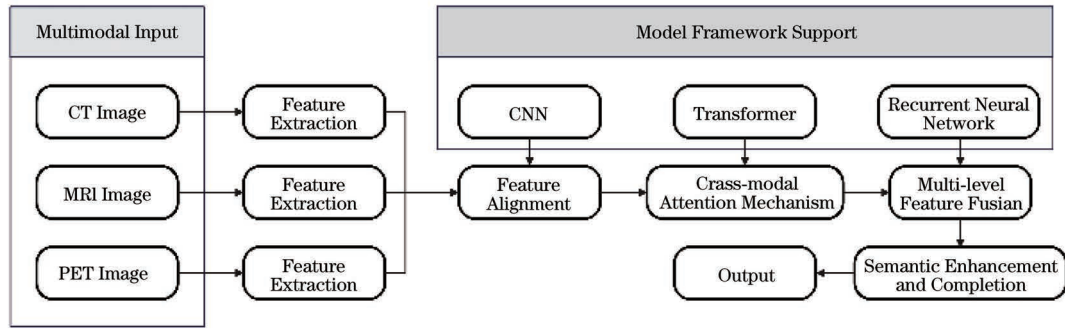


图 3 多模态融合方法流程图

Fig.3 Flowchart of multimodal fusion methods

3 常见评价指标及架构性能比较

3.1 常见指标

3.1.1 重叠度指标

DSC 是医学图像分割常用评价标准之一,衡量预测区域与真实区域的重叠程度。IoU 同样衡量交并比,但对不重叠区域惩罚更强。为直观对比其计算逻辑差异,如图 4^[37]所示,以预测分割集合 A、真实分割集合 B 为例,拆解了计算过程。其中, $D_{\text{Sensitivity}}$ 、 V_{PPV} 、 R_{IoU} 、 C_{DSC} 分别表示敏感度 (Sensitivity)、阳性预测值 (PPV)、IoU、DSC。

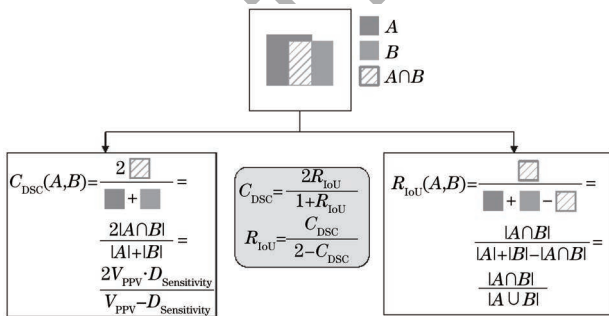


图 4 重叠度指标对比示意图

Fig.4 Schematic chart of overlap metric comparison

DSC 在实际应用中具有一定局限性,如图 5^[37]所示,包括问题表述不当、指标选择不合理以及指标应用不规范。

首先,DSC 对类别分布高度敏感,在类别不均衡的场景中容易产生显著偏差。如图 5(a)所示,在某些实际任务中,若参考标签中存在多个目标区域,但预测结果仅覆盖其中一个大目标区域,DSC 仍可能得到较高分数,而全面识别多个区域但稍有偏差的结果反而被赋予较低分数。因此,仅以 DSC 检测多个结构或小体积病灶极易导致语义层面的误判。

其次,DSC 不具备良好的边界敏感性,无法准确反映微小边界误差对整体分割质量的影响。图 5(b)展示了两个预测结果在分割边界上存在细

微差异 (Prediction 1 和 Prediction 2 仅 1 像素错位),但 DSC 分数分别为 0.80 与 0.79,差异被指标放大。因此,仅依赖 DSC 进行跨数据集或模型优劣排序,可能造成过度简化甚至误导结论。

最后,DSC 常被错误用于跨样本、跨层级的数据汇总,导致统计结果严重偏差。如图 5(c)所示,某医院 5 位患者包含不同数量图像样本,若直接对图像层面 DSC 值进行简单平均 ($\bar{O}_{\text{DSC}_{\text{img}}} = 0.8$),忽略患者层级的结构权重将导致结果被高样本量患者主导。按照患者为单位加权汇总,得到的真实平均值为 $\bar{O}_{\text{DSC}_{\text{patient}}} = 0.6$ 。在实际多中心、多病例任务评估中,层级结构被忽视极易造成泛化评估失真。

为增强指标的结构表征能力,研究者提出了 cDice(centerline Dice)^[38]以衡量分支结构的连通性。此外,Volume Dice 是 DSC 的体积扩展形式,用于 3D 医学图像中空间结构的整体分割质量评估。同时,Tversky 系数 (Tversky Index)作为 DSC 和 IoU 的推广形式,引入可调权重,在类别不平衡或小目标分割场景下应用广泛。

3.1.2 距离度量指标

豪斯多夫距离 (HD) 衡量预测边界与真实边界之间的最大距离,适合评估最坏情况边界偏移 (安全边界),其 95% 版本 HD95 可缓解极端点对整体评估的影响。

平均对称表面距离 (ASSD) 计算两个边界之间所有点的平均距离,提供更平滑的边界误差评估,常与 HD 联合使用以兼顾平均性能与最坏性能。

HD 及 HD95 常被用作主流解剖结构边界清晰度和空间定位精度的边界性能评价指标。然而,其在统计稳定性方面存在误用与过度依赖问题。

首先,HD95 并非真正鲁棒,易受局部误差影响,即便整体重叠度良好 (DSC 高),当分割存在集中漂移或小范围断裂时,其值也会显著上升。例如肺结节分割中两模型 DSC 相近 (约 0.88),因边缘微小跳跃,HD95 相差 3 倍以上,在小目标或复杂边

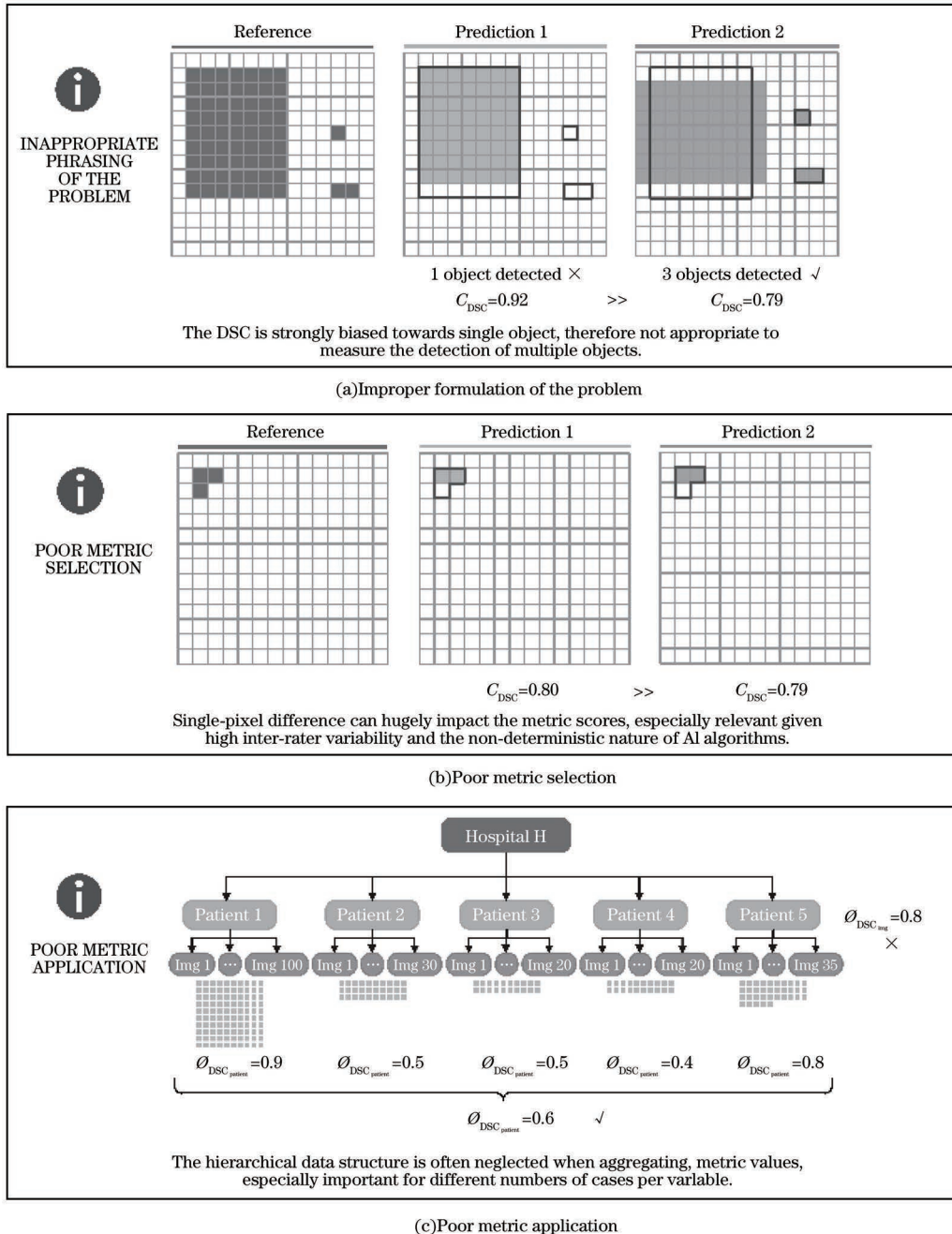


图 5 DSC 常见误用问题示意图

Fig.5 Schematic chart of DSC common misuse issues

界区域问题更明显。其次, HD95 难以解释语义问题, 仅反映单点最大误差近似, 无法体现整体边界结构完整性或拓扑连续性, 而临床更关注是否沿解剖边界正确贴合, 而非仅“最坏距离”。ASSD 则能更均衡度量边界误差, 适合评估整体贴合质量。此外, 边界指标应结合任务解读, 在小目标、边界模糊情形下, HD95 不宜作为唯一或首要指标。指标解释需回归目标本身, 而非单纯数值对比。

为克服 HD95 在小目标和局部错位中的波动问题, 文献[39]提出 Boundary F1 Score 以及局部化 HD(Localized HD), 更关注边界一致性与连接性,

贴近临床精准切边的需求评估。

3.1.3 体积误差指标

相对体积误差(RVE)通过计算预测体积与真实体积的相对差异, 评估分割结果的系统性偏倚。此外, 还可结合 Volume Dice 或绝对体积差(AVD)用于多尺度器官或病灶体积一致性验证, 从而更全面地表征全局定量性能与统计稳定性。

3.1.4 不平衡敏感性指标

F β -score 通过引入权重参数调整召回率与精确率的重要性, 以更关注小样本类别。BA(Balanced Accuracy)在多类别场景下可防止多数类主导总体

结果。MCC(Matthews Correlation Coefficient)和 Cohen's Kappa 通过同时考虑真阳性、真阴性、假阳性、假阴性,更适合类别比例差异显著的任务。此外,Precision-Recall 曲线下的面积(AUPRC)相较 ROC-AUC 更能反映正负样本比例失衡下的分类性能。在小病灶检测中应优先报告 $F\beta$ -score、MCC 或 AUPRC,而在多器官分割中则应兼顾 DSC 与 BA,以进行更全面的评估。

如表 1 所示,从重叠度、边界精度、体积一致性、

结构一致性 4 个维度,汇总了代表指标及用途。其中:Topology Error 指的是分割结果在结构连接性或形状完整性上与真实标签存在不一致;熵(Entropy)衡量预测结果不确定性或信息复杂度,适用于不确定性建模;期望校准误差(ECE)评估模型置信度与实际准确率的偏差,有助于提升临床可解释性和安全性;Uncertainty Dice 是将预测不确定性引入 DSC 评估,综合衡量分割准确性与模型输出不确定性之间的关系。

表 1 常见指标分类

Table 1 Common metric categories

Metric Dimension	Representative Metric	Typical Use Case
Overlap	Dice, IoU, Tversky Index	Basic prediction region accuracy
Boundary Accuracy	HD95, ASSD, Boundary F1 Score	Evaluates boundary errors in sensitive structures
Volume Consistency	Volume Dice, RVE	Assesses lesion volume
Structural Consistency	cdDice, Topology Error	Ensures connectivity and topology (e. g. , ventricles, blood vessels)
Uncertainty	Entropy, ECE, Uncertainty Dice	Reflects model uncertainty regarding input/parameters
Class Balance	Sensitivity, Specificity, BA, $F\beta$ -score, G-mean, AUPRC	Detection and recognition in class-imbalanced or few-shot scenarios

3.2 性能比较

本节系统梳理多种模型在多类医学影像数据集上

DSC、HD95 指标的表现,如表 2、表 3 所示,其中,空白表示原文献中未展示,“—”表示无相关测试结果。

表 2 模型-数据集-DSC 表

Table 2 Model-dataset-DSC table

Model	BTCV	ACDC	LiTS	BraTS	KiTS	AMOS	Synapse	%
A3DS-SegResNet ^[40]	80.69	90.69	79.28	90.79	81.11	87.27		
A3DS-SwinUNETR ^[40]	76.54	82.68	68.59	89.90	52.82	85.05		
CoTr ^[40]	81.95	90.56	79.10	90.73	84.59	88.02		
MedNeXt-L-k3 ^[40]	84.70	92.65	82.14	91.35	88.25	89.62		
MedNeXt-L-k5 ^[40]	85.04	92.62	82.34	91.50	87.74	89.73		
MISSFormer ^[41]		87.90						81.96
nnFormer ^[40-41]	80.86	92.40	77.40	90.22	75.85	81.55		86.57
nnU-Net ^[40]	83.08	91.54	80.09	91.24	86.04	88.64		
STU-Net-B ^[40]	83.05	91.30	79.19	90.85	86.32	88.46		
STU-Net-L ^[40]	83.36	91.31	80.31	91.26	85.84	89.34		
STU-Net-S ^[40]	82.92	91.04	78.50	90.55	84.93	88.08		
Swin-UNet ^[41]		90.00						79.13
Swin-UNETR ^[40-41]	78.89	91.29	76.50	90.68	81.27	83.81		83.48
TransBTS ^[41,42]	81.31			90.66	87.03	86.52		
TransUNet ^[3,41]	83.80	89.71						77.49
U-Mamba-Bot ^[40]	83.51	91.79	80.40	91.26	86.22	89.13		
U-Mamba-Enc ^[40]	82.41	91.22	80.27	90.91	86.34	88.38		
U-Net ^[41]								76.85
UNETR ^[41]	66.06	85.29		81.20	84.10	81.98		78.35

表 3 模型-数据集-HD95 表
Table 3 Model-dataset-HD95 table

Model	MSD-Spleen	MSD-Brain-WT	MSD-Brain-ET	MSD-Brain-TC	MSD-Brain-All	Synapse
AttUNet ^[3]	4.09	9.00	10.45	10.46	9.97	
CoTr w/o CNN ^[3]	4.75	11.49	9.59	12.58	11.22	
CoTr ^[3]	3.86	9.20	9.45	10.45	9.70	
MISSFormer ^[41]						18.20
nnFormer ^[41]						10.63
SETR-MLA ^[3]	4.09	15.50	10.24	14.72	13.49	
SETR-NUP ^[3]	4.12	14.42	11.72	15.19	13.78	
SETR-PUP ^[3]	4.11	15.25	11.76	15.02	14.01	
Swin-UNet ^[41]						21.55
Swin-UNETR ^[41]						10.55
TransBTS ^[3]	—	10.03	9.97	8.95	9.65	
TransUNet ^[3,41]	4.03	14.03	10.42	14.50	12.98	31.69
U-Net ^[3,41]	4.09	9.21	11.12	10.24	10.19	—
UNETR ^[3,41]	1.33	8.27	9.35	8.85	8.82	18.59
UNETR++ ^[41]						7.53

3.3 适应性分析

表 4 统计了模型在数据集中的综合表现。数据统计与归一化方法:缺失值填充,使用该列同行均值填充;Min-Max 归一化, $\text{Norm}(C_{\text{DSC}}) = (C_{\text{DSC}} - 52.82)/(92.65 - 52.82)$; $\text{Norm}(D_{\text{HD95}}) = (D_{\text{HD95}} - 1.33)/(31.69 - 1.33)$;加权综合, $S_{\text{Score}} = 0.6 \times \text{Norm}(C_{\text{DSC}}) + 0.4 \times (1 - \text{Norm}(D_{\text{HD95}}))$ 。其中, D_{HD95} 表示 HD95。

表 4 模型综合得分

Table 4 Model comprehensive score

Model	Comprehensive Score	$\text{Norm}(C_{\text{DSC}})$	$1 - \text{Norm}(D_{\text{HD95}})$
MedNeXt-L-k5	0.874	0.905	0.831
MedNeXt-L-k3	0.865	0.889	0.831
U-Mamba-Bot	0.860	0.861	0.859
STU-Net-L	0.859	0.867	0.847
CoTr	0.857	0.861	0.851
nnU-Net	0.855	0.848	0.863
STU-Net-B	0.854	0.845	0.863
TransUNet	0.853	0.851	0.855
U-Mamba-Enc	0.852	0.838	0.867
STU-Net-S	0.851	0.840	0.863
A3DS-SegResNet	0.849	0.832	0.875
TransBTS	0.848	0.826	0.887
nnFormer	0.847	0.830	0.875
Swin-UNETR	0.844	0.809	0.895
A3DS-SwinUNETR	0.834	0.778	0.907
UNETR	0.825	0.661	0.918

Transformer 架构(如 UNETR、Swin-UNet)在 3D 脑肿瘤分割(BraTS)中表现优异,凭借自注意力捕捉区域关系,DSC 和 HD95 表现优于传统 CNN。而 CNN 类模型(如 U-Net、nnU-Net)在小目标、边界模糊任务(如 ISIC 病变)中占优,能更好地提取局部细节,鲁棒性更高。扩散模型在边界精度方面表现出色,ASSD 通常低于 2.0 mm,在复杂区域重建中效果优于标准 CNN。

混合结构结合 CNN 局部感知与 Transformer/GNN 全局建模,大尺度结构分割泛化性强。Prompt 范式模型下的提示引导模型(如 SAM-Med)具备零样本/少样本推理能力,适用于跨模态、低标注或高变异性场景,虽不及定制模型,但其在 2D 中任务迁移灵活性高,人机交互潜力大。

然而,需注意模型性能受多种实验条件制约,包括损失函数设计、训练样本数量、超参数调优、数据预处理方式、输入分辨率等。不同研究中指标计算方法(如是否包含空白切片、边界粗糙度处理)也可能影响最终评估结果。因此,具体应用中应避免对单一指标孤立解读,需结合多维评价与任务场景(如器官大小、模态复杂度)进行综合判断。

4 基于子任务的指标误用分析与选择

4.1 指标误用分析

在医学图像分割任务的评估中,通用指标因定义简单、计算便捷已成为主流评估工具。然而,在任

务结构和语义差异显著的背景下,直接以其进行跨任务模型比较可能导致误解和错误结论^[38,43-44]。本节将从任务场景复杂度、判断标准模糊度、目标大小与形态难度方面展开分析。

医学图像分割任务结构异质性强。多器官任务目标往往形状规则、边界明确,而病灶或血管类任务则涉及小目标、不规则形状、分支拓扑等复杂结构,这类任务中常规 DSC 或 IoU 难以敏感地捕捉性能差异。例如 ISIC 数据集中的病灶分割结果即使边界明显错位^[45],仍可能获得高 DSC,掩盖实际语义误差。相比之下,拓扑敏感指标(如 cIDice)对血管等分支结构的预测解释力更高,提示指标应依据任务结构进行调整。

不同任务对模型能力要求不同,单一指标提升并不代表所有任务的泛化能力增强。例如 Transformer 架构在 BraTS 脑肿瘤任务中 DSC 显著提升(3%~5%),但在 ISIC 病灶任务中提升不足 1%。这并非模型性能下降,而是肿瘤任务对长程依赖建模的需求更高,Transformer 的全局建模能力在该语境下更具优势,而皮肤病变任务更依赖边界细节与局部感受野。这种误读风险在单一指标报告中尤为突出,可能造成“表面性能一致,实质机制不同”的认知偏差。

指标选择需匹配任务特性、临床目标,否则数值优也未必有效。小目标分割中 DSC 易不稳定,推荐 Boundary F1 Score 补充边界质量。多类别任务总体 DSC 易被主类主导,应使用 Macro Dice(各类别 DSC 均值,不考虑类别样本量,适合类别不均衡情况)或 Per-class IoU(单类别交叉比,反映某个类别预测区域与真实区域的重叠程度)避免类别不均衡产生掩盖效应。此外,拓扑任务应引入 cIDice/Connectivity Error^[46]以反映断裂、粘连等预测缺陷。

综上,应构建任务结构感知的指标体系:任务结构先行,根据目标大小、拓扑特征、语义结构分析需求;多维度组合,使用 DSC/IoU 捕捉面积的同时引入 cIDice、ASSD、Boundary F1 Score 等评估边界和结构;分类评估报告,对多类别任务采用 per-class 分析,避免主类误导;跨任务对比引入相对增益或归一化指标,辅助对模型迁移能力的公平评估。这种任务结构-指标匹配策略有助于提升指标解释力与评价外推性。

4.2 基于子任务的指标选择

医学图像分割任务种类繁多,广泛覆盖多个临床需求。由于目标结构在解剖形态、数据属性与临床价值方面存在显著差异,难以全面地衡量模型性

能。因此建立任务敏感型的评价体系显得尤为关键。

4.2.1 器官分割

器官分割任务聚焦于肝脏、肺、前列腺、胰腺等大体积、边界清晰的解剖结构。这类结构通常解剖稳定、组织差异明显,对模型的基本特征提取和结构泛化能力要求较高。其目标区域体积较大、边界相对规整、解剖结构标准化程度高,更适合评估模型整体分割性能,尤其是体素重叠度和几何一致性,强调结构整体的精准恢复与体积保持,而非边界模糊或微小器官所需的精细建模,广泛使用的指标包括 DSC、HD95/ASSD、RVE。

在器官分割场景中,不同网络性能表现有明显差异。CNN 模型凭借强大的局部建模与上下文融合能力,在器官结构中仍为主流选择。nnUNet 通过自适应超参数配置,已成为多器官公开基准的性能上限参考。Transformer 架构进一步增强了长程上下文建模能力,在处理形变较大的器官边界时结构一致性表现更优。混合模型兼顾 CNN 细粒度提取与 Transformer 全局语义表达,近年来取得领先效果。

4.2.2 肿瘤与病灶分割

肿瘤与病灶分割聚焦于识别形态多变、边界模糊、异质性显著的异常区域,但由于其结构小、边界模糊、类别极度不平衡,对模型的细粒度建模能力要求更高。病灶区域常表现出如下特性:体积小、形态复杂、边界不规则,与正常组织之间的灰度分布差异微弱;肿瘤内部异质性强,存在坏死区、增强区、实性区等不同生物学亚区;类不平衡严重,小病灶仅占图像总像素极少比例,常导致训练偏向背景。

为解决上述挑战,针对性评价指标包含:Tversky Index/FTL(Focal Tversky Loss,基于 Tversky Loss^[47]引入 focal 调制因子 γ 的改进损失)、DSC/Focal Dice(标准 DSC 加幂次 $\gamma > 1$ 的 focal 因子,使损失在易分割区域权重低、难分割区域权重高,提升网络对难样本的关注度)、HD95、Entropy/Uncertainty-aware Dice(将模型预测的不确定性因素融入 DSC 相关计算,常以损失函数形式体现)。

肿瘤分割任务具有强边界建模、小目标感知、类别不平衡鲁棒性等挑战。CNN 模型在多尺度融合与跳跃连接设计上持续优化,但对微小病灶或形变较大的区域仍有局限。Transformer 通过建模长程依赖提升对异质结构的全局理解,尤其在复杂脑肿瘤任务中检测能力优于 CNN。扩散模型用于复杂结构生成式建模,处理不确定边界。混合结构融合

CNN 局部细节捕捉与 Transformer 语义表达,取得了较好平衡。文献[48]提出基于注意力的多模态神经胶质瘤分割,该分割具有针对小强度差异的多注意力层,以关注小强度差异神经胶质瘤区域。

4.2.3 脑组织与皮层结构分割

脑组织与皮层结构分割旨在精确划分灰质、白质、脑室及大脑皮层等区域,为脑图谱构建、神经发育评估、神经疾病辅助诊断等提供基础。因涉及大量细粒度区域,需格外关注结构一致性与空间连通性。主要挑战包括:拓扑结构高度复杂,皮层折叠导致显著的边界不规则性和 3D 空间复杂度;解剖结构对称性强且个体差异显著,要求高泛化能力;部分区域体积极小,训练对标签误差高度敏感。常用指标为 DSC、cDice/TopoLoss^[49]、ASSD/HD。

针对这些特点,逐步发展出多种适配网络结构和优化策略。UNet 及变体因多尺度融合机制,能对复杂区域实现鲁棒建模,被广泛采用。Transformer 通过长距离建模能力,有效捕捉折叠结构的远程依赖关系,提升了解剖一致性。GNN 将脑区域显式建模为图结构,在保持区域拓扑关系方面显示出优越性;同时,引入拓扑保持损失函数在训练中^[50]显著提高了模型对复杂皮层区域的结构保持能力。

4.2.4 血管分割

血管分割任务在视网膜图像分析、脑血管结构重建、心血管疾病辅助诊断中意义重大,目标是精确分割出细长、分支繁多、拓扑复杂的血管结构,对连通性保持能力、小目标敏感性及边界连续性建模要求极高。血管结构普遍存在以下难点^[51]:目标细小且分支密集,传统卷积核易忽略细节导致小血管断裂或遗漏;血管连通性与拓扑结构是诊断关键,若出现分割断裂将严重影响临床判断;光照变化、成像模糊及病理区域伪影常造成边界模糊。常用评价指标包含 cDice、Precision/Recall、HD95/Boundary F1 Score。

CNN 早期多采用 U-Net 变体,增强残差连接和多尺度特征融合捕捉微小管道,但在高噪声环境下易发生断裂。拓扑感知网络结合拓扑保真度的损失函数 Topology-preserving Loss,显著减少预测断裂^[52]。GNN 将血管视为图,利用节点(支点)、边(血管段)显式建模,采用图注意力机制增强分支连通性,但推理复杂。Transformer 利用自注意力机制建模长程依赖,在血管分叉和交叉区域连通性修复能力更好。混合模型结合 CNN、Transformer 与扩散模型补全断裂片段,在噪声场景中恢复血管连续性,成为新兴趋势。

4.2.5 多器官分割

多器官分割任务多见于腹部 CT/MRI 等高复杂解剖场景,需同时识别多个不同尺度、形态和纹理特征的器官结构,常作为医学图像分割算法通用性的测试基准,在器官建模、术前规划、图谱配准等临床任务中具有重要的应用价值。其挑战在于:多类别异质性强,器官形状、大小、灰度值等方面差异显著,难以共享特征;类别不平衡,大器官占据主导区域,小器官精度易被忽略;解剖上下文复杂,需捕捉器官间的相对位置与结构约束。

常用评价指标包括:Mean Dice(Dice 平均值)/Macro-averaged IoU(每类单独计算 IoU,再对所有类别取算术平均)、HD、Per-class Sensitivity/Specificity^[53](每类器官分别评估识别能力,检测类别间性能差异)、Class-wise Metrics(按类分别计算的评价指标再汇总,展现每个类别的性能差异)。

UNet、VNet 等扩展至多通道输出以支持多类预测,但在类别间特征共享和不平衡处理上存在瓶颈。nnU-Net^[54]自动调参优化预处理与模型配置成为任务基线。TransBTS^[55]、SegMamba^[21,56]等引入状态空间模型 SSM,提升对器官间空间关系与上下文的建模能力。同时,文献[57]显式建模解剖结构约束,提升小器官检测能力。

4.2.6 交互式与弱监督分割

交互式与弱监督分割旨在降低高质量像素级标签依赖,结合用户提示(如点击、框选)、弱标签(如图像级标签、粗略掩码)或预训练模型提升性能,适用于少样本病灶分割、跨中心部署等。任务特征为不确定性高,输入提示变化、标签不完整,易导致输出结果不稳定。边界模糊,部分提示缺乏明确边界信息,难以捕捉细节。考虑人机交互,不仅要评估分割质量,还需衡量交互效率与响应准确度。

常用评价指标包括 Uncertainty-aware Dice/Entropy Map(评估模型在边界模糊区的信心表达能力)、Clicks vs Accuracy(度量达到目标精度所需用户提示数,是交互效率的核心指标)、Click-based Accuracy(以用户点击为提示时模型输出的分割准确率,衡量提示效率与模型响应能力)和 ECE。

SAM-Med^[58]迁移 SAM,微调适配灰阶结构与解剖约束。CLIP-MedSeg^[59]引导多模态分割,支持语言提示或弱标签驱动。MedClick^[60]支持点选、框选等交互方式,引入点到边界距离编码,有效提升多轮交互效率与精度。FastSAM-Med^[61]构建支持多模态提示交互范式,兼顾推理速度与响应灵活性。文献[62]使用注意力增强与伪标签挖掘策略,显著

提升了病灶弱标签的定位精度。

如表 5 所示,从分割任务出发,梳理特征并针对

性推荐评估指标,以便更清晰地结合不同医学图像

分割任务选择适配指标。

表 5 分割任务-特征-指标表

Table 5 Segmentation task-feature-metric table

Segmentation Task	Feature Keywords	Recommended Metrics
Organ Segmentation	Large volume, clear structure	Dice, HD95, RVE
Tumor/Lesion Segmentation	Small target, blurred boundary	Tversky Index, HD95, Entropy, Focal Dice
Brain Structure Segmentation	Complex topology, multiple brain regions	Dice, cDice, TopoLoss, ASSD
Vascular Segmentation	Slender branches, critical connectivity	cDice, Precision/Recall, Boundary F1 Score
Multi-Organ Segmentation	Multiple classes, large structural differences	Mean Dice, Class-wise Metrics, HD
Interactive/Weakly Supervised Segmentation	Uncertainty, interactive optimization	Entropy, ECE, Click-based Accuracy

5 结束语

医学图像分割研究正从以结构创新为核心的模型堆叠优化,逐步迈向网络架构、任务属性与评价体系三者协同发展的阶段。当前研究在任务范式上呈现“细化-融合”并行演进,分割对象从宏观器官拓展至微观结构^[63],交互式分割、弱监督与跨模态融合不断强化模型的语义理解与人机协作能力;在模型架构上,各类结构在表达能力、泛化性能与部署效率间持续权衡,轻量化与动态计算成为提升 3D 影像处理效率的重要方向,而提示驱动的分割模型虽具潜力,但在标签不足与模态偏移下仍面临适应性限制;在评价体系方面,传统指标在不同任务间的可比性与解释性存在差异,表面距离、拓扑保持性与不确定性评估等维度的重要性逐渐凸显,亟需构建面向任务特性的指标选择框架。总体来看,未来的发展应在高效灵活的网络设计与科学合理的评价体系之间形成协同,通过模型压缩^[64]、硬件感知优化、多指标组合评估等策略促进算法在临床中的可部署性与可靠性,从而加速医学图像分割技术从实验室性能向实际医疗价值的落地转化。

参考文献

[1] ANTONELLI M, REINKE A, BAKAS S, et al. The medical segmentation decathlon [J]. Nature Communications, 2022, 13(1): 4128.

[2] HE K L, GAN C, LI Z Y, et al. Transformers in medical image analysis[J]. Intelligent Medicine, 2023, 3(1): 59-78.

[3] HATAMIZADEH A, TANG Y C, NATH V, et al. UNETR: Transformers for 3D medical image segmentation [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Washington D. C., USA: IEEE Press, 2022: 1748-1758.

[4] WANG L, HU P, SHEN C, et al. TransBTS: multimodal brain tumor segmentation using Transformer [EB/OL]. [2025-08-11]. <https://arxiv.org/abs/2103.04430>.

[5] VALANARASU J M J, OZA P, HACIHALILOGLU I, et al. Medical Transformer: gated axial-attention for medical image segmentation [C] // Proceedings of the Medical Image

Computing and Computer Assisted Intervention. Berlin, Germany: Springer International Publishing, 2021: 36-46.

[6] NTANZI S, VIRIRI S. UNETR++ with voxel-focused attention: efficient 3D medical image segmentation with linear-complexity Transformers[J]. Applied Sciences, 2025, 15(20): 11034.

[7] GENG S, JIANG S, HOU T, et al. FEU-Diff: a diffusion model with fuzzy evidence-driven dynamic uncertainty fusion for medical image segmentation[J/OL]. IEEE Transactions on Neural Networks and Learning Systems: 1-16 [2025-08-11]. <https://ieeexplore.ieee.org/document/11165205>.

[8] GAO M, YANG Y, WANG H, et al. DDPM-UNet: denoising diffusion probabilistic models for 3D medical image segmentation[EB/OL]. [2025-08-11]. <https://arxiv.org/abs/2211.03364>.

[9] ZHENG H, ZHANG Y, ZHANG X, et al. Semantic Diffusion for weakly supervised medical image segmentation [C] // Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2023: 1-10.

[10] GAO Y, GLOCKER B, RUECKERT D. Graph U-Net for brain cortical parcellation [C] // Proceedings of MICCAI'19. Washington D. C., USA: IEEE Press, 2019: 1-8.

[11] ZHU M, XIAO Y, LIU J, et al. Graph attention U-Net for brain tumor segmentation[J]. Neurocomputing, 2021, 452: 360-372.

[12] CHAI S R, JAIN R K, MO S C, et al. A novel adaptive hypergraph neural network for enhancing medical image segmentation [C] // Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention. Berlin, Germany: Springer, 2024: 23-33.

[13] ZHOU X G, SUN Y Y, DENG M, et al. Robust semi-supervised multimodal medical image segmentation via cross modality collaboration [C] // Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention. Berlin, Germany: Springer, 2024: 57-67.

[14] ZHANG J, YE Z Y, CHEN M Y, et al. TransGraphNet: a novel network for medical image segmentation based on Transformer and graph convolution [J]. Biomedical Signal Processing and Control, 2025, 104: 107510.

[15] CHEN J N, MEI J R, LI X H, et al. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of Transformers[J]. Medical Image Analysis, 2024, 97: 103280.

[16] ZHANG Y D, LIU H Y, HU Q. TransFuse: fusing Transformers and CNNs for medical image segmentation [C] // Proceedings of the Medical Image Computing and Computer Assisted Intervention. Berlin, Germany: Springer, 2021: 14-24.

[17] CHEN T, WANG C H, CHEN Z H, et al. HiDiff: hybrid

- diffusion framework for medical image segmentation [EB/OL]. [2025-08-11]. <https://hub.baai.ac.cn/paper/ace71b6a-ccbc-42f1-8fca-eb9da0edcb>.
- [18] AL QURRI A, ALMEKKAWY M. Hybrid MultiResUNet with Transformers for medical image segmentation [J]. Biomedical Signal Processing and Control, 2025, 110: 108056.
- [19] XING Z H, WAN L, FU H Z, et al. Diff-UNet: a diffusion embedded network for robust 3D medical image segmentation [J]. Medical Image Analysis, 2025, 105: 103654.
- [20] NAGARE M, BUZZARD G T, BOUMAN C A. Texture matching GAN for CT image enhancement [J]. Journal of Mathematical Imaging and Vision, 2025, 67(4): 45.
- [21] XING Z H, YE T, YANG Y J, et al. SegMamba: long-range sequential modeling mamba for 3D medical image segmentation [C] // Proceedings of the Medical Image Computing and Computer Assisted Intervention. Berlin, Germany: Springer, 2024: 578-588.
- [22] LUO H Y, HE T, YI Z. A stable mapping of nmODE [J]. Artificial Intelligence Review, 2024, 57(5): 120.
- [23] NIU H, ZHOU Y X, YAN X H, et al. On the applications of neural ordinary differential equations in medical image analysis [J]. Artificial Intelligence Review, 2024, 57(9): 236.
- [24] HE Q, YAO X, WU J, et al. A lightweight U-like network utilizing neural memory ordinary differential equations for slimming the decoder [C] // Proceedings of IJCAI'24. Vienna, Austria: International Joint Conferences on Artificial Intelligence Organization, 2024: 821-829.
- [25] XU X, LUO H, YI Z, et al. A forward learning algorithm for neural memory ordinary differential equations [J]. International Journal of Neural Systems, 2024, 34(9): 2450048.
- [26] WANG Z, GU J, ZHOU W, et al. Neural memory state space models for medical image segmentation [J]. International Journal of Neural Systems, 2025, 35(1): 2450068.
- [27] MA J, ZHANG Y, GU S, et al. Segment anything in medical images with prompt-enhanced vision foundation model [J]. Nature Machine Intelligence, 2023, 5(9): 1135-1146.
- [28] WU J D, WANG Z Y, HONG M X, et al. Medical SAM Adapter: adapting segment anything model for medical image segmentation [J]. Medical Image Analysis, 2025, 102: 103547.
- [29] LI Z H, LI Y X, LI Q D, et al. LViT: language meets vision Transformer in medical image segmentation [J]. IEEE Transactions on Medical Imaging, 2024, 43(1): 96-107.
- [30] FISCHER M, BARTLER A, YANG B. Prompt tuning for parameter-efficient medical image segmentation [J]. Medical Image Analysis, 2024, 91: 103024.
- [31] CHENG Z H, WEI Q Y, ZHU H R, et al. Unleashing the potential of SAM for medical adaptation via hierarchical decoding [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2024: 3511-3522.
- [32] 赵欣, 李森, 李智生. 基于 CNN 和 Transformer 并行编码的腹部多器官图像分割 [J]. 吉林大学学报(理学版), 2024, 62(5): 1145-1154.
- ZHAO X, LI S, LI Z S. Abdominal multi-organ image segmentation based on parallel coding of CNN and Transformer [J]. Journal of Jilin University (Science Edition), 2024, 62(5): 1145-1154. (in Chinese)
- [33] BAI W, CHEN C, TARRONI G, et al. Semi-supervised learning for cardiac MRI segmentation via anatomical edge and shape constraints [Z]. 2022.
- [34] 赖小波, 许茂盛, 徐小媚. 多模态 MR 图像和多特征融合的胶质母细胞瘤自动分割 [J]. 计算机辅助设计与图形学学报, 2019, 31(3): 421-430.
- LAI X B, XU M S, XU X M. Automatic segmentation for glioblastoma multiforme using multimodal MR images and multiple features [J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(3): 421-430. (in Chinese)
- [35] PATRÍCIO C, NEVES J C, TEIXEIRA L F. Explainable deep learning methods in medical image classification: a survey [J]. ACM Computing Surveys, 2024, 56(4): 1-41.
- [36] TJOA E, GUAN C T. A survey on eXplainable Artificial Intelligence (XAI): toward medical XAI [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(11): 4793-4813.
- [37] MAIER-HEIN L, REINKE A, EISENMANN M, et al. Metrics reloaded: pitfalls and recommendations for image analysis validation [J]. Nature Methods, 2024, 21: 195-212.
- [38] TAHA A A, HANBURY A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool [J]. BMC Medical Imaging, 2015, 15: 29.
- [39] KERVADEC H, BOUCHTIBA J, DESROSIERS C, et al. Boundary loss for highly unbalanced segmentation [J]. Medical Image Analysis, 2021, 67: 101851.
- [40] ISENSEE F, WALD T, ULRICH C, et al. nnU-Net revisited: a call for rigorous validation in 3D medical image segmentation [C] // Proceedings of MICCAI'24. Berlin, Germany: Springer, 2024: 488-498.
- [41] SHAKER A, MAAZ M, RASHEED H, et al. UNETR++: delving into efficient and accurate 3D medical image segmentation [J]. IEEE Transactions on Medical Imaging, 2024, 43(9): 3377-3390.
- [42] ROY S, KOEHLER G, ULRICH C, et al. MedNeXt: Transformer-driven scaling of ConvNets for medical image segmentation [C] // Proceedings of MICCAI'23. Berlin, Germany: Springer, 2023: 405-415.
- [43] IRATNI M, ABDULLAH A, ALDHAHERI M, et al. Transformers for neuroimage segmentation: scoping review [J]. Journal of Medical Internet Research, 2025, 27: e57723.
- [44] 张啸成, 王涛, 田昕, 等. 基于移位窗口自注意力机制的新生儿脑区域图像分割 [J]. 吉林大学学报(理学版), 2024, 62(5): 1129-1137.
- ZHANG X C, WANG T, TIAN X, et al. Image region segmentation of neonatal brain based on self-attention mechanism of shifted windows [J]. Journal of Jilin University (Science Edition), 2024, 62(5): 1129-1137. (in Chinese)
- [45] 王斯豪, 张笃振, 杨昌昌. 基于双路径注意力机制和多尺度信息融合的皮肤病变图像分割 [J]. 计算机应用, 2025, 45(3): 978-989.
- WANG S H, ZHANG D Z, YANG C C. Skin Lesion image segmentation based on dual-path attention mechanism and multi-scale information fusion [J]. Journal of Computer Applications, 2025, 45(3): 978-989. (in Chinese)
- [46] O K S, GALDRAN A, RIERA-MARIN M, et al. Uncertainty aware segmentation quality assessment in medical images [C] // Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI). Athens, Greece: IEEE Press, 2024: 1-5.
- [47] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] // Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE Press, 2017: 2999-3007.
- [48] LIU X B, HOU S F, LIU S, et al. Attention-based multimodal glioma segmentation with multi-attention layers for small-intensity dissimilarity [J]. Journal of King Saud University (Computer and Information Sciences), 2023, 35(4): 183-195.
- [49] LO H, VASCONCELOS N, TRIVEDI M, et al. Topological constraints in deep learning for image segmentation [C] // Proceedings of CVPR'22. Washington D. C., USA: IEEE

- Press, 2022: 1-10.
- [50] WEN B, ZHANG H, BARTSCH D U G, et al. Topology-preserving image segmentation with spatial-aware persistent feature matching [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Washington D.C., USA: IEEE Press, 2025: 5762-5771.
- [51] LIU Z H, SUNAR M S, TAN T S, et al. Deep learning for retinal vessel segmentation: a systematic review of techniques and applications [J]. Medical & Biological Engineering & Computing, 2025, 63(8): 2191-2208.
- [52] SHIT S, PAETZOLD J C, SEKUBOYINA A, et al. clDice—a novel topology-preserving loss function for tubular structure segmentation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D.C., USA: IEEE Press, 2021: 1-15.
- [53] 邢莹, 闫晓华, 普程伟, 等. 全自动数字图像分析在外周血白细胞形态学复检中的临床应用 [J]. 中华医学杂志, 2016, 96(8): 634-639.
- XING Y, YAN X H, PU C W, et al. Clinical application of automatic digital image analysis in the review of peripheral blood leukocyte morphology [J]. National Medical Journal of China, 2016, 96(8): 634-639. (in Chinese)
- [54] 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.
- [55] WANG W X, CHEN C, DING M, et al. TransBTS: multimodal brain tumor segmentation using Transformer [C] // Proceedings of MICCAI'21. Berlin, Germany: Springer, 2021: 109-119.
- [56] 高路尧, 胡长虹, 肖树林. 基于超像素分割的图注意力网络的高光谱图像分类 [J]. 吉林大学学报(理学版), 2024, 62(2): 357-368.
- GAO L Y, HU C H, XIAO S L. Hyperspectral image classification based on superpixel segmentation with graph attention networks [J]. Journal of Jilin University (Science Edition), 2024, 62(2): 357-368. (in Chinese)
- [57] JEON Y S, YANG H, FU H, et al. Teaching ai the anatomy behind the scan: addressing anatomical flaws in medical image segmentation with learnable prior [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Washington D.C., USA: IEEE Press, 2025: 24024-24033.
- [58] MA J, HE Y T, LI F F, et al. Segment anything in medical images [J]. Nature Communications, 2024, 15(1): 654.
- [59] KOLEILAT T, ASGARIANDEHKORDI H, RIVAZ H, et al. MedCLIP-SAM: bridging text and image towards universal medical image segmentation [C] // Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention. Berlin, Germany: Springer, 2024: 643-653.
- [60] JI Y, ZHANG Z, ZHOU T, et al. MedClick: interactive medical image segmentation with click prompt [C] // Proceedings of MICCAI'23. Berlin, Germany: Springer, 2023: 1-12.
- [61] LUO Y, XU Q, FENG J, et al. Med-FastSAM: improving transfer efficiency of SAM to domain-generalised medical image segmentation [EB/OL]. [2025-08-11]. <https://openreview.net/pdf?id=ZL1Vb8LuE1>.
- [62] RAHMATI B, SHIRANI S, KESHAVERZ-MOTAMED Z. A hybrid approach for enhancing pseudo-labeling in medical images through pseudo-label refinement [J]. Scientific Reports, 2025, 15(1): 35161.
- [63] LIU X B, SONG L P, LIU S, et al. A review of deep learning-based medical image segmentation methods [J]. Sustainability, 2021, 13(3): 1224.
- [64] LIU X Y, DING X, YU L, et al. PQ-SAM: post-training quantization for segment anything model [C] // Proceedings of the International Conference on Computer Vision. Berlin, Germany: Springer, 2024: 420-437.

文字编辑 陆燕菲
栏目编辑 宋 圆