

硕士学位论文

基于 CT 影像的肺癌免疫治疗疗效预测算法
研究

Research on Lung Cancer Immunotherapy Efficacy
Prediction Algorithms Based on CT Images

一级学科

人工智能

二级学科

工业智能技术与系统

作者姓名

夏子清

指导教师

谢斌 教授

2026 年 5 月

中图分类号 TP391.41

学校代码 10533

UDC 004.9

学位类别 学术学位


硕士学位论文

基于 CT 影像的肺癌免疫治疗疗效预测算法研究

Research on Lung Cancer Immunotherapy Efficacy
Prediction Algorithms Based on CT Images

作者姓名	夏子清
一级学科	人工智能
二级学科	工业智能技术与系统
研究方向	智能感知与检测
二级培养单位	自动化学院
指导教师	谢斌教授

论文答辩日期 2026.5.22

答辩委员会主席 

中南大学

2026年5月

学位论文原创性声明

本人郑重声明，所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了论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，也不包含为获得中南大学或其他教育机构的学位或证书而使用过的材料。与我共同工作的同志对本研究所作的贡献均已在论文中作了明确的说明。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

学位论文作者签名：夏子清 日期：2026年5月28日

学位论文授权使用授权书

本学位论文作者和指导教师完全了解中南大学有关保留、使用学位论文的规定：即学校有权保留并向国家有关部门或机构送交学位论文的复印件和电子版，允许本学位论文被查阅和借阅。本人授权中南大学可以将本学位论文的全部或部分内内容编入有关数据库进行检索和公开传播，可以采用复印、缩印或其它手段保存和汇编学位论文。本人同意按《中国优秀博硕士学位论文全文数据库出版章程》规定享受相关权益。本人保证：毕业后以学位论文内容发表的论文作者单位注明中南大学；学位论文电子文档的内容和纸质学位论文的内容相一致。

延缓公开论文延缓到期后适用本授权书，涉密论文在解密后适用本授权书。

本学位论文属于：（请在以下相应方框内打“√”）

公开

延缓公开，延缓期限（__年__月__日至__年__月__日）

学位论文作者签名：夏子清

指导教师签名：谢心

日期：2026年5月28日

日期：2026年5月28日

（填写阿拉伯数字）

基于 CT 影像的肺癌免疫治疗疗效预测算法研究

摘要：肺癌是全球发病率和死亡率最高的恶性肿瘤之一，晚期患者的 5 年生存率长期处于较低的水平。目前，免疫治疗已成为肺癌治疗的重要手段，然而，并非所有患者都对免疫治疗敏感，有些甚至出现严重的不良反应。因此，治疗前有效预测肺癌患者接受免疫治疗的疗效具有重要意义。现阶段主要的预测手段高度依赖有创基因检测或病理分析，不仅侵入性强，而且检测费用昂贵。为了通过无创手段准确筛选出免疫治疗的潜在获益人群，本研究对肺癌患者的胸部 CT 影像进行特征挖掘，在实现自动化肺部肿瘤区域精准分割的基础上，进一步构建了基于肿瘤面积最大帧和基于多帧 CT 影像的免疫治疗疗效预测模型。本文的主要贡献如下：

(1) 针对临床中高质量标注数据获取难度大、标注成本高的问题，以及肺癌免疫治疗的疗效预测对肿瘤区域特征聚焦的需求，提出了一种基于双学生交叉指导与不确定性校正的半监督分割框架。为了打破单一学生网络在无标签数据上容易产生预测趋同的问题，本研究设计了单教师-双学生结构，并在训练过程中引入分歧损失和交叉指导机制；针对病灶边界模糊以及噪声干扰，通过多次随机推断捕捉像素级的预测波动性，构建不确定性掩码，并进一步提取特征空间的全局类别原型，利用语义一致性约束校正伪标签，提高肺部肿瘤的分割精度。在自建数据集 Xiangya-LIM 上进行了实验，20% 标注比例下 Dice 系数达到了 81.83%。

(2) 在肺部肿瘤自动化分割的基础上提出了一种基于肿瘤面积最大帧的肺癌免疫治疗疗效预测模型 MaxFrameNet。针对肺部肿瘤尺度跨度大的问题，本研究设计了自适应尺度感知门控模块以增强多尺度特征表征能力；针对肿瘤边缘结构与内部纹理异质性难以统一建模的问题，提出了空间域-频率域协同建模策略；同时，为缓解肺实质背景及炎症组织带来的噪声干扰，构建了结合全局语义建模与空间一致性约束的噪声抑制机制。在 Xiangya-LIM 数据集上进行了免疫治疗疗效预测实验，AUC 达到 0.7549；此外，在公开数据集 LIDC-IDRI 上的肺结节良恶性分类任务中，AUC 达到 0.8789，进一步验证了所提模型在提取肺部病理特征方面的有效性。

(3) 为了进一步提高肺癌免疫治疗疗效预测的精度, 提出了基于多帧影像信息的疗效预测模型 MultiFrameNet。针对全局特征提取不足的问题, 通过上述的 MaxFrameNet 分支提取肿瘤局部特征, 同时利用 VMamba 分支挖掘全局特征关联, 实现对局部-全局特征的深度解析; 针对肿瘤尺度不确定性与异构特征融合的问题, 引入跨尺度特征融合模块, 充分融合全局与局部特征。为充分利用帧与帧之间的空间信息, 设计了三维空间特征计算模块, 精确捕捉肿瘤的三维空间分布特征, 显著提高了模型的免疫疗效预测能力。MultiFrameNet 在 Xiangya-LIM 数据集上进行疗效预测实验, AUC 达到 0.8068; 在公开数据集 LIDC-IDRI 上进行了良恶性分类实验, AUC 达到 0.8910, 验证了所提出模型在肺部病灶形态特征与纹理信息提取方面具有良好的有效性与泛化能力。

图 20 幅, 表 16 个, 参考文献 92 篇

关键词: 肺癌; 免疫治疗疗效预测; 肿瘤分割; CT 影像
分类号: TP391.4

Research on Lung Cancer Immunotherapy Efficacy Prediction Algorithms Based on CT Images

Abstract: Lung cancer is one of the malignant tumors with the highest incidence and mortality worldwide, and the five-year survival rate of patients with advanced stages has long remained at a low level. In recent years, immunotherapy has become an important treatment strategy for lung cancer. However, not all patients respond favorably to immunotherapy, and some may even experience severe adverse reactions. Therefore, accurately predicting the therapeutic efficacy of immunotherapy for lung cancer patients before treatment is of great clinical significance. At present, most predictive approaches rely heavily on invasive genetic testing or pathological analysis, which are not only intrusive but also costly. To noninvasively identify patients who are likely to benefit from immunotherapy, this thesis explores feature representations from chest CT images of lung cancer patients. Based on the precise automated segmentation of lung tumor regions, prediction models for immunotherapy efficacy are further constructed using both the tumor maximum-area slice and multi-frame CT images. The main contributions of this thesis are summarized as follows.

(1) To address the challenges of limited availability of high-quality annotated clinical data and the high cost of manual annotation, as well as the need for tumor-focused feature extraction in predicting the efficacy of immunotherapy for lung cancer, a semi-supervised segmentation framework based on dual-student cross-guidance and uncertainty correction is proposed. To overcome the prediction homogenization problem that a single student network may encounter when learning from unlabeled data, this study designs a single-teacher dual-student architecture and introduces a discrepancy loss and cross-guidance mechanism during training. Furthermore, to mitigate the effects of ambiguous lesion boundaries and noise interference, multiple stochastic inferences are performed to capture pixel-level prediction variability, thereby constructing an uncertainty mask. Global class prototypes in the feature space are then extracted, and pseudo-labels are corrected through a semantic consistency constraint, improving the seg-

mentation accuracy of pulmonary tumors. Experiments conducted on the self-constructed Xiangya-LIM dataset demonstrate that the proposed method achieves a Dice coefficient of 81.83% with only 20% labeled data.

(2) Based on the automated segmentation of lung tumors, a lung cancer immunotherapy efficacy prediction model named MaxFrameNet is proposed, which utilizes the CT slice with the largest tumor area. To address the large scale variation of pulmonary tumors, an adaptive scale-aware gating module is designed to enhance multi-scale feature representation. To tackle the difficulty of jointly modeling tumor edge structures and internal texture heterogeneity, a spatial–frequency collaborative modeling strategy is proposed. Meanwhile, to alleviate noise interference caused by lung parenchyma background and inflammatory tissues, a noise suppression mechanism combining global semantic modeling and spatial consistency constraints is constructed. Immunotherapy efficacy prediction experiments conducted on the Xiangya-LIM dataset achieve an AUC of 0.7549. In addition, experiments on the publicly available LIDC-IDRI dataset for benign–malignant lung nodule classification obtain an AUC of 0.8789, further validating the effectiveness of the proposed model in extracting pathological features of lung lesions.

(3) To further improve the accuracy of lung cancer immunotherapy efficacy prediction, a multi-frame imaging-based prediction model named MultiFrameNet is proposed. To address the limitation of insufficient global feature extraction, the MaxFrameNet branch is employed to extract tumor local features, while a VMamba branch is introduced to mine global feature correlations, enabling deep analysis of local–global feature interactions. To cope with tumor scale uncertainty and heterogeneous feature fusion, a cross-scale feature fusion module is introduced to fully integrate global and local features. Furthermore, to effectively exploit spatial information across multiple frames, a three-dimensional spatial feature computation module is designed to accurately capture the 3D spatial distribution characteristics of tumors, significantly improving the model’s immunotherapy efficacy prediction capability. Experimental results show that MultiFrameNet achieves an AUC of 0.8068 for efficacy prediction on the Xiangya-LIM dataset, and an AUC of 0.8910 on the LIDC-IDRI dataset for benign–malignant classifi-

cation, demonstrating that the proposed model is effective and generalizable in extracting morphological and texture features of pulmonary lesions.

Keywords: Lung Cancer;Immunotherapy Efficacy Prediction;Tumor Segmentation;CT Images

Classification: TP391.4

目 录

第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	2
1.2.1 肺癌免疫治疗疗效预测现状	2
1.2.2 肺癌病灶分割研究现状	6
1.3 主要研究工作	8
1.4 本文组织结构	9
第 2 章 基于双学生交叉指导与不确定性校正的肺部肿瘤分割框架	10
2.1 引言	10
2.2 算法方案	10
2.2.1 总体思路	10
2.2.2 单教师双学生架构与交叉指导机制	11
2.2.3 基于不确定性估计的伪标签构建	14
2.2.4 基于特征原型的自适应语义一致性掩码校正	15
2.2.5 损失函数的设计	16
2.3 实验结果与分析	18
2.3.1 Xiangya-LIM 数据集	18
2.3.2 评价指标	19
2.3.3 实验设置	20
2.3.4 对比实验	21
2.3.5 消融实验	23
2.4 本章小结	26
第 3 章 基于最大帧多维度特征提取的肺癌免疫治疗疗效预测模型	27
3.1 引言	27
3.2 算法模型	27
3.2.1 网络整体结构	27
3.2.2 自适应尺度感知门控模块	28
3.2.3 空频双域特征增强模块	31

3.2.4	全局语义引导的噪声抑制模块	35
3.3	实验结果与分析	37
3.3.1	数据集	37
3.3.2	评价指标	39
3.3.3	实验设置	40
3.3.4	对比实验	40
3.3.5	消融实验	43
3.4	本章小结	46
第 4 章	基于肿瘤多帧融合的肺癌免疫治疗疗效预测模型	48
4.1	引言	48
4.2	算法模型	48
4.2.1	网络整体结构	48
4.2.2	双分支主干网络	49
4.2.3	跨尺度特征融合模块	53
4.2.4	空间特征计算模块	55
4.3	实验结果与分析	57
4.3.1	数据集	57
4.3.2	评价指标	58
4.3.3	实验设置	59
4.3.4	对比实验	59
4.3.5	消融实验	62
4.4	本章小结	65
第 5 章	总结与展望	66
5.1	总结	66
5.2	工作展望	67
	参考文献	69
	攻读学位期间主要的研究成果	78
	致 谢	79

第 1 章 绪论

1.1 研究背景与意义

肺癌是全球发病率和死亡率最高的恶性肿瘤之一，严重威胁着人类健康，并给全球医疗系统带来巨大的负担^[1]。2022 年，全球新增肺癌病例约 248 万例，死亡病例高达 181 万例，其高发病率和高死亡率使其成为公共卫生领域的一大挑战^[2]。由于肺癌的潜伏期较长，且初期症状不明显，许多患者在确诊时往往已经处于局部晚期或转移性阶段，这使其五年生存率一直处于较低水平。虽然近年来肺癌的诊疗技术取得了长足进展，手术、化疗、放疗等传统治疗方法得到了逐步优化，但对于晚期或转移性肺癌患者而言，这些治疗手段仍然存在显著的局限性，无法有效应对肿瘤异质性、治疗的耐药性等核心问题，导致治疗效果不佳，患者的生存质量也难以得到有效提升^[3]。

免疫治疗作为癌症治疗领域的一项重大突破，通过激活和增强患者自身免疫系统的抗肿瘤能力，能够实现对癌细胞的主动识别、攻击与清除，为晚期肺癌患者带来了新的治疗希望^[4]。目前，免疫检查点抑制剂已成为免疫治疗的核心手段，并广泛应用于肺癌的临床治疗中。部分患者通过免疫治疗可以实现肿瘤缩小、延长生存期，甚至获得长期缓解的理想效果^[5]。但需注意的是，免疫治疗并不适用于所有肺癌患者。相关研究表明，仅有 27% 至 45% 的肺癌患者能够从免疫治疗中明显获益^[6]。一些患者接受免疫治疗后，病情不仅没有得到好转，反而出现了疾病的进展^[7]；此外，免疫治疗还可能引发多种免疫相关的不良反应，如免疫性肺炎、结肠炎、内分泌紊乱等，严重时甚至威胁到患者的生命^[8]。免疫治疗的高昂费用也是一个不容忽视的问题，它不仅给患者及其家庭带来了沉重的经济压力，还在一定程度上限制了免疫治疗的广泛应用^[9]。更重要的是，目前临床上缺乏高效、便捷、无创的免疫治疗疗效预测方法，无法在治疗前精准识别哪些患者能够从治疗中获益，导致无效治疗的发生，不仅浪费医疗资源，还可能给患者造成不必要的身体和心理负担。

因此，如何在治疗前准确预测肺癌患者接受免疫治疗的疗效，已成为当前该领域亟待解决的核心问题。

目前，肺癌免疫治疗疗效预测的相关研究主要依赖基因组检测、蛋白标志物分析以及病理学检查等方法^[10-13]。这些方法通常需要通过穿刺或手术等手段获取肿瘤组织样本，属于有创操作，存在一定的出血和感染风险。同时，由于肿瘤具有明显的异质性，单一取样可能无法全面反映整体特征，存在一定的取样偏差。此外，这些检测方法成本较高、周期较长，也增加了患者的时间与经济负担，

难以满足临床快速筛查和大规模应用的需求。

计算机断层扫描（Computed Tomography, CT）作为肺癌临床诊断、分期以及疗效评估的常规影像学手段，具有无创、可重复、覆盖广等明显优势，为无创化免疫治疗疗效预测提供了重要的支持。目前，已有多项研究尝试采用影像组学方法来构建免疫治疗疗效预测模型^[14-16]。然而，传统影像组学方法通常依赖手工特征的设计，不仅对医生经验依赖较强、耗时较长，而且所构建的特征表达能力有限，难以充分刻画肿瘤复杂的形态结构与纹理异质性。同时，手工特征往往缺乏对高阶语义信息的建模能力，在不同数据集或扫描条件下也容易产生较大的分布偏移，从而限制了模型的泛化能力与预测性能。因此，本研究直接基于原始 CT 影像数据构建深度学习模型，通过端到端的特征学习方式自动挖掘与免疫治疗疗效相关的影像表征，避免对人工特征工程的依赖，以期更加客观、高效地实现肺癌免疫治疗疗效预测。

在此基础上，本研究进一步认识到，精准且自动化的肿瘤区域分割是提升疗效预测性能的重要前提。传统的肺部肿瘤分割通常依赖放射科医生手动勾画，这种方式不仅费时费力、效率低下，无法满足大规模临床数据处理和标准化诊疗的需求。因此，引入深度学习方法实现自动化的肺部肿瘤分割，能够有效提高分割的效率与一致性，为后续肿瘤特征提取和免疫治疗疗效预测提供更加精确且稳定的区域基础，从而进一步提升整体预测模型的性能与可靠性。

为了解决上述临床痛点和研究短板，本项目与湘雅医院开展合作研究，利用深度学习算法对肺癌患者的胸部 CT 影像进行特征挖掘。在实现自动化肺部肿瘤区域分割的基础上，准确筛选出免疫治疗的潜在获益人群，为肺癌个体化免疫治疗提供科学、有效的智能化临床决策支持，并推动肺癌精准诊疗水平的提升。

1.2 国内外研究现状

1.2.1 肺癌免疫治疗疗效预测现状

（1）基于机器学习的肺癌免疫疗效预测方法

传统的肺癌免疫治疗疗效预测方法主要依赖于统计模型，通过临床特征如吸烟史、年龄等或者 CT 图像上的纹理特征来进行预测。例如，Shen 等人^[17]收集了 63 名接受免疫治疗的患者共计 72 个病灶的 CT 图像，使用 Mazda 软件从这些图像中提取了一些纹理特征。这些纹理特征包括直方图、绝对梯度、运行长度矩阵、灰度共生矩阵、自回归模型和小波变换等多种参数，使用 Fisher 系数、互信息量和分类错误概率最小化结合平均相关系数的方法进行纹理特征选择，运用主成分分析、线性判别分析和非线性判别分析等方法来预测治疗效果，准确性达到了 81.9%。Benzekry 等人^[18]使用支持向量机、逻辑回归等十种机器学习方法

预测晚期非小细胞肺癌（Non-Small Cell Lung Cancer, NSCLC）患者对免疫治疗反应的决定因素，随机森林算法的准确率可以达到 0.68，AUC 达到 0.74。Wei 等人^[19] 招募了 123 和 99 名接受抗 PD-1/PD-L1 单药治疗或联合化疗的 NSCLC 患者，通过检测患者在开始治疗前和治疗六周后的血液样本，使用随机森林预测患者生存期，AUC 达到 0.725。Li 等人^[20] 采用包括逐步 Cox 回归、套索回归、岭回归、Cox 模型的偏最小二乘回归等 10 种机器学习算法，组合成 101 种算法，预测五年生存期的准确率达到 0.67 至 0.8。Liu Zichao 等人^[21] 使用 XGBoost 来捕获突变基因和免疫治疗优势之间的非线性关系，区分接受免疫治疗治疗的患者中的反应者和非反应者的 AUC 达到 0.67。

尽管这些机器方法在一定程度上能够预测肺癌免疫治疗的疗效，但它们往往受到特征选择的准确性和模型复杂性等因素的制约，导致预测结果的稳定性和可靠性不足，大多数研究的 AUC 指标都较低。为了提高免疫治疗疗效预测的精度，克服这些局限，研究者们开始积极探索新型算法和技术，特别是深度学习方法，希望构建复杂的神经网络模型来自动化地提取从医学影像数据、基因组数据、临床检查数据中的特征。

（2）基于深度学习的肺癌免疫治疗疗效预测

Li 等人^[22] 开发了一种基于 Vision Transformer (ViT) 网络的深度学习模型，通过分析患者的病理切片图像来预测他们对 PD-1/PD-L1 抑制剂治疗的反应，准确率达到 81%。Yang 等人^[23] 在基于概率的高斯混合模型的基础上引入 ResNet，构建了 DeepLRHE 框架以处理肺癌的 H&E 病理切片图像，结果显示，AUC 可以达到 0.87。Patkar 团队^[24] 提出了 Histo TME 模型，在由 895 名 NSCLC 患者的 H&E 病理切片图像及转录组数据上训练，预测免疫治疗的反应，测试集的 AUC 可以达到 0.75。还有不少研究^[11, 25] 发现通过深度学习方法分析 H&E 染色全切片图像中的肿瘤微环境等特征展现出了优于传统生物标志物如 PD-L1 的性能。

这些研究的数据基础主要是肿瘤的病理切片图像和基因组数据，通过深度学习模型识别出基因突变和表达模式，进而实现肺癌免疫治疗的疗效预测。然而，要获取肿瘤的组织图像或基因组数据往往需要做病理切片和基因检测，这不仅会对肺癌患者造成一定的伤害，还可能带来显著的经济负担。为了在无创的基础上对患者提供个性化的治疗方案参考，一些学者开始探索基于 CT 影像和临床数据的免疫治疗疗效预测的研究模式。这种方法的优势在于，CT 影像是一种非侵入性的检查手段，患者接受度高，且成本相对较低。通过分析 CT 影像，医生可以获得关于肿瘤大小、形状、位置和生长速度等关键信息，这些信息对于评估肿瘤的生物特性预测治疗效果至关重要。

Yang 等人^[26] 研究分析了 200 名接受单一抗 PD-1/PD-L1 药物治疗的晚期

NSCLC 患者基于 CT 图像的放射组学和临床数据，开发了一种基于全连接层和 SimTA 模块的模型用于区分出哪些患者对免疫治疗有反应，AUC 可以达到 0.80。He 等人^[27] 基于 CT 图片提取的影像组学联合建模无进展生存期和总生存期风险，在 236 例患者组成的数据集上验证，综合预测晚期 NSCLC 患者对免疫检查点抑制剂治疗的疗效，在回顾性队列中实现 C-index 0.75-0.80 的性能。Liao 团队^[28] 通过自动化高通量特征提取与深度神经网络层次化特征学习的联合优化，实现肺癌患者免疫治疗反应的精准个体化预测，对 206 例接受免疫治疗的肺癌患者队列进行了回顾性分析，AUC 可以达到 0.80 的性能。Chiu 团队^[29] 使用治疗前和治疗后的影像组学数据，提取和分析影像特征的动态变化，以预测非小细胞肺癌患者对免疫疗法的反应和临床预后，AUC 达到 0.81。Saad 等人^[30] 在 MD Anderson 癌症中心和斯坦福大学收集了 976 名转移性、EGFR/ALK 阴性的 NSCLC 患者接受免疫治疗前的 CT 扫描数据，并构建了一个 Deep-CT 的深度学习模型，用这些扫描图像来预测患者的总体生存率和无进展生存率。Cécile 等人^[31] 基于 CT 扫描的影像学特征来预测接受免疫治疗的转移性肺腺癌患者的无进展生存期（PFS），最好的 DeepSurv 模型的 C 指数为 0.631。

这些研究主要是基于 CT 图像提取出来的放射组学，但是也有一些研究证明深度学习特征可能比手工影像组学特征可以从影像中捕捉到更细微的信息。Kanakarajan 团队^[32] 提出的联合模型中前 10 个最具预测性的特征包括 WHOPS、吸烟状态、两个剂量特征和五个深度学习衍生特征，没有影像组学特征。影像组学特征缺失于顶部特征表明，手工制作的影像组学特征对联合模型的性能几乎没有贡献。Zhu 等人^[15] 发现在晚期非小细胞肺癌免疫治疗疗效问题上，单独的 ResNet34 深度学习模型在测试集上 AUC 达到 0.731，优于影像组学模型的 0.712，而且敏感性更高（0.761 vs 0.690），更适合识别获益患者。

She 等人^[33] 收集了来自四个不同医院的 274 名接受治疗的患者的数据，使用 ShuffleNetV2x05 的深度学习模型，分析患者在免疫治疗前两周内拍摄的 CT 图像，以区分患者是否会有良好的治疗反应，AUC 达到 0.75。Wu 等人^[34] 提出了一种多视图自适应加权图卷积网络用于预测 NSCLC 免疫治疗效果，在 107 例非小细胞肺癌患者中进行了评估，准确率达到 77.27%，AUC 达到 0.7780。现有的大多数研究主要采用影像组学特征和图像特征融合的方法，或是纯放射影像组学的方法，纯端到端深度学习直接处理 CT 图像的研究相对较少。

（3）3D 肺部 CT 图像分类

尽管当前关于肺癌免疫治疗的疗效预测的研究对 2D CT 影像的分析更普遍，尤其是通过影像组学与深度学习的结合，已有显著的预测成果。然而，随着技术的发展，3D CT 图像的使用逐渐成为肺癌免疫治疗预测领域的新方向。有研究证

明三维图像可以提供更多的信息^[35]。与 2D 图像相比, 3D 图像能够提供肿瘤在空间上的完整结构信息, 更加全面地反映肿瘤的体积和形态变化。

Xu 等人^[36]使用了一种结合卷积神经网络和递归神经网络的深度学习模型, 来分析肺癌患者在接受免疫治疗前以及治疗后 1 个月、3 个月和 6 个月的 CT 扫描图像, 预测患者的生存率、癌症是否会进展、是否会发生远处转移以及是否会局部复发。在预测 2 年生存率时, 模型的效果达到了较高的准确性, AUC 达到 0.74。He 等人^[37]收集了 327 名患者的 CT 图像, 使用 3D-Densenet 进行分析, 以区分 TMB 高的患者和 TMB 低的患者, 从而预测肺癌免疫治疗效果, AUC 达到 0.81。Xu 等人^[38]提出了一种先进的 2.5D 神经网络, 结合 2D 切片内特征提取和 3D 切片间特征聚合, 作者提出一种基于注意力机制的切片间特征的重新校准和累积梯度反向传播策略, 允许在整个卷上进行内存高效的参数优化, 预测肺癌免疫治疗疗效的准确率达到 95.80%。

除了肺癌免疫治疗的疗效预测之外, 在肺部的其他疾病上, Marentakis 等人^[39]将 Inception 和 LSTM 相结合来对非小细胞肺癌的两种常见病例腺癌和鳞状细胞癌进行分类, 在不需要详细的人工肿瘤分割前提下提供超过其他模型以及放射专家诊断 7% 至 25% 的优越性能。Yan 等人^[35]通过一系列实验发现, 3D CNN 比 2D CNN 更适合肺结节的分类任务, 可以更充分地利用 CT 切片之间的空间信息, 从而提高分类性能。Yan^[40]等人提出了 3DCE 网络, 将相邻的 CT 影像切片分别送入 2D 卷积网络中提取特征, 再将特征聚合起来进行预测, 增强了网络对切片的 3D 上下文信息的利用。Dou 等人^[41]提出使用 3D CNN 来降低肺结节检测任务中的高假阳率问题, 并在 LUNA2016 比赛的降假阳率模块中获得最优。Zhu 等人^[42]提出了一种基于 3D 深度双路径网络的结节分类框架 3D MLF-DCNN, 该方法用于检测肺部 CT 图像并将其分类为恶性或良性, F-1 评分为 99.25%, 准确率、敏感性和特异性均为 99.20%, 特异性为 99.17%。Tekade 等人^[43]提出了 3D 多径 VGG-like 网络, 将 U-Net 的预测和 3D 多径 VGG-like 网络相结合来对肺结节进行分类并检测恶性肿瘤水平, 准确率为 95.60%。Guo 等人^[44]构建了 3D ProNet 模型, 以三维 CT 图像作为输入, 区分肺腺癌、鳞状细胞癌和小细胞肺癌, AUC 达到 0.840。Zhang 等人^[45]提出了一种基于卷积组合单元 (CCU) 的三维卷积神经网络, 同时利用空间和光谱信息来对显微镜下高光谱图像中的肺癌进行分类。其中, CCU 用于融合不同卷积尺度获取的特征, 实验结果显示, 该模型准确率达到 0.962, 优于 2D-VGGNet。

在肺癌医学图像分类领域, 三维图像能够提供更为全面和连续的组织结构信息, 使得模型能够捕捉到更多与疾病相关的细微特征和模式。然而, 尽管三维图像在分类任务中显示出巨大潜力, 目前关于利用这些信息来预测肺癌免疫治疗疗

效的研究还相对较少。

1.2.2 肺癌病灶分割研究现状

(1) 全监督分割

全监督学习方法作为医学图像分割的主流范式，在肺部肿瘤分割领域取得了显著进展。Ronneberger 等人^[46]提出的 U-Net 通过编码器-解码器结构和跳跃连接，实现了多尺度特征的融合，在医学图像分割任务中表现优异，U-Net 及其变体，也成为了肺部肿瘤分割的基础架构。

针对肺癌的分割问题，许多研究者提出了多种改进的 U-Net 架构。Agnes 等人^[47]提出 Wavelet U-Net++，通过增强的跳跃连接实现了平均 Dice 系数 93.7% 的肺结节分割性能。Wu 等^[48]开发了 RAD-UNet 分割模型，专门解决分割不足和过度分割问题，在两个数据集上分别达到 93.56% 和 93.72% 的 Dice 系数。Zhang 等^[49]通过引入全局和局部密集连接以及扩张卷积来扩大感受野并整合多尺度信息，其研究实现了 92.13% 的平均交并比，较基线 U-Net 大大提升。Hou 等^[50]将残差结构、Transformer 模块和多尺度特征融合集成到 U-Net 中，在 LIDC-IDRI 数据集上达到 91.87% 的 Dice 系数和 86.88% 的 IoU。此外，多网络架构也被用于提升肺部肿瘤的分割性能。Zhao 等^[51]提出的 DSU-Net 采用级联双 U-Net 框架，而 Banu 等^[52]开发的 AWEU-Net 将注意力机制引入级联架构。Ming 等人^[53]提出双流编码器-解码器结构，解决 CNN+Transformer 融合浅、交互弱问题，在 LIDC-IDRI 数据集上进行肺结节的分割，Dice 系数达到 93%。Yan 等人^[54]提出了一种基于改进的 U-Net 和 Segment Anything 模型 (SAM) 的可提示分割方法，在 LUAN16 数据集上进行肺结节的分割实验，Dice 系数可达 80.12%。

全监督深度学习分割模型对大量精细标注数据的依赖一直是制约临床应用的瓶颈。以 CT 影像中的肺结节或肿瘤分割为例，专业放射科医师需要在每个切片上逐像素勾勒病灶边界，一个病例的完整标注往往耗费较长时间。这种高昂的标注代价使得大规模高质量标注数据的获取变得十分困难，从而在一定程度上限制了基于大规模标注数据的有监督模型训练与应用。

(2) 半监督分割

半监督学习为标注数据稀缺的困境提供了可行的解决方案，它利用少量标注数据建立监督信号，同时借助大量无标注数据挖掘数据的内在结构，从而在标注资源受限的情况下提升模型泛化能力^[55]。

对于肺癌或者肺结节的半监督分割，早期的研究主要依赖生成对抗网络生成合成样本以扩充训练集。Jin 等人^[56]将 GAN 合成样本混入训练集，增强 P-HNN 分割模型在肺边缘结节、小结节的鲁棒性，在 LIDC-IDRI 上进行了实验，Dice

系数达到 89.2%。Qin 等人^[57] 采用条件 GAN 生成肺结节 CT 合成图像与语义标签, 采用少量标注、GAN 合成样本和未标注数据共同训练 3D CNN 分割网络在 LIDC-IDRI 数据集上的 Dice 可以达到 0.8483。

这种基于 GAN 网络的方法受限于生成质量的稳定性, 如果 GAN 网络无法生成稳定的图像和标签, 反而会带来噪声干扰, 不少学者转向了自训练 (Self-training) 范式。Liu 等人^[58] 仅用自训练和知识蒸馏的方式, 用粗标注生成高质量伪标签来指导训练, 在 LUNA16 数据集 10% 的标注下, Dice 达到 0.864。Fredriksen 等人^[59] 提出一种基于师生架构的纯自训练肺部肿瘤分割方法, 利用少量像素级标注训练教师模型, 为大量仅含边界框的弱标注数据生成伪掩码, 再联合真实标注与伪掩码训练学生模型, 在无一致性正则约束下有效提升分割性能, 40 例精细标注和 1000 例框标注的情况下 Dice 达到 0.82。

由于自训练固有的确认偏差问题, 一些研究者开始引入一致性正则化技术。Liu 等人^[60] 提出了结合主观逻辑不确定性估计的双一致性正则化方法, 通过强弱扰动双重约束与不确定性加权, 有效提升半监督肺癌 CT 分割的鲁棒性与精度, 10% 标注的 LIDC-IDRI 数据集的 Dice 可以达到 0.813。Tarvainen 团队^[61] 提出 Mean Teacher 模型, 通过维护一个学生网络和一个指数移动平均 (EMA) 更新的教师网络, 强制学生网络与教师网络在输入扰动下保持预测一致性, 该框架被证实能有效降低模型认知不确定性。在肺部肿瘤分割任务中, 这一框架被广泛应用和改进。Wang 等人^[62] 提出了扰动增强均值教师框架 PE-MT, 通过解耦增强与不确定性聚焦, 在 LIDC-IDRI 数据集上标注 10% 的数据, Dice 可以达到 0.808。Su 等人^[63] 构建了双学生均值教师网络 DSMT, 通过异构学生互补与伪标签优化, 可以在极少量标注下实现肺肿瘤的高精度分割。

肺癌或者肺结节分割这一特定任务中, 纯一致性正则化方法面临解剖结构约束缺失的局限, 难以区分结节与血管等灰度相似结构。近期研究转向将一致性正则化与形状先验相结合, CoreFormer^[64] 利用扩散模型学习结节拓扑特征, 将一致性约束限制在解剖合理区域内, 在 LIDC-IDRI 数据集 5% 标签设置下将 Dice 提升至 88.7%。Liu 等人^[65] 提出了一种通用的半监督分割方法 UGCL, 在全局数据层面增强类内紧凑性与类间可分性, 构建稳定原型, 施加全局原型对比约束, 对病灶边界等不确定性区域施加一致性约束, 充分利用可靠互补信息。Lu 等人^[66] 等人提出了一种医学通用半监督分割方法 UPCoL, 在 Mean Teacher 架构下, 对无标签数据的扰动预测保持一致, 感知不确定性降噪, 同时学习紧凑、判别性的类别原型, 强制原型空间一致。

这一趋势表明, 任务特定的特征约束与通用一致性正则化的融合是当前肺癌半监督分割的主流发展方向。

1.3 主要研究工作

本研究以肺癌免疫治疗的疗效精准预测为核心目标，围绕 CT 影像开展深度学习方法的构建与验证，实现肿瘤区域的半监督自动化分割，并在此基础上分别建立基于肿瘤面积最大单帧影像与多帧切片的免疫治疗疗效预测模型，旨在辅助临床医生在治疗方案制定时做出更为合理的决策，从而提高患者的生存概率。

本文的主要研究思路和工作如图1-1所示，包括以下三点：

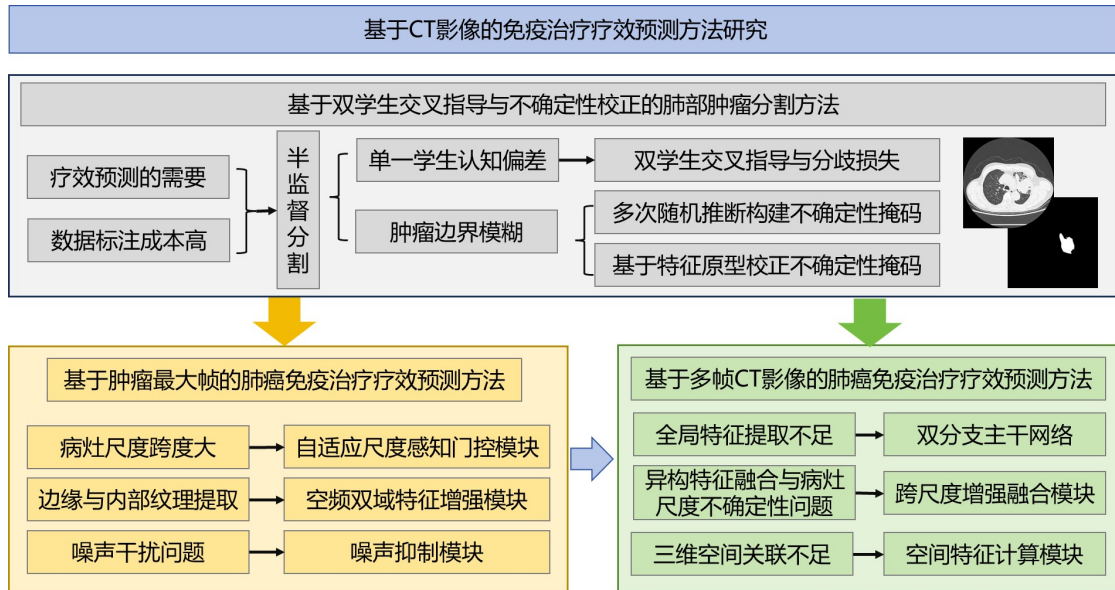


图 1-1 主要研究内容框架图

(1) 基于双学生交叉指导与不确定性校正的肺部肿瘤分割框架

针对临床中高质量标注数据获取难度大、标注成本高的问题，以及肺癌的疗效预测对肿瘤区域特征聚焦的需求，本研究提出了一种基于双学生交叉指导不确定性校正的半监督分割框架。针对单一学生网络在无标签数据上的预测同质化困境，本研究设计了双学生单教师结构，通过引入分歧损失和交叉指导机制，鼓励两个学生网络在不同的特征空间中进行探索，从而增强模型对未标注样本的多样性表征。针对肿瘤边界模糊引发的噪声干扰，本框架利用多步随机推断捕捉像素级的预测波动性，构建了不确定性掩码，并进一步通过提取特征空间的全局类别原型的语义一致性约束构建特征选择掩码，实现对肺部肿瘤区域的精准分割。

(2) 基于肿瘤区域面积最大帧的肺癌免疫治疗疗效预测

在肿瘤区域分割基础上，从单帧影像层面展开研究。肿瘤面积最大的轴位层面通常包含最具代表性的形态、密度及纹理信息，能够在有限数据条件下快速构建高可靠性的预测模型。针对肺部肿瘤尺度跨度大的问题，设计了基于多膨胀率卷积的多尺度特征提取结构，并引入轻量化门控模块实现尺度权重的动态分配。面向肿瘤边缘结构与内部纹理异质性难以统一建模的问题，提出了空间域与频率

域协同建模策略。针对肺实质背景及炎症等噪声干扰问题，构建了融合全局语义建模与空间一致性约束的噪声抑制机制。

(3) 基于多帧 CT 影像的肺癌免疫治疗疗效预测

针对单帧切片难以完整反映肿瘤在三维空间上的异质性特征和跨层空间关联信息的问题，本研究进一步构建基于多帧 CT 影像的疗效预测模型。在单帧疗效预测网络的基础上，引入全局特征提取网络 Vmamba，形成局部-全局特征提取双分支主干网络，针对异构特征的融合以及病灶尺度不确定性问题，提出了跨尺度融合与增强模块，针对三维空间特征关联不足的问题，提出了空间特征计算模块，强化肿瘤区域在不同层面的特征关联，来提升对免疫治疗疗效的判别能力，弥补单帧模型在空间信息表达上的不足。

1.4 本文组织结构

第一章为绪论。首先阐述了肺癌免疫治疗疗效预测的影像分析在临床诊疗中的迫切需求以及肺部肿瘤分割的重要意义；随后系统回顾了肺癌免疫治疗疗效预测、肺部肿瘤分割等方面的研究现状，并通过对比分析总结了现有方法的局限性与不足；在此基础上，明确了本课题的研究目标和三个研究内容，最后概括论文的结构安排。

第二章为基于双学生交叉指导与不确定性校正的半监督肺部肿瘤分割框架研究，详细描述了框架的设计思想、关键技术及其创新之处，重点介绍了双学生并行结构、分歧损失、交叉指导机制、多次随机推断、原型特征计算等技术细节，并在自建数据集 Xiangya-LIM 上验证了模型的分割性能。

第三章为基于肿瘤区域面积最大帧进行肺癌免疫治疗疗效预测的方法。针对肺部肿瘤尺度跨度大，肿瘤边缘结构与内部纹理异质性难以统一建模，以及肺实质背景及炎症等噪声干扰问题，在 ResNet34 的基础上进行改进，提出了 MaxFrameNet，在自建数据集 Xiangya-LIM 上验证了所提出的模型在疗效预测问题上的性能，并在 LIDC-IDRI 公开数据集良恶性分类实验，验证了模型在肺部肿瘤影像表征学习方面的稳定性与适应性。

第四章为基于多帧 CT 影像的肺癌免疫治疗疗效预测。针对全局特征提取不足、异构特征融合和病灶尺度不确定性问题，以及三维空间关联特征提取不足的问题，在第三章提出的 MaxFrameNet 的基础上构建了 MultiFrameNet 模型，在自建数据集 Xiangya-LIM 上验证了所提出的模型在疗效预测问题上的性能，并在 LIDC-IDRI 公开数据集上证明了模型稳健的肺部肿瘤影像理解能力。

第五章总结了全文的主要工作，展望了下一阶段的研究方向。

第2章 基于双学生交叉指导与不确定性校正的肺部肿瘤分割框架

2.1 引言

肺部肿瘤的大小、形态和边界等特征与患者的接受免疫治疗的疗效密切相关，而准确地分割出肿瘤区域是获取这些关键信息的第一步。传统的肿瘤分割方法需要放射科医师手动勾画，耗时费力，效率低下，难以满足大规模临床数据处理的需求。深度学习推动了医学影像处理的进步，凭借其强大的特征学习能力与端到端建模方式，促进了自动化分割的发展。目前大多数研究基于全监督学习的方法开展的，取得了不错的性能，但是同时也存在高度依赖大规模的标注数据的问题。

半监督学习（Semi-Supervised Learning, SSL）的出现为解决“标注数据稀缺”问题提供了有效途径。这种学习方式通过利用少量有标签数据引导模型学习核心特征，同时借助大量无标签数据挖掘数据内在分布规律，能够在降低标注成本的同时提升模型的泛化能力与鲁棒性。

本章围绕肺部肿瘤的分割问题展开研究，针对标注数据稀缺的问题，基于 Mean Teacher 架构，设计了一种双学生交叉指导与不确定性校正的框架。该框架通过引入双学生-单教师的结构，来增强特征学习的多样性，从而避免了单一网络容易陷入同质化问题。此外，框架利用多次随机推断的方式量化模型的不确定性，从而筛选出高质量的伪标签，并通过背景类和肿瘤类的特征距离计算进一步校正这些伪标签。同时，本章设计了多种损失函数的组合策略，包括有监督损失、一致性损失、交叉指导损失，以及双学生模型之间的分歧损失。

2.2 算法方案

2.2.1 总体思路

图2-1所示为基于双学生交叉指导与不确定性校正的半监督肺部肿瘤分割框架，这个算法主要由一个指数移动平均（EMA）教师模型（Teacher Model）和两个结构相同但参数独立的学生模型（Student 1, Student 2）构成。在训练过程中，有标签数据提供基础的监督信号，而无标签数据则通过教师模型生成伪标签（Pseudo-labels）。为解决伪标签可能带来的噪声累积问题，本文引入了以下三大核心机制：一是基于交叉指导和分歧损失的双学生特征解耦机制；二是基于多步随机推断的不确定性估计与伪标签构建；三是基于类别特征原型的自适应语义一致性掩码校正。

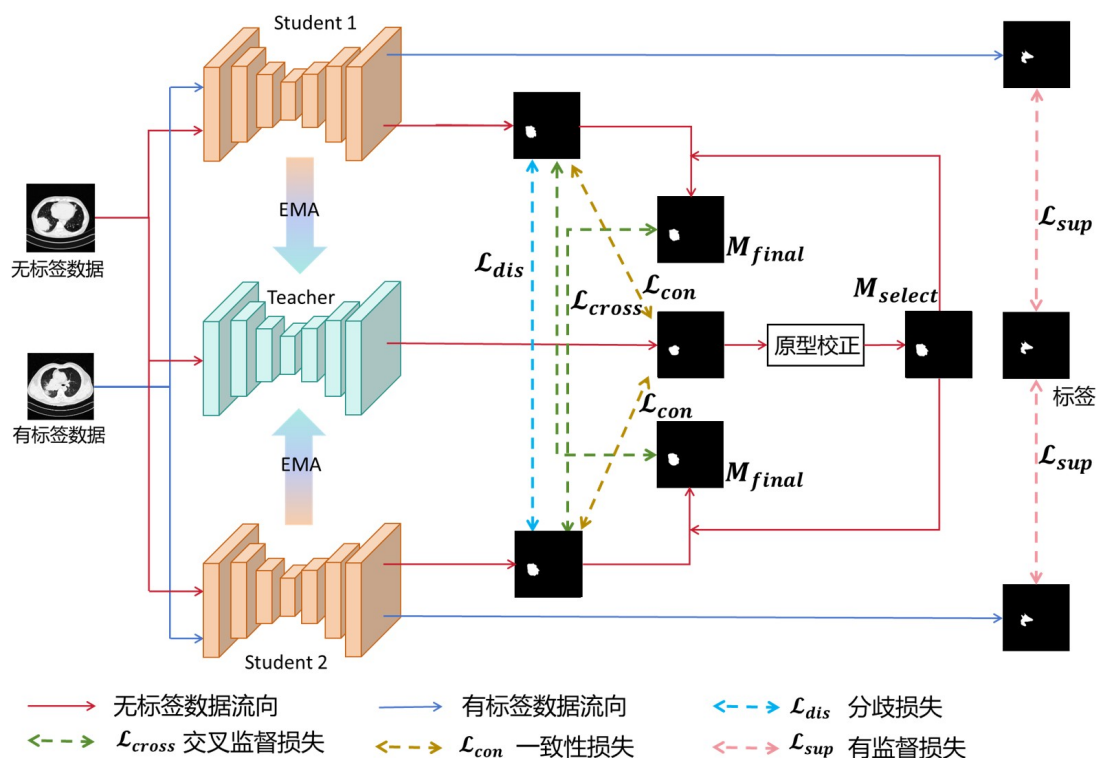


图 2-1 基于双学生交叉指导与不确定性校正的半监督肺部肿瘤分割框架

2.2.2 单教师双学生架构与交叉指导机制

传统的 Mean Teacher 框架由于学生与教师之间的强耦合，容易陷入“确认偏差”（Confirmation Bias）。针对这一局限性，本文设计了单教师-双学生架构（Single Teacher Double Students, STDS），引入分歧损失（Discrepancy Loss）和交叉指导机制（Cross-Teaching），有效地缓解了这一问题。

两个学生模型（Student 1, S_1 和 Student 2, S_2 ）以及教师模型（Teacher, T ）均采用了经典的 UNet^[46]（U 形卷积神经网络）。

UNet 是一个典型的 Encoder-Decoder 结构，分为编码器和解码器两个部分，如图 2-2 所示。编码器通过一层一层的卷积操作，从输入的原始 CT 图像中提取不同尺度的特征。随着卷积的进行，图像的尺寸会逐渐缩小，但是特征的通道数会增加，最后把提取的特征传递到一个残差块中。接着，解码器部分会通过反向操作将这些特征恢复成原始 CT 图像的大小。UNet 的亮点在于它的跳跃连接机制（Skip Connections），编码器中的特征会直接传递到解码器的对应位置，可以保留和传递各个尺度的特征，既能保留高层次的语义信息，又能捕捉到低层次的细节，避免了在图像缩小过程中丢失重要信息。这对于保留细节特别有用，尤其是在医学图像这种对精度要求非常高的任务中。所以在医学影像分割中，尤其是在肺部肿瘤分割等任务中，UNet 被广泛应用并取得了很好的效果。

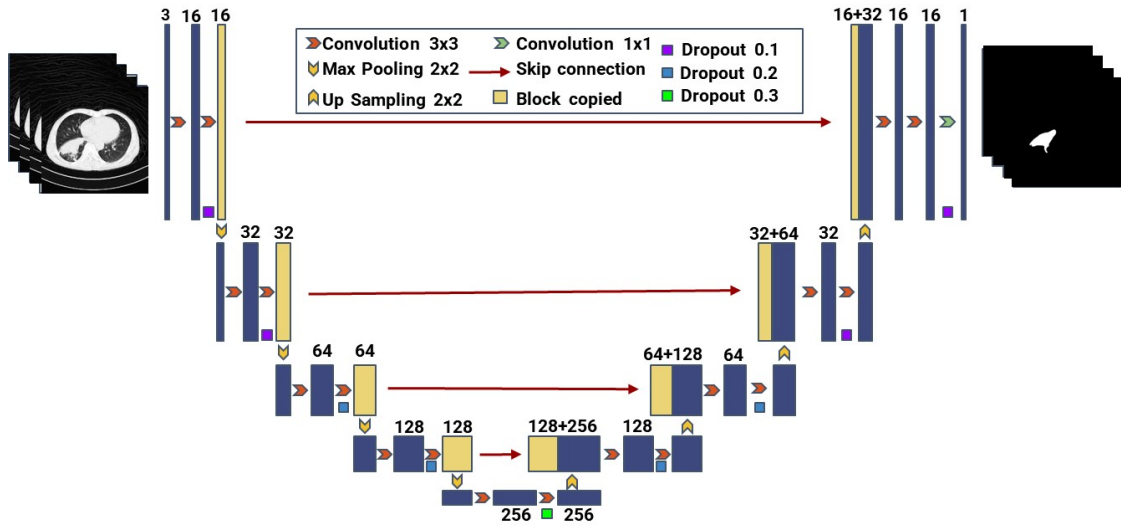


图 2-2 UNet 网络结构

三个网络的参数分别独立定义。学生模型 S_1 的参数集合记为 θ_{s_1} ，学生模型 S_2 的参数集合记为 θ_{s_2} ，两者具有完全相同的网络拓扑结构，但采用不同的随机初始化种子，确保在训练初期具有不同的特征映射空间。教师模型 T 的参数集合记为 θ_t ，其不参与反向传播的梯度计算，仅通过动量更新获取学生模型的知识。

将输入图像定义为 $X \in R^{C \times H \times W}$ ，其中 C 为输入通道数， H 和 W 为图像的高和宽。在网络的前向传播阶段，给定同一批次的无标签输入图像 X_u ， S_1 和 S_2 分别输出对应的未归一化对数概率。随后经过 Softmax 激活函数转化为像素级的类别概率分布 P_1 和 P_2 ，如公式2-1和公式2-2所示：

$$P_1 = \text{Softmax}(f(X_u; \theta_{s_1})) \quad (2-1)$$

$$P_2 = \text{Softmax}(f(X_u; \theta_{s_2})) \quad (2-2)$$

为了促使两个学生网络 S_1 和 S_2 避免陷入同质化，能够从不同的特征子空间中学习多样化的表征，本节在无标签数据上引入了基于余弦相似度（Cosine Similarity）的分歧损失函数 \mathcal{L}_{dis} 。对于图像中的每一个像素点 i ，计算两个概率向量的相似度并求均值，如公式2-3所示：

$$\mathcal{L}_{dis} = \frac{1}{N} \sum_{i=1}^N \cos(P_1^{(i)}, P_2^{(i)}) \quad (2-3)$$

其中， $N = H \times W$ ，是单张图像的像素总数， \cos 表示余弦相似度。在优化过程中，通过最小化 \mathcal{L}_{dis} ，网络被强制要求拉大 P_1 与 P_2 之间的距离，从而保持两个学生网络预测结果的差异性与互补性。

在保证模型具备差异性后，如何有效利用无标签数据进行监督是提升分割精度的关键。本章摒弃了传统的自我监督（Self-training），采用了交叉指导策略。

利用学生网络 S_1 输出的概率分布生成硬伪标签 (Hard Pseudo-labels) \hat{y}_1 , 用于指导学生网络 S_2 的学习; 反之, 利用 S_2 生成的硬伪标签 \hat{y}_2 指导 S_1 的学习。伪标签的生成过程基于 Argmax 操作, 如公式2-4和公式2-5所示:

$$\hat{y}_1^{(i)} = \arg \max_{l \in \{0,1\}} \left(P_{1,l}^{(i)} \right) \quad (2-4)$$

$$\hat{y}_2^{(i)} = \arg \max_{l \in \{0,1\}} \left(P_{2,l}^{(i)} \right) \quad (2-5)$$

为了防止低质量的伪标签造成错误梯度的反向传播, 本章使用教师模型 T 输出的确定性掩码 (Certainty Mask, 记为 $M_{ct} \in \{0,1\}^{B \times H \times W}$, 其计算过程详见 2.2.3 节) 和基于特征原型的空间特征选择掩码 (Spatial Feature Selection Mask, 记为 $M_{select} \in \{0,1\}^{B \times H \times W}$, 其计算过程详见 2.2.4 节) 共同校正产生的 M_{final} 对交叉监督信号进行严格筛选。仅对被判定为高可靠区域 (即 M_{final} 上的肿瘤目标像素点) 计算交叉监督的损失, 如2-6所示:

$$\mathcal{L}_{cross} = \frac{1}{\sum_{i=1}^N M_{final}^{(i)} + \varepsilon} \sum_{i=1}^N M_{final}^{(i)} \cdot \left(\mathcal{L}_{CE} \left(P_1^{(i)}, \hat{y}_2^{(i)} \right) + \mathcal{L}_{CE} \left(P_2^{(i)}, \hat{y}_1^{(i)} \right) \right) \quad (2-6)$$

其中, ε 为防止分母为零的平滑常数, 取值为 10^{-6} , \mathcal{L}_{CE} 为交叉熵损失函数, 如公式2-7所示:

$$\mathcal{L}_{CE}(\hat{y}_i, y_i) = -\frac{1}{N} \sum_{i=1}^N [y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)] \quad (2-7)$$

其中, \hat{y}_i 表示模型输出的预测值, y_i 表示真实的标签。

通过这种交叉监督机制, S_1 能够学习到 S_2 挖掘到的可靠特征, 反之亦然, 有效抵御了单一网络极易产生的确认偏差。

与学生模型通过反向传播算法更新参数不同, 教师模型 T 的参数 θ_t 扮演着“全局知识库”的角色。为了使教师模型能够融合两个学生模型的优势表征, 本章设计了针对双学生网络的连续指数移动平均 (Exponential Moving Average, EMA) 更新规则。

更新系数 α 并不固定, 而是随全局训练步数 (step) 动态自适应衰减。在训练初期, 赋予学生模型较大的更新权重以促使教师模型快速拟合; 在训练中后期, α 趋近于设定的基准值 (本章设定 $\alpha_{base} = 0.99$), 从而保证伪标签生成的绝对稳定性, 更新系数 α 的更新公式如2-8所示:

$$\alpha_{step} = \min \left(1 - \frac{1}{step + 1}, \alpha_{base} \right) \quad (2-8)$$

在每个迭代周期 (Iteration) 结束时, 教师模型的参数 θ_t 依次接收来自 S_1 和 S_2 的动量更新, 如公式2-9和2-10所示:

$$\theta_t \leftarrow \alpha_{step}\theta_t + (1 - \alpha_{step})\theta_{s1} \quad (2-9)$$

$$\theta_t \leftarrow \alpha_{step}\theta_t + (1 - \alpha_{step})\theta_{s2} \quad (2-10)$$

这种双重 EMA 更新策略, 确保了教师模型的权重空间由两个学生模型参数的历史集成, 从而能够为无标签数据输出更为鲁棒、平滑的概率分布特征与不确定性掩码。

2.2.3 基于不确定性估计的伪标签构建

为了过滤教师模型在组织边界或病灶模糊区域生成的错误伪标签, 本章引入了基于不确定性的像素级可靠性评估。对于每一个无标签样本, 在不关闭网络随机性的前提下, 通过教师模型 T 进行 K 次独立的前向传播, 得到一组预测概率集合 P, 如2-11所示:

$$P = P_t^{(1)}, P_t^{(2)}, \dots, P_t^{(K)} \quad (2-11)$$

其中 $P_t^{(K)} \in R^{L \times H \times W}$, L 为总类别数, 在肺部肿瘤分割问题中, $L = 2$, 包括肿瘤目标和背景两类。为了量化这 K 次预测的波动程度, 计算每个像素在类别 l 上的预测均值 μ_l 与方差 σ_l , 如公式2-12和2-13所示:

$$\mu_l = \frac{1}{K} \sum_{k=1}^K P_{i,l}^{(k)} \quad (2-12)$$

$$\sigma_l = \sqrt{\frac{1}{K} \sum_{k=1}^K (P_{i,l}^{(k)} - \mu_l)^2} \quad (2-13)$$

方差 σ_l 反映了模型对该像素属于类别 l 的预测分歧。方差越大, 说明模型在该区域的判断越不稳定。由于不同类别的基准概率存在差异, 例如, 与肺部的肿瘤区域相比, 背景占比最大, 预测结果为背景的基准概率也就偏大, 直接使用绝对方差难以建立统一的阈值标准。因此, 本章沿类别维度对方差图进行 Softmax 归一化处理, 获取相对波动响应 $\bar{\sigma}_l$, 如公式2-14所示:

$$\bar{\sigma}_l = \frac{\exp(\sigma_l)}{\sum_{j=1}^L \exp(\sigma_j)} \quad (2-14)$$

为了综合衡量所有类别的预测混乱度, 将归一化后的方差 $\bar{\sigma}_l$ 代入香农信息熵 (Shannon Entropy) 公式, 计算得到最终的标量不确定性度量图 $U \in R^{H \times W}$, 如公式2-15所示:

$$U = -\frac{1}{\ln(L)} \sum_{l=1}^L \bar{\sigma}_l \cdot \log(\bar{\sigma}_l + \varepsilon) \quad (2-15)$$

式中，引入 $\ln(L)$ （对于肺部肿瘤前景-背景二分类任务， $L=2$ ）作为分母，将不确定性值域严格映射至 $[0, 1]$ 区间，便于后续的超参数控制； $\varepsilon = 10^{-6}$ 为平滑常数，用于防止对数函数在 σ_l 趋近于零时发生梯度溢出。

通过预设的不确定性容忍阈值 τ ，对 U 进行二值化截断，生成交叉指导中使用的确定性掩码矩阵 $M_{ct} \in \{0, 1\}^{H \times W}$ ，用于区分高可信区域与高不确定区域，如公式2-16所示：

$$M_{ct}^{(i)} = \begin{cases} 1, & U^{(i)} < \tau \\ 0, & U^{(i)} \geq \tau \end{cases} \quad (2-16)$$

其中， τ 为不确定性阈值，本章设定为 0.1。

当且仅当像素预测的不确定性低于该阈值时，该像素才被认为是可信的，并允许参与后续伪标签生成与监督过程。当某像素点的不确定性 $U \geq \tau$ 时，对应的掩码值设置为 0，该位置的伪标签被判定为“不可靠”，其产生的误差梯度将被强制屏蔽。这一基于不确定性过滤的校正机制，赋予了网络在模糊组织边界处自动摒弃低质量监督信号的能力，可以有效过滤掉预测不稳定的像素区域，避免将高噪声伪标签引入训练过程，从而显著提升了双学生模型在无标签数据上的学习鲁棒性。

2.2.4 基于特征原型的自适应语义一致性掩码校正

尽管基于不确定性的过滤机制能够有效剔除高信息熵的噪声标签，但 Softmax 层输出的概率分布本质上侧重于局部像素的响应，往往缺乏对全局语义结构和上下文特征的宏观感知。为了进一步从深度特征空间挖掘像素间的语义一致性，本章提出了一种基于特征原型（Feature Prototype）的自适应掩码生成与二次校正策略。

在深度度量学习的视角下，同一类别的像素在经过深层网络映射后，应在高维特征空间中呈现出高度的聚类特性。本章首先提取教师网络 T 在解码器之前的深层特征图 $F \in R^{B \times C' \times H' \times W'}$ ，其中 B 为批次大小（Batch Size）， C' 为特征通道维度， H' 和 W' 为空间分辨率。为了便于像素级计算，将空间与批次维度展平，特征图可表示为包含 N 个特征向量的集合，如式2-17所示：

$$f_i \in R^{C' \times N} \quad (i = 1, 2, \dots, N; N = B \times H' \times W') \quad (2-17)$$

为了计算具有代表性的类别聚类中心（即特征原型），本章利用 2.2.3 节生成的初始高置信度伪标签掩码 $M_{ct}^{(l)} \in \{0, 1\}^N$ （类别 $l \in \{fg, bg\}$ ， fg 代表前景肿瘤目标， bg 代表背景）以及对应的最大类别概率置信度图 $W^{(l)} \in [0, 1]^N$ 作为空

间权重。全局类别特征原型 Z_l 的计算公式如2-18所示：

$$Z_l = \frac{\sum_{i=1}^N M_{ct}^{(i,l)} \cdot W^{(i,l)} \cdot f_i}{\sum_{i=1}^N M_{ct}^{(i,l)} \cdot W^{(i,l)} + \varepsilon} \quad (2-18)$$

式中， $\varepsilon = 10^{-6}$ 为防止分母为零的平滑常数。

通过引入概率置信度 $W_i^{(l)}$ 作为软权重（Soft Weight），类别特征原型的计算被强制倾向于特征空间中最确定、最典型的像素点，从而避免了模糊边界像素对全局聚类中心的稀释与污染。

获取当前 Batch 的肿瘤目标原型 Z_{fg} 与背景原型 Z_{bg} 后，它们分别代表了当前高维空间中两类语义特征的聚类中心。接着，本章计算特征图上每一个像素的深层特征向量 f_i 到这两个原型的欧式距离平方（Squared Euclidean Distance），如公式2-19和2-20所示：

$$D_{fg}^{(i)} = \|f_i - Z_{fg}\|^2 \quad (2-19)$$

$$D_{bg}^{(i)} = \|f_i - Z_{bg}\|^2 \quad (2-20)$$

基于特征空间中的最近邻原则，如果一个像素的特征向量距离肿瘤目标原型更近，则其在语义层面上应更倾向于属于前景肿瘤目标，反之则属于背景，由此能够生成一个完全独立于分类器概率输出的空间特征选择掩码矩阵（Spatial Feature Selection Mask） $M_{select} \in \{0, 1\}^N$ ，如公式2-21所示：

$$M_{select}^{(i)} = \begin{cases} 1, & D_{fg}^{(i)} < D_{bg}^{(i)} \\ 0, & D_{fg}^{(i)} \geq D_{bg}^{(i)} \end{cases} \quad (2-21)$$

利用 M_{select} 对原本基于局部概率生成的伪标签 \hat{y} 进行二次空间约束与一致性校验。只有当分类器输出的概率伪标签类别与高维特征空间划分的类别高度吻合时，该像素才被认为是绝对安全的监督信号。本章通过计算逻辑与（Logical AND）操作，生成最终的联合校正掩码 M_{final} ，如公式2-22所示：

$$M_{final}^{(i)} = M_{ct}^{(i)} \wedge M_{select}^{(i)} \quad (2-22)$$

通过这种概率分布和特征空间的双重门控机制，网络不仅在局部过滤了高不确定性的噪声，还在全局范围内剔除了游离于主流特征簇之外的离群点（Outliers），为学生网络的交叉指导提供了极高纯度的伪标签。

2.2.5 损失函数的设计

本算法的最终训练目标由监督损失 \mathcal{L}_{sup} 和无监督损失 \mathcal{L}_{unsup} 联合驱动。

(1) 监督损失 \mathcal{L}_{sup}

在有标签数据流中,采用结合类别权重 (Intra-weights) 的交叉熵损失与 Dice 损失的组合形式 (即 DSCLoss) 如公式2-23所示:

$$\mathcal{L}_{sup} = \frac{1}{2} \sum_{j=1}^2 ((1 - w_{inter}) \mathcal{L}_{CE}(P_j, Y) + w_{inter} \mathcal{L}_{Dice}(P_j, Y)) \quad (2-23)$$

其中, P_j 为学生模型 S_j 的预测概率分布, Y 为对应的真实标签, w_{inter} 为类别权重, 本章设置为 0.5。

(2) 无监督损失 \mathcal{L}_{unsup}

无监督损失包括三部分: 一致性损失 \mathcal{L}_{con} , 学生交叉监督损失 \mathcal{L}_{cross} 、双学生网络之间的分歧损失 \mathcal{L}_{dis} 。

一致性损失 \mathcal{L}_{con} 采用均方误差约束实现。对输入图像施加高斯噪声扰动后, 计算学生模型与教师模型在特征距离判定为背景的掩码区域内输出的均方误差, 以约束模型在扰动下保持预测一致性, 提升半监督训练的稳定性。如公式2-24所示:

$$\mathcal{L}_{con} = \frac{1}{|M_{bg}|} \sum_{k=1}^2 \sum_{i \in M_{bg}} \left\| \text{Softmax}(P_{stu_k, i}) - \text{Softmax}(P_{tea, i}^{noisy}) \right\|^2 \quad (2-24)$$

其中, M_{bg} 表示背景掩码区域, $|M_{bg}|$ 两个学生网络对应的背景掩码有效像素总数的和, $P_{stu_k, i}$ 表示第 k 个学生模型在像素 i 处的预测输出, $P_{tea, i}^{noisy}$ EMA 教师模型经噪声增强后的预测输出, $\text{Softmax}(\cdot)$ 表示类别概率归一化函数, $\|\cdot\|^2$ 表示均方误差。

由于肿瘤前景区域在 CT 序列中具有较强的形态多边性, 因此仅在结构稳定的背景区域施加一致性约束, 以保证模型既能稳定学习通用特征, 又不限制对肿瘤动态变化的建模能力。

学生交叉监督损失 \mathcal{L}_{cross} 和双学生网络之间的分歧损失 \mathcal{L}_{dis} 分别见公式2-6和公式2-3。

(3) 总体目标函数

整合上述所有机制, 网络整体优化的多维损失函数定义如下:

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda(step) \cdot \mathcal{L}_{con} + \lambda_{cross} \mathcal{L}_{cross} + \lambda_{dis} \mathcal{L}_{dis} \quad (2-25)$$

其中, $\lambda(step)$ 是一致性损失的权重, $step$ 表示当前训练迭代步数, $maxstep$ 表示最大训练迭代步数, 最终权重 λ_{max} 为 0.1, 采用 Sigmoid 型的动态 Ramp-up 衰减策略, 控制方式如公式2-26所示, 使得模型在训练初期主要依赖有标签数据, 在具备一定特征提取能力后逐步增强对无标签数据的探索; λ_{cross} 和 λ_{dis} 分别设定为 0.1 和 0.001。

$$\lambda(step) = \lambda_{max} \cdot e^{-5\left(1 - \frac{step}{maxstep}\right)^2} \quad (2-26)$$

2.3 实验结果与分析

2.3.1 Xiangya-LIM 数据集

本研究所使用的临床影像数据集源自湖南省湘雅医院放射科，数据集涵盖了334例确诊为肺癌并接受免疫治疗的患者的胸部CT扫描序列及其对应的标注信息，命名为Xiangya-LIM (Xiangya Lung cancer IMmunotherapy)。原始数据以DICOM格式存储，CT图像尺寸为 512×512 ，包含完整的元数据，如患者信息、扫描参数、图像序列号等。

利用pydicom库读取每个DICOM文件，提取像素数据、层厚、像素间距、扫描方位等元数据。将像素值转换为标准的亨氏单位(Hounsfield Unit, HU)，以确保不同设备、不同扫描参数下的CT值具有统一的物理意义，为后续的灰度标准化奠定基础。

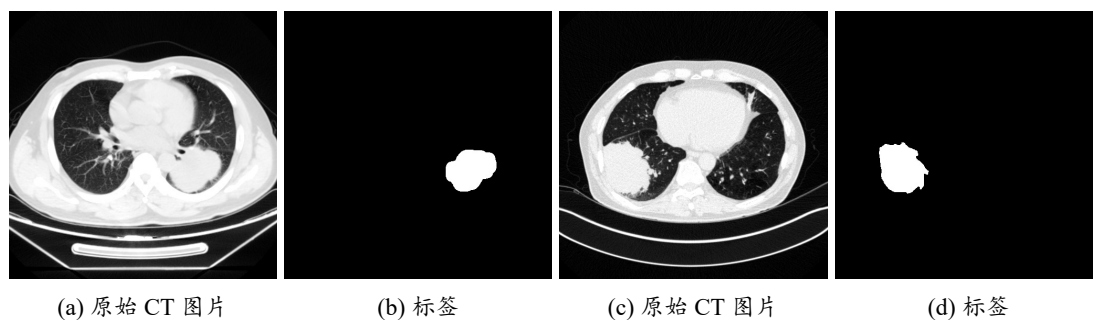


图 2-3 Xiangya-LIM 数据集示例

由于不同CT扫描协议的层厚和像素间距存在差异，如果直接将数据输入网络，容易导致空间尺度不一致的问题。因此，采用三线性插值方法将所有样本重采样，使其具有统一的空间分辨率。本研究设定目标体素间距为 $1.0 \times 1.0 \times 1.0 \text{ mm}^3$ ，该分辨率在保留解剖细节的同时兼顾了计算效率。重采样过程中，图像数据采用三线性插值，而标注掩膜则采用最近邻插值以避免标签值的非整数化。

为增强肺实质与结节的对比度，采用基于窗宽窗位的强度截断与归一化策略。设定肺窗的窗宽为1500 HU、窗位为-600 HU，将HU值进行截断，并将有效灰度线性映射至 $[0, 255]$ 范围，抑制骨骼、金属植入物等高亮结构的干扰，突出肺组织内的密度变化，有利于模型对肺部病灶的识别。

每位患者的CT序列通常包含大量切片，在进行分割任务时，仅保留包含肿瘤区域的切片，以减少无关信息对模型训练的干扰。

经过上述流程，本研究共构建了包含7099张高质量CT影像的数据集，为后续深度学习模型的训练与验证奠定了数据基础。数据集的示例如图2-3所示。

以患者为单位对数据集进行划分，将全部患者数据按照8:2的比例划分为

训练集和测试集，以确保同一患者的影像数据不会同时出现在训练集和测试集中，从而避免数据泄漏并提高模型评估结果的可靠性，如表2-1所示：

表 2-1 数据集划分情况

数据集	患者数	CT 切片数量
训练集	267	5740
测试集	67	1359
总计	334	7099

2.3.2 评价指标

为了客观、全面地评估所提出的半监督分割模型在肺癌病灶区域的分割精度，本章采用了医学影像分割领域公认的三项核心评价指标：Dice 系数（Dice Similarity Coefficient）、Jaccard 系数（Jaccard Index）以及 95% 豪斯多夫距离（95% Hausdorff Distance, HD95）。

(1) Dice 系数 (Dice Similarity Coefficient)

Dice 系数是衡量预测分割结果与人工标注金标准（Ground Truth）之间重叠程度的最常用指标。其值域为 [0,1]，数值越接近 1，表示模型分割结果与专家标注越吻合。其数学定义如下：

$$\text{Dice} = \frac{2|P \cap G|}{|P| + |G|} \quad (2-27)$$

其中， P 代表模型生成的预测分割二值图， G 代表由放射科医生手动标注的金标准， $|\cdot|$ 表示集合中像素点的数量。

(2) Jaccard 系数 (Jaccard Index)

Jaccard 系数又称为交并比（Intersection over Union, IoU），用于计算两个集合交集与并集的比值。它与 Dice 系数呈正相关，但对分割错误的惩罚比 Dice 更为严苛。其公式表示为：

$$\text{Jaccard} = \frac{|P \cap G|}{|P \cup G|} = \frac{|P \cap G|}{|P| + |G| - |P \cap G|} \quad (2-28)$$

在分析中，结合使用 Dice 和 IoU 可以更稳健地反映模型对病灶边界的捕捉能力。

(3) 95% 豪斯多夫距离 (95% Hausdorff Distance)

上述两项指标主要关注重叠面积，而豪斯多夫距离（HD）则侧重于评估两个集合边界之间的空间距离，对解剖结构的形状相似性非常敏感。原始的豪斯多

夫距离定义为两个集合边界点之间最大距离的最小值：

$$HD(P, G) = \max\{\sup_{p \in P} \inf_{g \in G} d(p, g), \sup_{g \in G} \inf_{p \in P} d(g, p)\} \quad (2-29)$$

由于原始 HD 对离群噪声点极其敏感，在医学影像评价中通常采用 95% 豪斯多夫距离（HD95）。该指标在计算时排除前 5% 的最大距离，取剩余距离中的最大值，从而能更稳定地反映分割边界与真实边界的紧密程度。HD95 的单位通常为毫米（mm），其数值越小，代表分割结果的轮廓越接近真实病灶。

2.3.3 实验设置

在实验运行环境方面，硬件环境和软件框架如表2-2所示。

表 2-2 实验环境配置

环境配置	环境参数
操作系统	Ubtunu 20.04.6
CPU	Intel(R)Xeon(R)Platinum 8336c cPu @ 2.30
GPU	NVIDIA GeForce RTX 4090
Python 环境	Python 3.10
深度学习环境	Pytorch 2.1.1

表 2-3 超参数设置

超参数名称	具体设置
教师网络更新系数 α	0.99
优化器	SGD
初始学习率 lr	0.001
权重衰减 Weight Decay	0.0001
最大迭代次数	40000
Batch Size	8
随机前向传播次数 K	10
不确定性过滤阈值 τ	0.1
学生间分歧损失权重 λ_{dis}	0.001
双学生交叉监督损失权重 λ_{cross}	0.1

在超参数设置方面，教师网络 T 的参数通过学生网络 S_1 和 S_2 的权重以指数移动平均（EMA）方式更新，更新系数 α 设定为 0.99；优化器使用了 SGD 优化器，初始学习率设定为 0.01，动量（Momentum）为 0.9，权重衰减（Weight Decay）

为 0.0001，学习率采用多项式衰减策略。训练的最大迭代次数设定为 40000 次，Batch Size 设定为 8（其中 4 个有标签样本，4 个无标签样本）。随机前向传播次数 K 设定为 10，不确定性过滤阈值 τ 设置为 0.1，学生间分歧损失权重 λ_{dis} 设置为 0.001， λ_{cross} 设定为 0.1，数值稳定性平滑项 ε 设置为 10^{-6} ，如表2-3所示。

2.3.4 对比实验

为了验证所提出方法的有效性，本章选取了当前医学图像半监督分割领域具有代表性的算法作为对比方法，包括：MT^[61]（Mean Teacher）、UAMT^[67]（Uncertainty-aware Mean Teacher）、DTC^[68]（Dual Task Consistency）、BCP^[69]（Bidirectional Copy-Paste）、CML^[70]（Cross Model Learning）、DyCON^[71]（Dynamic Consistency）。所有对比方法均在相同数据集、相同网络结构（UNet）下进行训练，以确保实验公平性。不同标注比例（50% 和 20%）下的对比实验结果如表2-4所示。

表 2-4 不同模型在 LIDC-IDRI 数据集上的性能对比

Model	有标签比例	无标签比例	Dice(%) \uparrow	Jaccard(%) \uparrow	HD95(mm) \downarrow
UNet ^[46]	100%	0%	84.80	73.61	4.31
MT ^[61]	50%	50%	78.14	65.45	11.94
UAMT ^[67]	50%	50%	81.06	68.89	7.59
DTC ^[68]	50%	50%	82.28	70.49	6.88
BCP ^[69]	50%	50%	81.54	69.62	7.25
CML ^[70]	50%	50%	83.18	71.74	5.13
DyCON ^[71]	50%	50%	79.96	67.58	9.21
Ours	50%	50%	83.66	72.45	4.76
MT ^[61]	20%	80%	72.01	59.26	15.47
UAMT ^[67]	20%	80%	76.39	63.67	11.28
DTC ^[68]	20%	80%	79.33	66.74	8.65
BCP ^[69]	20%	80%	77.72	65.45	9.94
CML ^[70]	20%	80%	80.55	68.25	7.88
DyCON ^[71]	20%	80%	75.38	62.61	13.23
Ours	20%	80%	81.83	70.03	6.22

如表2-4所示，本章提出的方法在 Dice 系数、Jaccard 系数和 HD95 指标上均优于现有半监督分割方法。随着标注数据比例的降低，所有方法的性能均出现不同程度的下滑。实验结果表明，在 50% 标注场景下，本章方法表现最优，Dice 系数达到 83.66%，仅比 100% 全监督的 UNet（84.80%）低 1.14%；Jaccard 系数达

到 72.45%，仅比全监督 UNet (73.61%) 低 1.15%。这证明了本框架在仅需一半标注成本的情况下，能够通过无标签数据的有效利用，达到逼近全监督性能的水平。当标签比例骤降至 20% 时，一些传统半监督方法（如 MT）性能严重退化，Dice 跌破 77%。而本章方法依然维持在 81.83%，且 HD95 指标 (6.22mm) 远优于其他对比方法（如 DyCON 的 13.23mm）。这凸显了本算法在缺乏专家标注时的鲁棒性。

在肺部影像分割对比实验中，本章提出的基于双学生交叉指导与不确定性校正的分割框架 (Ours) 与近年来主流的半监督框架进行了系统性评价，这些方法均采用 UNet 作为基础网络。表 2-4 的定量结果揭示了不同半监督框架在处理肺部肿瘤分割中的标签稀缺问题时的性能表现。

从整体趋势观察，所有半监督方法在 Dice、Jaccard 及 HD95 指标上均展现出随标注比例降低而性能衰减的特性，这符合数据驱动深度学习模型的客观规律。本章方法在 50% 标注比例下的 Dice 系数达到了 83.66%，显著优于经典的 Mean Teacher (78.14%) 及最新的 DyCON (79.96%)，且与使用 100% 全部标签的全监督 UNet (84.80%) 相比，性能差距已缩小至 1.14%。这一结果初步验证了本章方法通过双学生架构对无标签数据进行深度挖掘，能够有效缓解标注压力并逼近全监督的分割精度。

针对 50% 标注比例的结果进行分析，本章所提方法通过引入不确定性估计与类别原型提取机制，在分割精度上显著超越了 CML 等现有先进半监督分割方法。在实现过程中，通过对 EMA 教师模型执行多次随机推断并计算预测方差以构建不确定性图谱，是提升伪标签质量的关键环节。传统 Mean Teacher 框架因缺少有效的反馈与校正机制，教师模型在迭代过程中易出现确认偏差，导致伪标签噪声累积。为此，本章方法借助不确定性掩码筛选高置信度预测区域，并在特征空间中分别学习目标肿瘤原型与背景原型，通过特征距离度量自适应生成选择掩码。该机制使模型不再单纯依赖概率平滑约束，而是学习到更具判别性与鲁棒性的特征表示，在应对肺部复杂组织结构与模糊边界时表现出更强的分割能力。最终模型在对应设置下 Jaccard 系数达到 72.45%，充分验证了预测分割结果与真实标注之间具有较高的空间一致性。

在更具挑战性的 20% 极低标注比例场景下，本章方法的优势得到了进一步放大。当训练样本从 50% 锐减至 20% 时，诸如 MT 和 DyCON 等方法出现了明显的性能塌陷，Dice 值分别跌落至 72.01% 和 75.38%，且 HD95 指标急剧恶化，说明模型在极少量监督信号下难以通过单一的一致性约束维持目标的拓扑结构。相比之下，本章方法依然保持了 81.83% 的 Dice 分数，甚至优于部分对比方法在 50% 标签下的表现。这归功于双学生交叉监督策略与伪标签校正的协同作用。交

叉监督策略通过引入两个具有不同初始化参数的学生模型（Student 1 & Student 2），并促使二者相互学习对方生成的高质量伪标签，从而形成协同优化的训练机制。该策略能够有效提升模型对数据噪声的鲁棒性，同时缓解单模型在训练过程中容易陷入局部最优解的问题。

从衡量边界解剖结构准确性的 95% 豪斯多夫距离（HD95）来看，本章方法展现了较好的边界感知能力。在肺癌分割任务中，肿瘤的浸润性边缘往往与周围组织界限模糊，HD95 的降低直接反映了算法对边缘区域预测的平稳性。在 20% 标注下，本章方法的 HD95 为 6.22mm，远低于 MT 的 15.47mm 及 BCP 的 9.94mm。本章方法利用背景掩膜动态调整一致性权重，并结合带有噪声扰动的输入进行约束，这种精细化的扰动一致性学习使得模型生成的分割边缘更加连续且贴合实际解剖轮廓。

图2-4提供了本章方法与其他方法在 Xiangya-LIM 数据集上的可视化结果对比。第一列是原始的 CT 图像，红色区域表示肿瘤区域，第二列为专业医生人工标注的金标准，在所有方法的分割结果图中均用红色轮廓标注了金标准的肿瘤区域，方便对比。如图2-4所示，无论是处理小目标或具有复杂边界的分割任务，本章方法的分割结果均与真实标注更加接近，其在目标边界识别和精确提取方面明显优于其他半监督学习方法。这一优势源于本章的方法能有效利用无标注数据。相比之下，其余方法可能未能充分提取无标注数据中的有效信息，导致在处理肺部肿瘤的边界模糊及特征信息不足时表现不佳。

2.3.5 消融实验

为了验证本文提出的单教师双学生架构（STDS）、基于不确定性的掩码过滤（ M_{ct} ）以及基于特征原型的语义一致性校正（ M_{select} ）的有效性，本章在自建数据集 Xiangya-LIM 上进行了全因子消融实验。实验统一采用 20% 的有标签数据作为监督信号，其余 80% 为无标签数据。实验结果如表2-5所示。

相比于基础 Mean Teacher（MT）框架，引入 STDS 架构后，Dice 提升了约 3.71%。这表明传统的单学生模型极易与教师模型产生预测同质化，进而陷入“确认偏差”。而双学生架构通过交叉指导（Cross-Teaching）和分歧损失，强制两个网络在不同的特征子空间中进行探索。这种机制增强了模型对未标注样本的表征多样性，从而在训练初期就建立了更鲁棒的基准。

与 Exp 1 相比，Exp 3 的 HD95 指标从 15.47mm 下降至 11.19mm，Dice 系数上升了 4.28%。这说明对模型预测波动性的捕捉，能够在一定程度上约束模型识别出教师网络在组织边界、病灶阴影等低对比度区域的争议状态。通过屏蔽这些具有高度争议的伪标签，网络避免了由模糊边界引发的错误梯度回传。

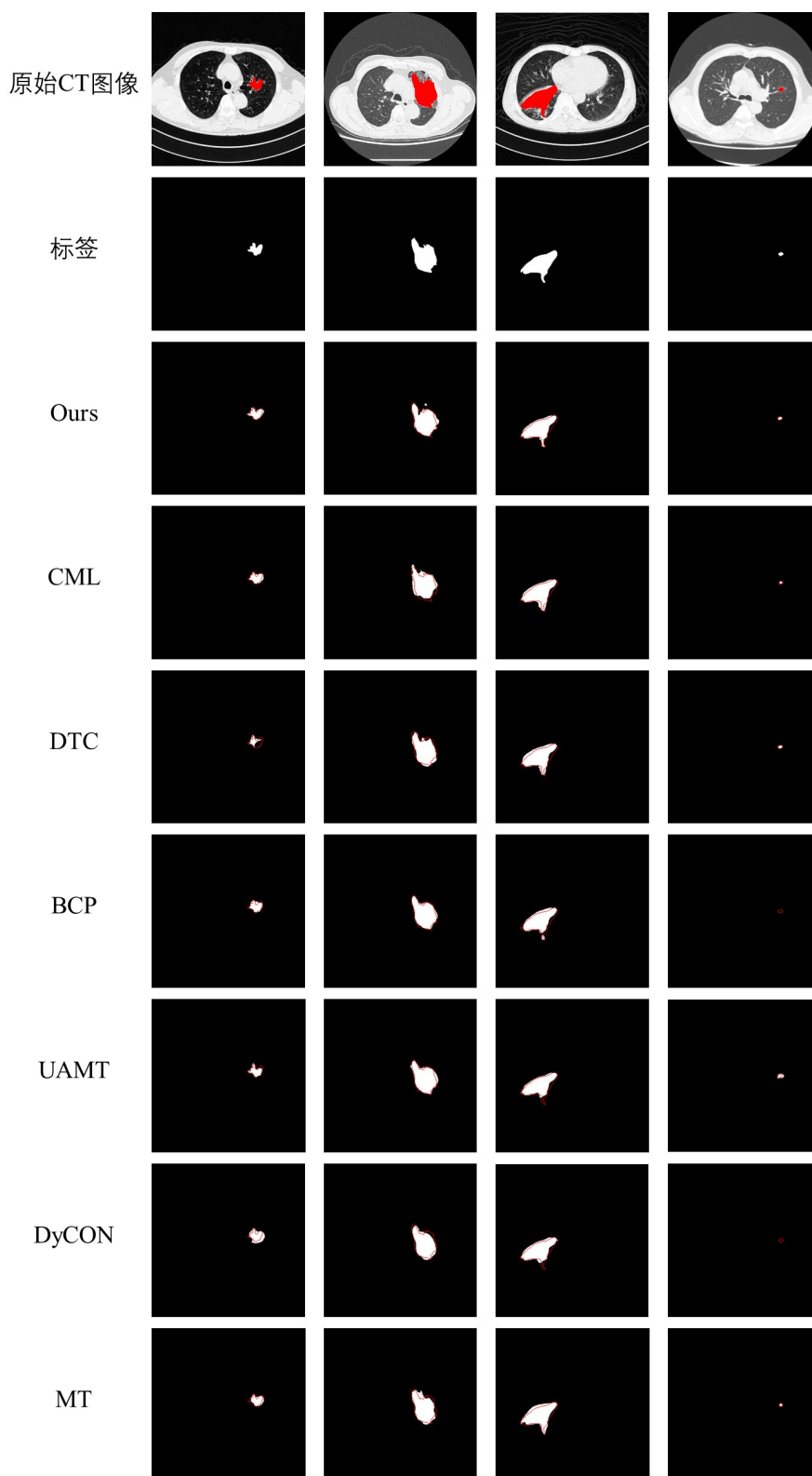


图 2-4 不同半监督方法在 Xiangya-LIM 数据集上的可视化示例

为了验证特征原型校正的作用，直接利用教师模型输出的原始伪标签作为掩码，计算所有像素的特征均值作为原型。如表2-5所示，Exp 4 同样取得了增益。

表 2-5 20% 标注下各个模块的消融实验结果

Exp.	基础框架 MT	STDS	M_{ct}	M_{select}	Dice(%)↑	Jaccard(%)↑	HD95(mm)↓
Exp 1	✓	×	×	×	72.01	59.26	15.47
Exp 2	✓	✓	×	×	75.72	63.23	11.99
Exp 3	✓	×	✓	×	76.92	63.91	11.19
Exp 4	✓	×	×	✓	78.87	66.45	10.03
Exp 5	✓	✓	✓	×	79.28	66.95	10.80
Exp 6	✓	✓	×	✓	79.59	66.98	9.95
Exp 7	✓	×	✓	✓	80.61	68.31	7.91
Exp 8	✓	✓	✓	✓	81.83	70.03	6.22

这说明即便在不依赖 M_{ct} 的情况下，直接从特征空间计算聚类中心仍能纠正部分分类器的逻辑错误，表明深层特征空间具备比分类概率层更强的语义一致性。

Exp 7 结合了不确定性过滤 M_{ct} 与特征原型校正 M_{select} ，其性能优于任何单一掩码模块，相比与 Exp 3，Dice 系数提升了 3.69%，相比 Exp 4，Dice 系数提升了 1.74%。由于 M_{ct} 先行过滤了边缘离群像素，使得计算出的特征原型更加趋向于类别的“几何中心”，从而生成了更高纯度的 M_{select} 。这种由局部概率到全局语义的级联校正策略，实现了对伪标签的深度净化。

当所有模块共同作用时，如 Exp 8 所示，模型达到了最优性能（Dice: 81.83%，HD95: 6.22mm）。相比 Baseline，HD95 降低了近 60%。这充分证明了：STDS 提供了高质量的模型分歧； M_{ct} 提供了像素级的动态过滤； M_{select} 提供了全局语义的空间约束。三者形成了一个从架构设计到数据净化的完整闭环，显著提升了半监督框架在肺癌病灶分割上的可靠性。

为直观验证不确定性掩码过滤与特征原型语义校正两种优化策略的实际作用，本章选取多组临床病例开展可视化对比分析，结果如图 2-5 所示。图中第一列为肺部原始 CT 影像，红色区域表示肿瘤区域，第二列为对应真实分割标签，第三列与第四列依次为不确定性掩码过滤热力图、特征原型伪标签校正可视化结果。

由第三列可视化结果可见，模型对肿瘤病灶区域预测不确定性整体偏高，其中肿瘤边缘交界地带不确定性最为突出；肺部的正常背景区域不确定性水平较低。该分布特征与医学影像分割领域普遍存在的不确定性变化规律契合，证实本章所采用的不确定性估计方法具备合理性与有效性。如第四列所示，经过特征原型完成语义一致性校正后，模型特征聚焦区域与真实病灶标签匹配度显著提升，

由此表明基于特征原型开展伪标签二次校正能够有效修正初始掩码偏差，进一步优化分割表征效果。

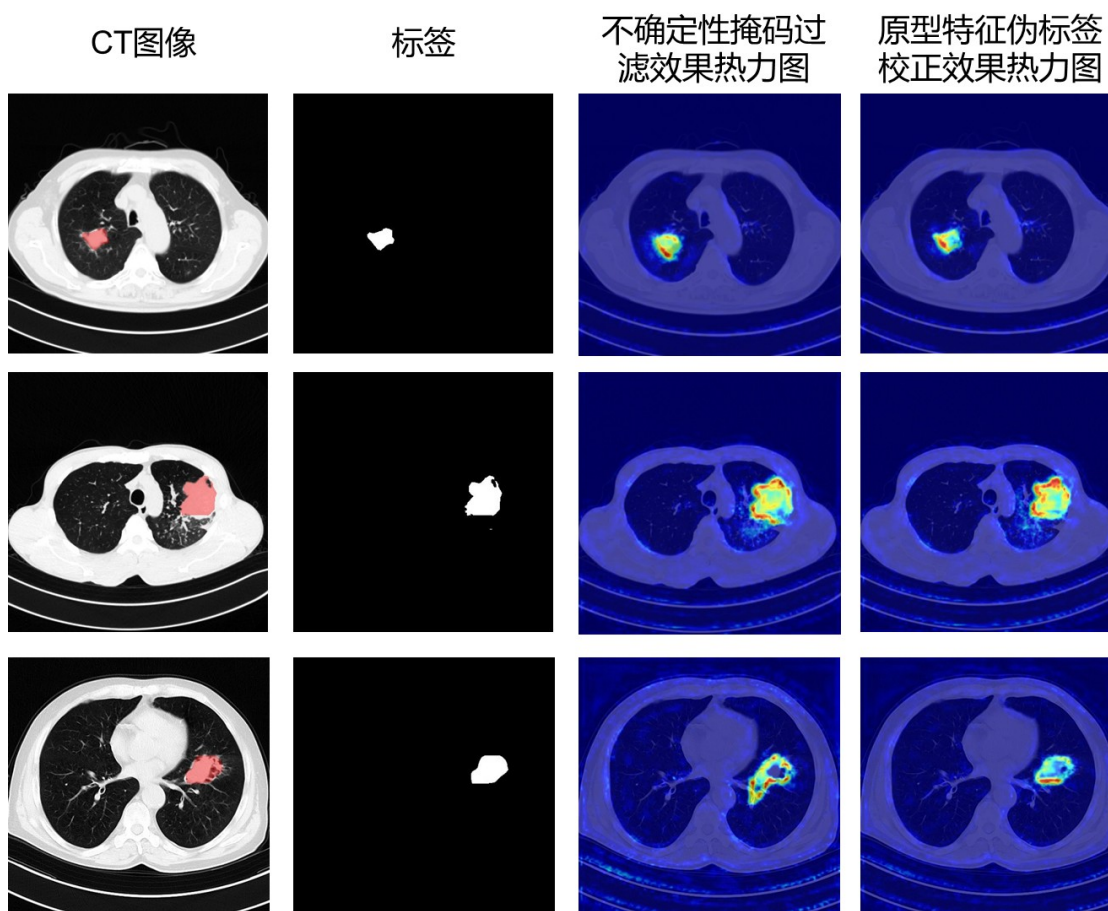


图 2-5 不确定性掩码过滤与特征原型语义校正可视化对比

2.4 本章小结

本章针对肺部肿瘤分割任务中标注成本高导致的“有标签数据稀缺”问题，提出了一种基于双学生交叉指导与不确定性校正的半监督分割框架。为了打破单一学生网络在无标签数据上容易陷入预测同质化的困境，设计了双学生并行结构，通过引入分歧损失和交叉指导机制，迫使两个学生网络在不同的特征空间进行探索，增加模型对未标注样本的表征多样性。针对病灶边界模糊导致的噪声干扰，通过多次随机推断捕捉像素级的预测波动性，构建不确定性掩码 M_{ct} ，并进一步提取特征空间的全局类别原型，利用语义一致性约束构建特征选择掩码 M_{select} ，实现对伪标签从“概率分布”到“高维特征”的校正。在 Xiangya-LIM 数据集上的对比实验表明，本章方法相较于现有主流算法在多项评价指标上均取得领先。同时，消融实验进一步验证了所提架构中各功能模块对提升分割性能的独立贡献及其协同效应。

第3章 基于最大帧多维度特征提取的肺癌免疫治疗疗效预测模型

3.1 引言

在肺癌患者的 CT 影像中，肿瘤区域面积最大的轴位层面通常包含最具代表性的形态、密度及纹理信息。已经有研究证明了基于肺部肿瘤最大帧 CT 图像提取特征能够在有限的条件下快速构建出高可靠性的预测模型^[72, 73]。但是仍然存在以下问题：一是肺部肿瘤尺度跨度大的问题，二是肿瘤的边缘结构和内部的纹理难以统一建模的问题，三是肺实质、血管等背景会带来噪声干扰的问题。

本章围绕肺癌免疫治疗的疗效预测问题展开研究，在第二章肺部肿瘤分割方法的基础上选出肺癌患者 CT 序列中的肿瘤区域最大帧最为输入，构建了 MaxFrameNet 模型。针对肺部肿瘤尺度跨度大的问题，设计了基于多膨胀率卷积的多尺度特征提取结构，并引入轻量化门控模块实现尺度权重的动态分配；面向肿瘤边缘结构与内部纹理异质性难以统一建模的问题，提出了空间域与频率域协同建模策略；对于肺实质背景及炎症等噪声干扰问题，构建了融合全局语义建模与空间一致性约束的噪声抑制机制。该模型适用于临床的辅助评估，在无创条件下为影像医生提供初步的免疫治疗疗效预测提示，尤其适用于基层医院或资源有限环境中肺癌免疫治疗潜在获益人群的初筛。

3.2 算法模型

3.2.1 网络整体结构

本章所提出的 MaxFrameNet 以 ResNet34^[74] 为基础骨干网络，创新性地提出了自适应尺度感知门控模块 (Adaptive Scale-Aware Gating, ASAG)、空频双域特征增强模块 (Spatial-Frequency dual-domain feature Enhancement Module, SFEM)、全局语义引导的噪声抑制模块 (Global Semantic-guided Noise Refiner, GSNR) 来增强模型特征提取的能力，提升肺癌免疫治疗疗效预测的准确性，如图3-1所示。

ResNet34 骨干网络包含四个阶段，对应图 3-1 中的 ResNet Layer1 至 ResNet Layer4，网络每经过一个阶段，特征图的通道数逐步增加，空间维度逐步降低，实现特征的逐步抽象与精炼，浅层网络聚焦于提取肿瘤的边缘、纹理等基础局部特征，深层网络则聚焦于提取肿瘤的高级语义特征。第一阶段输出 64 通道特征图，第二阶段输出 128 通道特征图，第三阶段输出 256 通道特征图，第四阶段输出 512 通道特征图，各阶段通过步长为 2 的卷积层实现空间维度的下采样，同时通过 1×1 卷积层实现通道数的调整，确保残差连接的维度匹配。每个阶段都由若

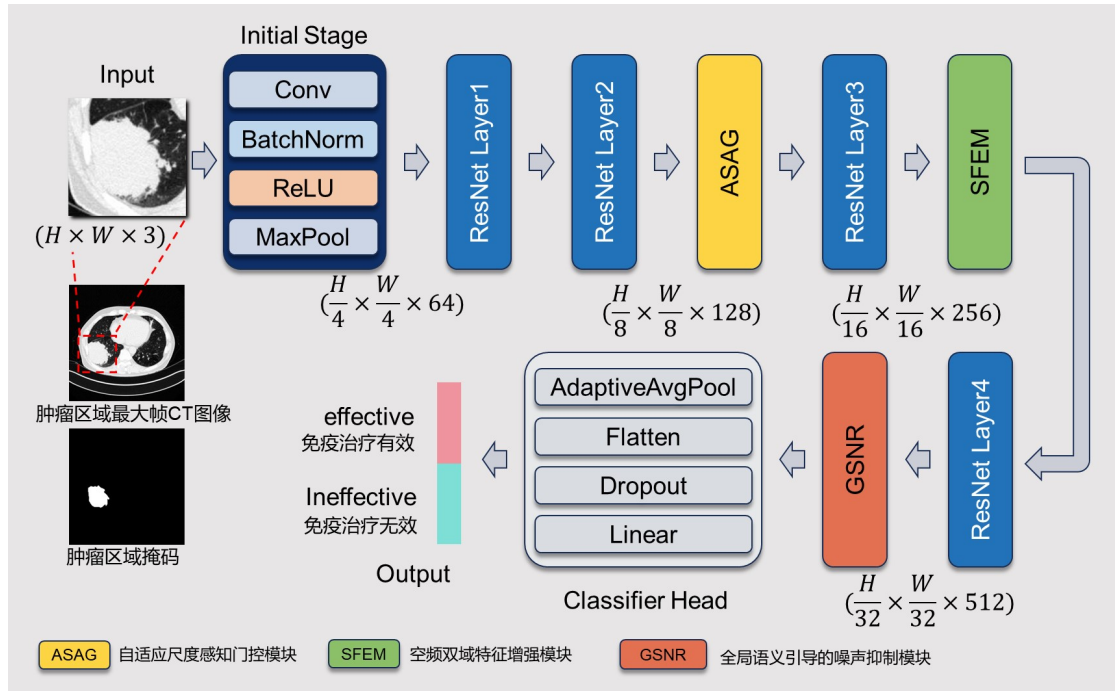


图 3-1 MaxFrameNet 网络结构图

干个残差块组成，这四个阶段包含的残差块数量分别为 [3, 4, 6, 3]。每个残差块的计算过程可通过公式3-1表示：

$$h = \delta (Conv_2 (\delta (Conv_1 (x)))) + x \quad (3-1)$$

其中， x 表示残差块的输入特征张量， $Conv_1$ 和 $Conv_2$ 分别表示残差块中的两次卷积操作（均采用 3×3 卷积核，填充值设置为 1 以保证特征图尺寸不变）， δ 表示 GELU 激活函数，兼顾梯度稳定性与计算效率， h 表示残差块的输出特征张量。

3.2.2 自适应尺度感知门控模块

在 CT 影像中，肺部的肿瘤通常呈现出显著的尺寸异质性，从数毫米的微小病灶到数厘米的大结节，其尺度跨度较大。传统卷积神经网络采用固定感受野的卷积核，难以同时有效捕捉不同尺度的肺部肿瘤的特征信息。在免疫治疗疗效预测中，肿瘤的大小和形态特征可能与其生物学行为相关^[24]。

针对免疫治疗疗效预测任务中肺部肿瘤尺寸跨度大，微小肿瘤至较大肿瘤并存、常规的卷积感受野固定，无法动态适配不同尺度目标的问题，本章创新性地设计了自适应尺度感知门控模块（Adaptive Scale-Aware Gating, ASAG）如图3-2所示。这个模块通过多膨胀率卷积来构建多尺度感受野，覆盖肺部肿瘤的典型尺度范围；通过门控网络学习不同尺度特征的重要性权重，实现多尺度特征的动态选择与融合；同时通过残差连接保留原始特征信息，避免尺度自适应过程

中的特征丢失，保证特征流的完整性，提升模块的训练稳定性。

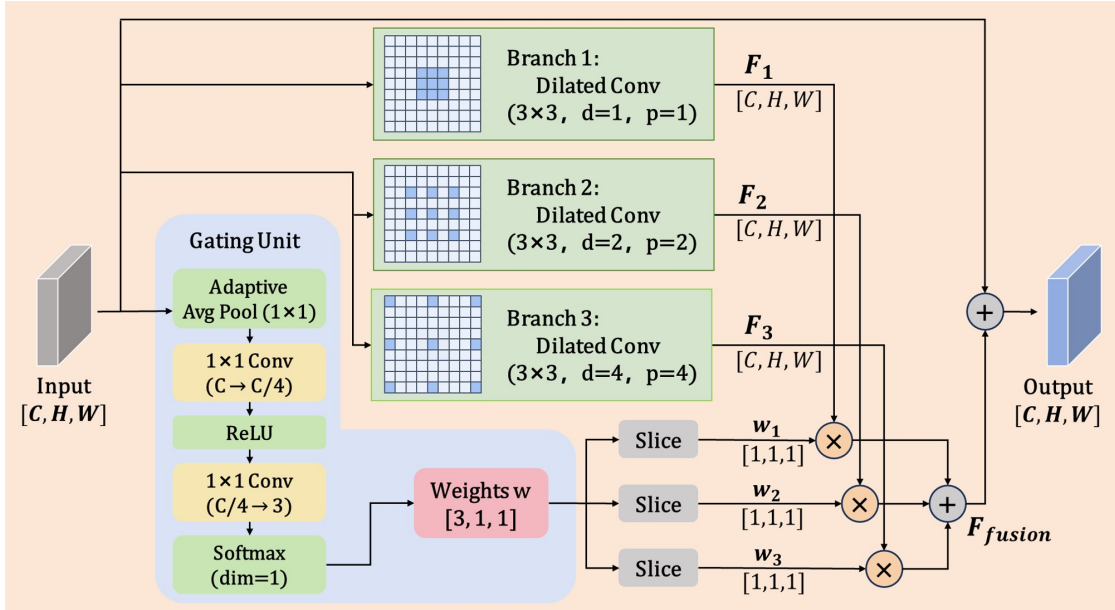


图 3-2 自适应尺度感知门控模块网络结构图

ASAG 模块整体由多尺度特征提取分支、门控权重生成和特征融合两部分组成，模块输入为通道数 C 的特征图 $X \in R^{B \times C \times H \times W}$ ，其中 B 为批次大小， H 和 W 分别为特征图的高、宽，输出为同尺寸的尺度自适应增强特征图 $Y \in R^{B \times C \times H \times W}$ 。该模块嵌入在 ResNet34 layer2 之后（128 通道特征层），该层为网络中层特征层，既保留了足够的空间细节，又具备初步的语义信息，是尺度自适应特征提取的最优阶段。

(1) 多尺度特征提取分支

为捕捉不同尺度的肺部肿瘤的特征，模块设计了 3 个并行的膨胀卷积分支，所有分支均采用 3×3 卷积核，通过设置不同的膨胀率（dilation rate）构建不同大小的感受野，且每个分支均保持输入输出通道数一致、特征图尺寸不变，确保后续特征融合的可行性。膨胀卷积的核心优势是在不增加卷积核参数和计算量的前提下，通过扩大卷积核的感受野范围，捕捉更广泛的上下文信息，其感受野大小计算如公式 3-2 所示：

$$RF = k + (k - 1) \times (d - 1) \quad (3-2)$$

其中， RF 为卷积实际感受野大小， k 为卷积核尺寸， d 为膨胀率。

细粒度尺度分支设置膨胀率 $d = 1$ ，卷积核为标准 3×3 卷积，由公式 3-2 计算得感受野 $RF = 3 \times 3$ 。该分支无膨胀，感受野最小，主要捕捉微小病灶的细粒度空间细节特征，如肺部肿瘤边缘的微小毛刺、局部纹理，同时保留特征图的高分

分辨率空间信息，避免小目标特征被稀释。其特征提取过程如公式3-3所示：

$$F_1 = Conv_{3 \times 3}^{d=1}(X) \quad (3-3)$$

其中， $X \in R^{B \times C \times H \times W}$ ，表示模块的输入特征图， C 为通道数， B 为批次大小， H 和 W 分别为特征图的高、宽。

中等尺度分支设置膨胀率 $d = 2$ ，计算得感受野 $RF = 5 \times 5$ 。该分支通过轻量膨胀扩大感受野，在保留一定空间细节的基础上，捕捉结节与周围肺实质的局部上下文关系，避免因感受野过小导致结节特征提取不完整。其特征提取过程如公式3-4所示：

$$F_2 = Conv_{3 \times 3}^{d=2}(X) \quad (3-4)$$

粗粒度尺度分支设置膨胀率 $d = 4$ ，由公式计算得感受野 $RF = 9 \times 9$ 。该分支为最大膨胀率，感受野最大，主要捕捉较大的肿瘤的全局语义特征，同时覆盖肿瘤周围的肺组织等背景信息，建立肿瘤内部以及肿瘤与其微环境的关联，提升对较大肿瘤的特征表征能力。其特征提取过程如公式3-5所示：

$$F_3 = Conv_{3 \times 3}^{d=4}(X) \quad (3-5)$$

三个分支并行处理输入特征图 X ，最终输出 3 个同尺寸、同通道的多尺度特征图 F_1 、 F_2 、 $F_3 \in R^{B \times C \times H \times W}$ ，分别对应细、中、粗三种尺度的肺部肿瘤特征，实现对不同尺寸结节的全覆盖特征提取。

(2) 门控权重生成分支

为了避免多尺度特征直接拼接或者相加导致的特征冗余和无效尺度干扰，例如，针对肺部的微小病灶时，大感受野分支提取的背景信息就成为了噪声，本章设计了轻量型门控权重生成分支。

这个分支以特征图 $X \in R^{B \times C \times H \times W}$ 为输入，采用自适应全局平均池化进行全局信息压缩，生成 $X_{glo} \in R^{B \times C \times 1 \times 1}$ ，聚合特征图的全局空间信息来避免局部信息主导权重生成，确保门控权重能反映整个特征图的尺度分布，即当前输入中占主导的肿瘤尺度。之后使用 1×1 卷积层将 X_{glo} 的通道数从 C 降维至 $C/4$ ，在保留全局尺度信息的前提下，简化后续权重映射的学习难度，降低特征冗余，避免因通道数过多导致门控网络过拟合，输出特征为 $X_{reduce} \in R^{B \times C/4 \times 1 \times 1}$ 。在通道降维后引入 ReLU 激活函数，为门控网络引入非线性表达能力，使门控权重能够学习复杂的“尺度-特征”映射关系，避免线性变换导致的权重表达能力不足。之后再次设置 1×1 卷积层，将降维后的特征通道数从 $C/4$ 映射为 3，与多尺度分支的数量保持一致，输出原始权重特征 $w_{raw} \in R^{B \times 3 \times 1 \times 1}$ ，实现从特征空间到权重空间的转换。最后在通道维度上执行 Softmax 归一化操作，将原始权重

w_{raw} 归一化至 $[0, 1]$ 区间，且满足 3 个权重之和为 1，输出最终的门控权重矩阵 $w = [w_1, w_2, w_3]^T \in R^{B \times 3 \times 1 \times 1}$ ，实现分支权重的动态分配，使门控网络能够明确选择当前输入中最有效的尺度分支，同时保证权重的可解释性（权重值越大，对应分支的特征越重要）。这一过程的公式如3-6所示：

$$w = \text{Softmax}(\text{Conv}_{1 \times 1}(\text{ReLU}(\text{Conv}_{1 \times 1}(\text{GAP}(X)))))) \quad (3-6)$$

其中，GAP 表示全局平均池化。

将门控子网络生成的动态权重 w 与对应分支的多尺度特征进行逐元素相乘，再对加权后的特征进行逐元素相加，实现多尺度特征的自适应融合，如公式3-7所示：

$$F_{fusion} = w_1 \odot F_1 + w_2 \odot F_2 + w_3 \odot F_3 \quad (3-7)$$

其中， \odot 表示广播后的逐元素相乘操作。

在动态融合特征 F_{fusion} 的基础上，引入恒等残差连接，将融合特征与模块原始输入特征 X 进行逐元素相加，得到 ASAG 模块最终的输出特征 Y ，其数学表达为公式3-8所示：

$$Y = F_{fusion} + X \quad (3-8)$$

这样设计一方面是保证特征信息的完整性，在尺度自适应融合过程中，部分细节特征可能会被弱化或丢失，而残差连接可将原始特征直接传递至输出端，从而尽量保留这些关键信息，避免不可逆的信息损失。另一方面是提升模块的训练稳定性，在反向传播过程中，残差连接为梯度提供了一条更直接的传播路径，有助于缓解深层网络中常见的梯度消失问题，使 ASAG 模块能够与主干网络 ResNet34 更稳定地联合训练。

3.2.3 空频双域特征增强模块

在肺癌免疫治疗的影像学评估中，病灶的形态学特征（如毛刺、分叶等边缘结构）和内部纹理异质性（如坏死区、毛玻璃样成分）都是用于判断疗效的重要影像信息。传统的卷积神经网络主要侧重于提取空间特征，但对频率域中包含的细粒度纹理信息利用不足。基于这一问题，本章设计了空频双域特征增强模块（Spatial-Frequency dual-domain feature Enhancement Module, SFEM），用于从空间域和频率域两个角度对 CT 影像进行特征分解与建模，并结合可学习向量和通道注意力实现特征的增强和融合，如图3-3所示。

SFEM 模块由空间梯度增强分支、频域纹理调制分支、通道注意力机制和特征融合模块四部分组成。通过局部均值差分提取空间域中的高频梯度信息，强化病灶的形态学边缘特征；通过二维快速傅里叶变换将特征映射至频率域，利用多

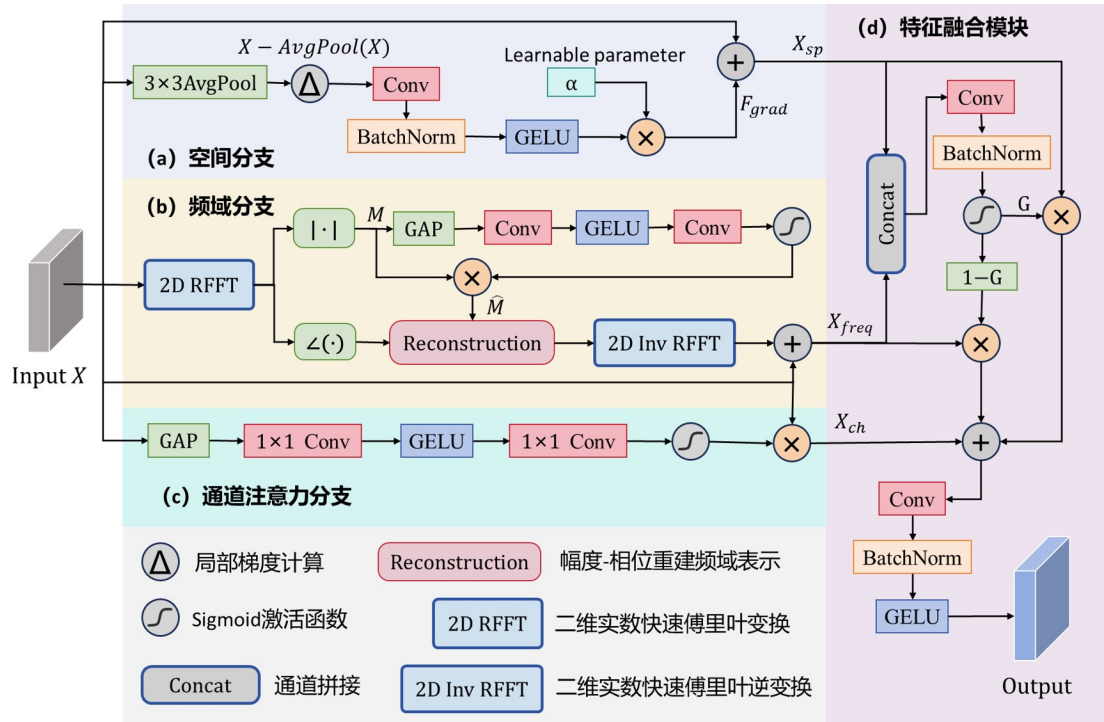


图 3-3 空频双域特征增强模块网络结构图

层感知机动态自适应增强关键频带能量，捕捉病灶内部深层异质性纹理；通过全连接网络提取通道注意力权重进行特征筛选；引入基于特征拼接的自适应门控机制计算空间域特征和频域特征的分配比例，完成空间和频率双分支特征的跨域软合并。该模块嵌入在 ResNet34 Layer3 后（256 通道特征层），该层为网络中高层特征，既保留了病灶边缘的梯度细节，又具备一定的语义区分度，是纹理特征增强的最优阶段。

(1) 局部空间梯度增强分支

特征图中某一个位置的纹理显著性由其与周围局部区域的特征值的差异决定。肿瘤边缘位置的特征值与邻域均值差异大，梯度值高，肺实质等平滑背景区域的特征值与邻域均值差异小，梯度值低。局部空间梯度增强分支的主要目标是强化肺部肿瘤的边缘结构与高频形态学特征。给定输入特征图 $X \in R^{B \times C \times H \times W}$ ，本模块通过局部平均池化操作提取低频平滑背景，然后通过逐元素差分的无参数操作模拟梯度计算，得到近似的局部梯度 ∇X ，如公式3-9所示：

$$\nabla X = X - AvgPool_{3 \times 3}(X) \quad (3-9)$$

这个梯度特征图凸显了病灶边缘的变化信息。跟传统的手工设计梯度算子，比如 Sobel 和 Prewitt 等相比，这样计算的局部梯度 ∇X 无需人工设计核参数，而是能根据网络特征图的分布规律自适应调整；相较于可学习的卷积梯度提取方法，这种方法可以避免引入额外参数，从而减少了过拟合的风险，而且还能保证

梯度提取过程快速高效，不会增加额外的计算负担。

为进一步编码梯度特征并恢复通道维度，模块采用深度可分离卷积（Depth-wise Convolution）构建轻量级特征提取单元。与标准卷积相比，深度可分离卷积将空间相关性与通道混合性解耦，通过独立的单通道卷积核提取各通道的空间特征，再以 1×1 的点卷积实现跨通道信息融合，在保持相近表征能力的同时显著降低参数量与计算开销。梯度特征 ∇X 依次经过核大小为 3×3 的深度可分离卷积、批归一化（Batch Normalization, BN）以及 GELU 激活函数 $\delta(\cdot)$ 进行非线性编码，如公式3-10所示：

$$F_{grad} = \delta(BN(Conv_{dw}(\nabla X))) \quad (3-10)$$

其中，GELU 激活函数相比 ReLU 具有更平滑的梯度特性，有助于缓解梯度稀疏问题，提升深层训练稳定性。

考虑到不同通道的梯度特征对最终预测的贡献存在差异，模块引入了可学习的空间权重因子 $\alpha \in R^{1 \times C \times 1 \times 1}$ 对编码后的梯度特征进行逐通道缩放，实现特征选择的自适应性。 α 的形状设计为 $[1, C, 1, 1]$ 的原因是：单参数对应单个通道，所有空间位置共享同一权重，既实现通道级差异化加权，又避免空间维度的过拟合，同时大幅减少参数量。最后将自适应选择的特征结合残差连接生成空间域特征表示 X_{sp} ，如公式3-11所示：

$$X_{sp} = \alpha \odot F_{grad} + X \quad (3-11)$$

其中， \odot 表示逐元素乘法。

（2）频域纹理调制分支

肺癌病灶内部有复杂的异质性，在空间域用小尺寸卷积核很难完全捕获，但是在频率域中，这些异质性表现为特定的幅度谱能量分布。为了解决这个问题，本章采用了二维实数快速傅里叶变换（2D Real Fast Fourier Transform, 2D RFFT）公式如3-12所示：

$$F(u, v) = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} f(x, y) e^{-j2\pi(\frac{ux}{H} + \frac{vy}{W})} \quad (3-12)$$

其中， $f(x, y)$ 是输入的二维图像或信号， $F(u, v)$ 是傅里叶变换后的频域表示， $e^{-j2\pi(\frac{ux}{H} + \frac{vy}{W})}$ 是傅里叶核函数，表示频域分量的复数域旋转。

将空间特征 X 映射至频率域，分解为幅度谱 M 和相位谱 P ，公式如3-13所示：

$$F(X) = M \cdot e^{jP} \quad (3-13)$$

对与免疫疗效相关的频带进行自适应增强，提取了幅度谱的全局平均能量，如公式3-14所示：

$$m_{global} = GAP(M) \quad (3-14)$$

通过多层感知机（MLP）生成频率门控权重 w_{freq} ，如公式3-15所示：

$$w_{freq} = \sigma(MLP(m_{global})) \quad (3-15)$$

其中， σ 表示 Sigmoid 激活函数。

利用权重 w_{freq} 对原始幅度谱 M 进行残差式增强，获得重校准后的幅度谱 \hat{M} ，而相位谱 P 保持不变，如公式3-16所示：

$$\hat{M} = M \odot (w_{freq} + 1) \quad (3-16)$$

最后通过二维逆实数快速傅里叶变换（2D Inverse Real Fast Fourier Transform, 2D Inv RFFT）将增强后的频域信号转换回空间域，形成频域分支输出 X_{freq} ，如公式3-17所示：

$$X_{freq} = X + F^{-1}(\hat{M} \cdot e^{jP}) \quad (3-17)$$

（3）通道注意力机制

在提取空、频双域特征时，本模块设计了一条基于通道注意力（Channel Attention）的直连通路，用于对输入 X 的通道特征进行全局特征筛选，筛选后的特征图为 X_{ch} ，如公式3-18、3-19所示：

$$\omega_{ch} = \sigma(MLP_{ch}(GAP(X))) \quad (3-18)$$

$$X_{ch} = X \odot \omega_{ch} \quad (3-19)$$

其中，GAP 表示全局平均池化， σ 表示 Sigmoid 激活函数。这个操作能够在融合前，滤除冗余的特征通道，保留对分类任务最具判别力的特征子集。

（4）特征融合模块

考虑到不同患者影像中空间形态特征和频率纹理特征的重要性可能存在差异，本章设计了一种基于通道拼接的自适应门控机制计算空频分配比例 G ，如3-20所示：

$$G = \sigma(BN(Conv_{1 \times 1}([X_{sp}, X_{freq}]))) \quad (3-20)$$

其中 $[X_{sp}, X_{freq}]$ 表示沿通道维度的拼接操作，BN 表示批归一化。

对空间域和频域的双分支特征 X_{sp} 和 X_{freq} 进行软合并。具体公式如3-21所示：

$$X_{fuse} = G \odot X_{sp} + (1 - G) \odot X_{freq} \quad (3-21)$$

其中， \odot 表示逐元素乘法。

融合后的特征 X_{fuse} 进一步与通道注意力分支的结果 X_{ch} 叠加，经过最终的映射层输出精炼后的特征图 Y ，如公式3-22所示：

$$Y = \delta(BN(Conv_{1 \times 1}(X_{fuse} + X_{ch}))) \quad (3-22)$$

该动态门控机制确保了模型能够根据不同肿瘤的内部异质性与边界清晰度，自适应地分配空域和频域特征的决策权重。

3.2.4 全局语义引导的噪声抑制模块

在肺癌免疫治疗疗效预测的深度学习任务中，模型往往面临严峻的“噪声干扰”挑战。肺实质背景、由于炎症引起的影像学阴影等等，在特征表示上可能与真实的肿瘤免疫治疗相关的特征产生混淆。为了解决这一问题，本节设计了全局语义引导的噪声抑制模块（Global Semantic-guided Noise Refiner, GSNR），利用反映病灶整体生物学属性的“全局语义”作为先验，通过计算全局语义与局部空间像素的一致性，结合背景抑制掩码，强制模型聚焦于病灶关键区域，从而剔除肺实质背景噪声，GSNR 的网络结构图如图3-4所示。

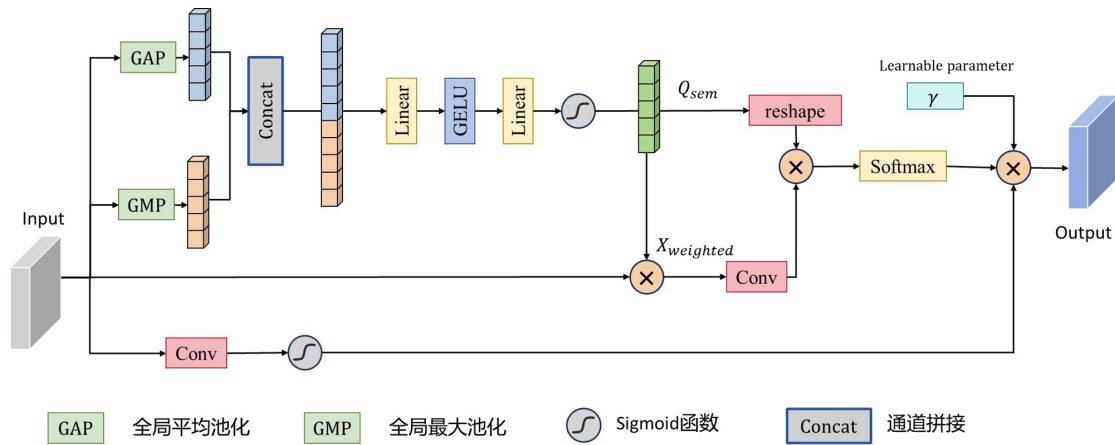


图 3-4 全局语义引导的噪声抑制模块网络结构图

为了捕获病灶特征的统计分布规律，GSNR 从输入特征图 $X \in R^{C \times H \times W}$ 中提取双路全局特征，通过并行的全局平均池化（GAP）和全局最大池化（GMP），分别获取通道均值 \bar{x} 和最大响应值 \hat{x} ，如公式3-23所示：

$$\bar{x}_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{c,i,j}, \hat{x}_c = \max_{i,j} X_{c,i,j} \quad (3-23)$$

将 \bar{x} 和 \hat{x} 两者拼接并映射至语义空间，通过非线性变换生成全局语义查询向量（Global Semantic Query） Q_{sem} ，如公式3-24所示：

$$Q_{sem} = \sigma(FC(GELU(FC([\bar{x}; \hat{x}])))) \quad (3-24)$$

其中, $[\bar{x}; \hat{x}]$ 表示特征拼接, FC 为线性全连接层, σ 为 Sigmoid 激活函数。利用 Q_{sem} 对输入特征进行初步通道加权, 生成增强感兴趣区域后的特征 $X_{weighted}$, 如公式3-25所示:

$$X_{weighted} = X \odot Q_{sem} \quad (3-25)$$

单一的通道加权无法抑制同通道内的空间噪声, 所以本模块中引入了“语义-空间”一致性校准机制, 将全局语义向量 Q_{sem} 与局部投影后的特征进行互相关运算。定义局部特征映射为 $\Phi_{value}(X_{weighted})$, 如公式3-26所示, 则空间注意力分布 $A_{spatial} \in R^{1 \times H \times W}$ 的计算过程如公式3-27:

$$\Phi_{value}(X_{weighted}) = Conv_{1 \times 1}(X_{weighted}) \quad (3-26)$$

$$A_{spatial} = Softmax\left(\frac{Q_{sem}^T \cdot \Phi_{value}(X_{weighted})}{\sqrt{d}}\right) \quad (3-27)$$

其中 d 为缩放因子。这一过程在几何意义上计算了每个局部像素点与全局病理属性的相关程度。只有在空间上符合全局肿瘤判别特征的像素点, 才会在 $A_{spatial}$ 中获得较高的激活权重, 从而实现基于语义的一致性校准。

GSNR 引入了噪声抑制模块来处理肺部影像中复杂的实质背景。通过轻量化的 1×1 卷积层对输入 X 进行空间维度建模, 直接学习背景噪声分布概率 $M_{mask} \in [0, 1]^{1 \times H \times W}$, 如公式3-28所示:

$$M_{mask} = \sigma(Conv_{1 \times 1}(X)) \quad (3-28)$$

掩码 M_{mask} 主要用于识别影像中的低置信度区域。最终, 模块通过将空间一致性权重、语义增强特征与噪声抑制掩码三者进行级联复合, 得到精炼后的特征表示, 如公式3-29所示:

$$X_{refined} = (A_{spatial} \odot X_{weighted}) \odot M_{mask} \quad (3-29)$$

引入一个初始值为 0 的可学习缩放参数 γ , 将噪声抑制后的残差信息补偿回原始特征, 保证训练初期的稳定性, 如公式3-30所示:

$$Y = \gamma \cdot X_{refined} + X \quad (3-30)$$

通过这一多级过滤机制, GSNR 可以实现对肺实质等噪声的物理降权, 提升了模型在复杂背景下对病灶异质性的表征精度。

3.3 实验结果与分析

3.3.1 数据集

本章在自建数据集 Xiangya-LIM 与公开的 LIDC-IDRI^[75] 肺结节数据集上进行了实验，其中自建数据集用于模型的核心训练与肺癌免疫治疗疗效预测的任务，LIDC-IDRI 数据集主要用于验证 MaxFrameNet 网络的特征提取与分类能力，两个数据集均对 CT 影像进行标准化处理，选取肺部肿瘤或者肺结节的最大帧作为模型输入，保证实验的一致性。

(1) Xiangya-LIM 数据集

与第二章的数据一致，本研究的数据来源于湖南省湘雅医院放射科，选择单独接受免疫治疗或免疫治疗联合化疗，且治疗前完成胸部全层 CT 扫描的患者，去除随访资料不完整，无法判定疗效的病例等，经过严格筛选，共纳入 334 例符合标准的患者作为研究对象。其中，175 例患者经免疫治疗后，靶病灶最大径之和减少 $\geq 30\%$ ，且疗效持续时间 ≥ 4 周，设为阳性组，即免疫治疗有效组；其余 159 例患者免疫治疗后靶病灶最大径之和减少 $< 30\%$ ，部分患者甚至出现病灶增大或新发病灶，归为阴性组，即免疫治疗无效组。

在第二章对原始数据的处理以及肺部肿瘤分割的基础上，在每个患者的 CT 序列中选择肿瘤区域最大的一帧。肿瘤区域大小以单帧内肿瘤掩码有效像素面积为量化指标，遍历患者所有 CT 切片对应的肿瘤分割掩码，计算每帧肿瘤像素的总面积，对比数值大小，取值最大者即为该患者的肿瘤区域最大帧。同时，以肿瘤区域外接矩形的中心点为中心，将 512×512 的图像裁剪出 150×150 的区域，既保留了肿瘤区域，又包含了肿瘤生长的微环境。

如图3-5所示为肺癌免疫治疗有效和无效的患者 CT 图像示例，第一列是原始肿瘤最大帧 CT 图像，第二列是肿瘤区域的掩码，第三列是裁剪后的肿瘤区域。

(2) LIDC-IDRI 数据集

LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) 是由美国国家癌症研究所等机构联合建立的公开胸部 CT 影像数据集，是肺结节检测、良恶性分类领域的经典基准数据集，被广泛应用于深度学习模型的性能验证。该数据集包含 1010 名患者的胸部螺旋 CT 序列，涵盖不同年龄、性别、吸烟史的受试者，经过解析，共有 875 例患者有 1 个或多个肺结节，在 CT 影像上共标注出 2619 个肺结节，所有肺结节均由 4 名具有专业资质的放射科医师独立进行良恶性程度评分，为肺结节良恶性分类提供了权威、客观的标注依据，每位医师对肺结节的良恶性程度按 1 至 5 分进行评分。评分标准为：1 分表示高

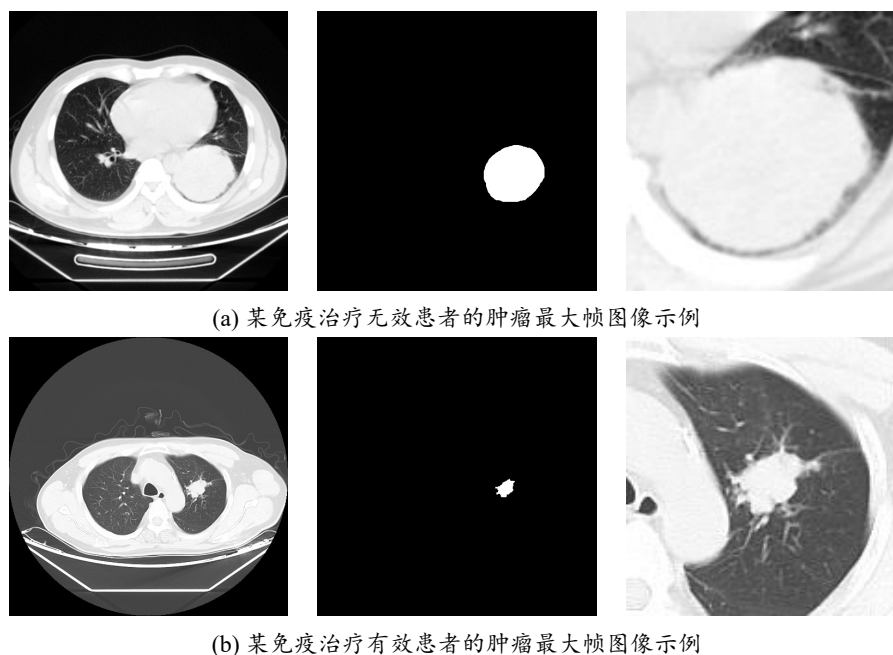


图 3-5 Xiangya-LIM 数据集中免疫治疗无效和有效患者的图像示例

度良性可能，2 分表示中度良性可能，3 分表示良恶性不确定，4 分表示中度恶性可能，5 分表示高度恶性可能。本研究对原始评分进行标准化整合：将 1-3 分合并为良性肺结节，4-5 分合并为恶性肺结节；对于同一肺结节的 4 名医师评分，采用多数投票法确定最终良恶性标签，若 4 名医师评分出现平局，选取评分的平均值进行判定，保证标注结果的合理性与客观性。

与自建数据集保持一致的处理流程，将 LIDC-IDRI 数据集中原始 DICOM 格式的 CT 序列重采样后，设定肺窗的窗宽为 1500 HU、窗位为 -600 HU，将 HU 值进行截断，并将有效灰度线性映射至 $[0, 255]$ 范围，筛选出包含肺结节的影像，并以肺结节标注区域的外接矩形的中心点为中心，将 512×512 的原始 CT 图像裁剪出 50×50 的区域，将得到的影像的像素值转换为 8 位无符号整型 (uint8) 格式，并以 PNG 格式进行存储。

数据集按照 8: 2 的比例划分训练集和测试集，如表 3-1 所示，其中训练集共有 2096 张图片，测试集有 523 张肺结节图片。

表 3-1 LIDC-IDRI 数据集划分情况

数据集	良性	恶性	总计
训练集	1640	456	2096
测试集	410	113	523
总计	2050	569	2619

如图3-6所示为 LIDC-IDRI 数据集中良性和恶性肺结节的最大帧 CT 图像示例。

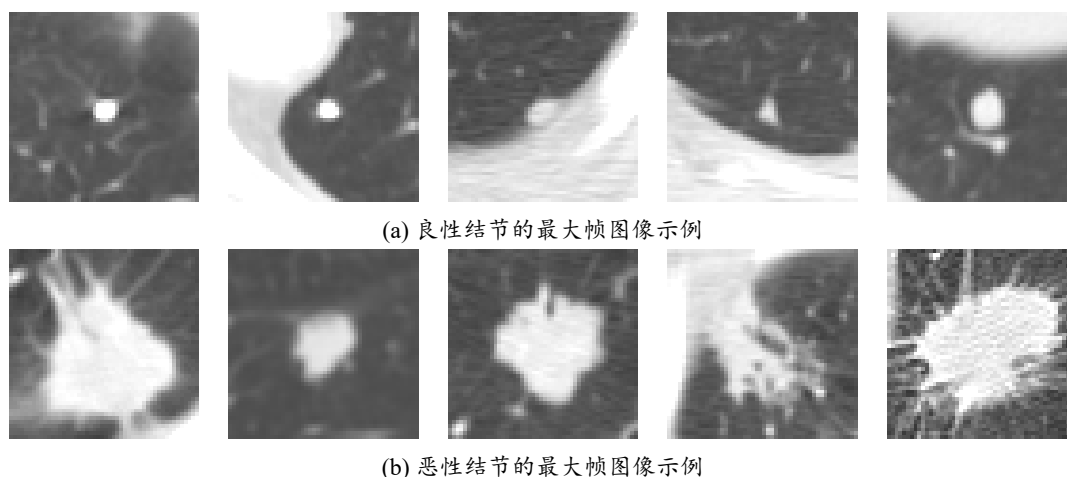


图 3-6 LIDC-IDRI 数据集中良性和恶性肺结节的图像示例

3.3.2 评价指标

为全面、科学地评估模型效能，本节采用了四项核心评价指标：受试者曲线下面积（Area Under the Curve, AUC）、准确率（Accuracy, ACC）、灵敏性（Sensitivity）、特异性（Specificity）。

AUC 表示受试者工作特征曲线（ROC curve）下的面积，衡量分类模型在所有可能的分类阈值下的整体性能，越高的 AUC 值表示模型的整体预测性能越好。

准确度是所有正确预测的样本数占总样本数的比例，表示模型整体预测的准确程度。公式如3-31所示：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3-31)$$

其中， TP (True Positive) 表示真正例，即模型正确预测为阳性的样本数； TN (True Negative) 表示真反例，即模型正确预测为阴性的样本数； FP (False Positive) 表示假正例，即模型错误预测为阳性的样本数； FN (False Negative) 表示假反例，即模型错误预测为阴性的样本数。

敏感性表示在所有实际为阳性的样本中，模型正确预测为阳性的比例。它反映了模型找到所有阳性样本的能力。公式如3-32所示：

$$Sensitivity = \frac{TP}{TP + FN} \quad (3-32)$$

特异性表示在所有实际为阴性的样本中，模型正确预测为阴性的比例。它反映了模型找到所有阴性样本的能力。公式如3-33所示：

$$Specificity = \frac{TN}{TN + FP} \quad (3-33)$$

3.3.3 实验设置

在实验运行环境方面，硬件环境和软件框架和第二章的实验环境一致，见2.3.3节的表2-1。

在超参数设置方面，模型训练共进行200个Epoch，批次大小（Batch Size）设为8。学习率调度采用余弦退火策略，初始学习率设为 10^{-3} ，并逐渐降至 10^{-7} 。优化器选用Adam，以保证模型可以较快且稳定地收敛。

在数据预处理阶段，将肿瘤或结节区域图像统一缩放为 224×224 的输入尺寸，以满足模型的输入要求。为了提高模型的泛化能力并减少过拟合，在训练过程中加入了随机旋转（ $\pm 20^\circ$ ）等数据增强操作。这些增强仅在训练阶段使用，测试阶段保持原始图像不变，以保证评估结果的客观性。所有数据增强均以在线随机方式生成，因此训练数据并不是一个固定扩展的数据集，而是在训练过程中通过数据增强形成的一种动态数据分布。

3.3.4 对比实验

为了全面评估MaxFrameNet的性能，本研究在Xiangya-LIM数据集和LIDC-IDRI数据集上进行了验证。由于目前医学影像领域的大多数方法都是在深度学习经典模型的基础上进行改进和适配，所以本章选取了近年来具有代表性的模型作为对比方法，以更全面地验证所提出模型的有效性与优越性。

(1) Xiangya-LIM

本研究将提出的MaxFrameNet与当前主流的深度学习模型（包括经典卷积神经网络、轻量化网络、基于Transformer的模型以及基于Mamba的模型）在肺癌免疫治疗疗效预测任务上进行了对比实验，以五折交叉验证的平均值和方差作为各项评估指标的结果。实验结果如表3-2所示。

如表3-2所示，实验结果表明，本章所提出的方法在AUC、ACC、敏感性和特异性上均优于其他模型，AUC达到75.49%，相较于次优模型Overlock-XT提升了2.52%；准确率ACC达到70.09%，比次优模型Overlock-XT提升了0.91%。此外，本章提出的方法在灵敏度（0.7098）与特异性（0.7067）之间达到了高度平衡，灵敏度相比次优模型StartNet-s1提升了0.23%，特异性相比次优模型Overlock-XT提升了0.59%，这表明该模型在减少肺癌免疫治疗疗效有效性的漏诊与误诊方面具有显著的临床应用潜力。

传统的卷积神经网络模型，如Inception-v3在特征传递与融合过程中无全局语义约束，易将胸腔正常组织、扫描伪影、肺部无关背景噪声同步纳入特征学习流程，冗余干扰信息严重影响病灶有效特征表达。相比之下，MaxFrameNet的全局语义引导的噪声抑制模块依托高层全局语义精准定位病灶区域，定向过滤无关

表 3-2 Xiangya-LIM 数据集下各模型性能对比结果

Model	AUC	ACC	Sensitivity	Specificity
SwimTransformer ^[76]	0.6410± 0.0001	0.6286± 0.0001	0.6336± 0.0007	0.6269± 0.0001
ConvNext-T ^[77]	0.6583± 0.0000	0.6301± 0.0002	0.6360± 0.0002	0.6265± 0.0002
MedViT-T ^[78]	0.6580± 0.0006	0.6466± 0.0008	0.6814± 0.0007	0.6315± 0.0012
Inception-v3 ^[79]	0.6745± 0.0000	0.6557± 0.0001	0.6682± 0.0006	0.6486± 0.0000
She 等 ^[33]	0.6861± 0.0003	0.6562± 0.0006	0.6525± 0.0004	0.6587± 0.0012
MobileNet-v3 ^[80]	0.7079± 0.0007	0.6537± 0.0008	0.6407± 0.0013	0.6751± 0.0007
MedMamba-T ^[81]	0.7094± 0.0013	0.6722± 0.0007	0.6547± 0.0003	0.6929± 0.0018
StartNet-s1 ^[82]	0.7111± 0.0016	0.6960± 0.0012	0.7075± 0.0011	0.6861± 0.0016
EffcientNet-b0 ^[83]	0.7174± 0.0011	0.6857± 0.0013	0.6775± 0.0005	0.7006± 0.0016
Overlock-XT ^[84]	0.7297± 0.0008	0.6968± 0.0012	0.6930± 0.0008	0.7008± 0.0018
MaxFrameNet (Ours)	0.7549± 0.0003	0.7079± 0.0009	0.7098± 0.0003	0.7067± 0.0014

噪声，从而提升预测模型特征纯度与判别性能。

基于 Mamba 的模型 MedMamba，依赖固定卷积核与全局扫描，尺度采样静态、无法随病灶大小、形态动态调整。对肺部肿瘤这种多尺度、跨尺度、不规则、边界模糊的病灶，易丢失小病灶细节、大病灶全局语义。相比之下，MaxFrameNet 通过引入自适应尺度感知模块，能够动态调整不同层次特征的感受野大小，更好地捕捉肺部肿瘤的多尺度异质性特征，从而提升了模型的精度。

Transformer 架构在捕获长距离依赖方面具有优势，但在本任务中 SwinTransformer 和 MedViT 的表现都较差。Transformer 结构通常需要海量的数据进行预训

练。肺癌免疫治疗的数据由于标注获取的成本非常高、样本量相对较小，难以支撑 Transformer 复杂的参数训练，导致其在小样本场景下容易陷入局部最优。此外，免疫疗效的相关特征可能隐藏在微小的局部纹理中，Transformer 的全局注意力机制有时会忽略这些细微的局部生物标志物。

轻量化模型如 MobileNet-v3、EfficientNet-b0 等表现出了较好的灵活性，AUC 突破了 0.70。但是为了轻量化，模型深度、通道数、特征图分辨率被大幅压缩，高层语义特征抽象能力不足，对微小肺结节、低密度病灶、模糊边界的细节捕捉能力弱，而这些正是肺癌免疫治疗疗效预测的关键判别特征。相比之下，MaxFrameNet 通过引入空频双域特征增强模块，能多方面地提取肿瘤的内部纹理特征、边缘梯度特征等，从而提升了模型的判别能力。

(2) LIDC-IDRI 数据集

在针对肺癌免疫治疗疗效预测的研究中，模型对肿瘤病灶特征的捕捉能力是决定预测准确性的核心因素。鉴于目前临床上尚缺乏大规模、标准化的肺癌免疫疗效公开影像数据集，本研究首先在国际公认的肺结节评估数据集 LIDC-IDRI 上开展了良恶性分类性能的基准测试。本实验的设计目的是通过在一个具有高度临床共识和精细标注的大样本库中进行验证，评估本研究提出的 MaxFrameNet 在提取肺部病理形态学特征方面的有效性。

由于 LIDC-IDRI 数据集中包含的图片量较大，为在保证训练样本规模的同时提高实验效率，本章按照 8:2 的比例将数据集划分为训练集和测试集进行模型训练与性能评估，因此未采用五折交叉验证策略。

表 3-3 不同模型在 LIDC-IDRI 数据集上的性能对比

Model	AUC	ACC	Sensitivity	Specificity
SwimTransformer ^[76]	0.7418	0.6902	0.6726	0.6951
ConvNext-T ^[77]	0.7943	0.7247	0.7699	0.7122
MedViT-T ^[78]	0.8086	0.7992	0.6814	0.8317
Inception-v3 ^[79]	0.8150	0.7476	0.7522	0.7463
She 等 ^[33]	0.8074	0.7897	0.7168	0.8098
MobileNet-v3 ^[80]	0.8135	0.7610	0.7522	0.7634
MedMamba-T ^[81]	0.8403	0.7839	0.7080	0.8049
StartNet-s1 ^[82]	0.8288	0.8011	0.7788	0.8073
EfficientNet-b0 ^[83]	0.8563	0.8164	0.7699	0.8293
Overlock-XT ^[84]	0.8611	0.8011	0.8407	0.7731
MaxFrameNet (Ours)	0.8789	0.8184	0.8053	0.8220

实验结果如表3-3所示，通过对实验数据的对比分析可见，本章提出的模型

(MaxFrameNet) 在所有评价维度上均展现出了较好的实验结果。

在衡量模型整体分类性能的曲线下面积 (AUC) 指标上, 本模型达到了 0.8789, 这一数值不仅超过传统的卷积网络 Inception-v3 (0.8150) 和基于 Transformer 架构的网络 SwimTransformer (0.7418), 也优于近年来表现突出的轻量化网络 MobileNet-v3 (0.8135) 和针对医疗影像优化的 MedViT-T (0.8086)、MedMamba-T (0.8403)。从准确率 (ACC) 的角度来看, 本模型以 0.8184 的精度位列第一, 较经典模型 Inception-v3 提升了 6.39%。这种性能的提升, 有力地证明了本研究在网络架构设计中引入的特征融合机制能够更有效地整合全局语义信息与局部病灶细节, 从而在复杂的肺部解剖背景中精准定位并识别恶性风险信号。

在医学影像诊断的临床实践中, 敏感度 (Sensitivity) 与特异性 (Specificity) 的平衡是评价算法实用性的核心准则。本研究模型在保持高特异性 (0.8220) 的同时, 敏感度达到了 0.8053。对比实验中其他模型表现, 可以观察到一种普遍存在的权衡困境。Overlock-XT 模型虽然在敏感度上表现较好, 达到了 0.8407, 但其特异性仅为 0.7731, 这意味着该模型在减少漏诊的同时, 也带来了较高的误诊率, 在实际临床决策中可能引发不必要的医疗干预。反观 MedViT-T, 其特异性高达 0.8317, 但敏感度却低至 0.6814, 说明该模型在预测时更为保守, 容易出现漏诊情况。本研究提出的模型通过对正负样本特征空间的对齐, 实现了两项指标的协同优化。这种稳健的分类能力表明模型不仅能识别出明显的恶性特征, 也能对不明显的早期病变保持敏锐。

3.3.5 消融实验

(1) Xiangya-LIM 数据集

为了探究模型中不同模块的作用, 本节在 Xiangya-LIM 数据集上进行了消融实验。表3-4展示了消融实验的结果, 实验中我们系统性地移除模型的各个模块, 并采用五折交叉验证的平均值来评估模型性能。结果可见, 移除任意一个模块均会导致模型性能显著下降, 这表明每个模块在模型捕获相关特征的过程中均发挥着关键作用。

从实验结果可以看出, 在 Base 模型的基础上, 无论单独添加哪一个模块, 模型的性能均得到了不同程度的提升。加入 ASAG 后, AUC 从 0.6717 提升至 0.6791, ACC 提升了约 2.7%。这表明多尺度特征提取增强了模型对局部细节特征的提取能力。单独加入 SFEM 对 AUC 的提升最为显著, 从 0.6717 提升至 0.6934, 说明该模块通过整合空间细节特征与频域特征, 构建了一种互补性的表达方式, 形成了更为全面的特征表示。加入 GSNR 后, 模型的敏感度提升明显, 从 0.6453 跃升至 0.6818, 表明该模块通过减少噪声干扰, 显著降低了漏诊率。

表 3-4 不同模块在 Xiangya-LIM 数据集上的消融实验结果

Exp.	Base	ASAG	SFEM	GSNR	AUC	ACC	Sensitivity	Specificity
1	✓	×	×	×	0.6717	0.6447	0.6453	0.6417
2	✓	✓	×	×	0.6791	0.6617	0.6667	0.6582
3	✓	×	✓	×	0.6934	0.6455	0.6380	0.6658
4	✓	×	×	✓	0.6855	0.6567	0.6818	0.6515
5	✓	✓	✓	×	0.7163	0.6742	0.6817	0.6689
6	✓	✓	×	✓	0.7190	0.7061	0.7060	0.7064
7	✓	×	✓	✓	0.7272	0.6866	0.7067	0.6773
8	✓	✓	✓	✓	0.7549	0.7079	0.7098	0.7067

对比组合模块的实验数据，可以观察到叠加效应：当任意两个模块组合使用时，各项指标均优于单一模块模型。例如，Base+SFEM+GSNR 的组合将 AUC 提高至 0.7272，证明了不同模块在特征处理上具有良好的互补性，能够从多维度构建更稳健的特征表达。当同时集成 ASAG、SFEM 和 GSNR 三个模块时（即本章提出的最终模型 MaxFrameNet），其性能达到了最优水平：AUC 达到 0.7549，相比基础模型提升了约 8.3%。ACC 达到 0.7009，敏感度与特异性表现均衡且处于较高水平。

为了进一步证明肿瘤区域的分割对免疫治疗疗效预测的重要性，本章还设计了额外的消融实验，结果如表3-5所示，使用肿瘤区域作为模型输入与使用原始 CT 图像作为输入相比，AUC 从 0.7032 提升至 0.7549，准确率从 0.6609 提升至 0.7079，敏感性和特异性也有所提升。这表明，通过分割将模型的输入聚焦于肿瘤病灶本身，可以有效减少肺部背景组织及其他无关结构的干扰，使模型能够提取到更加具有判别性的肿瘤特征。因此，这种基于病灶区域引导的特征学习方式对于更准确地捕捉关键病灶信息、强化肿瘤相关影像表征，并进一步提升模型的整体判别能力和预测性能具有重要作用。

表 3-5 原始 CT 与肿瘤区域输入的消融实验结果

输入图像	AUC	ACC	Sensitivity	Specificity
原始 CT 图像	0.7032±0.0008	0.6609±0.0007	0.6611±0.0005	0.6604±0.0007
肿瘤区域图像	0.7549±0.0003	0.7079±0.0009	0.7098±0.0003	0.7067±0.0014

图3-7 展示了以原始 CT 图像作为输入与以肿瘤区域作为输入时模型可视化结果。第一列为原始 CT 图像，其中红色框标出了肿瘤所在区域；第二列为直接以原始 CT 图像作为输入时得到的 CAM（Class Activation Map）热力图；第三列

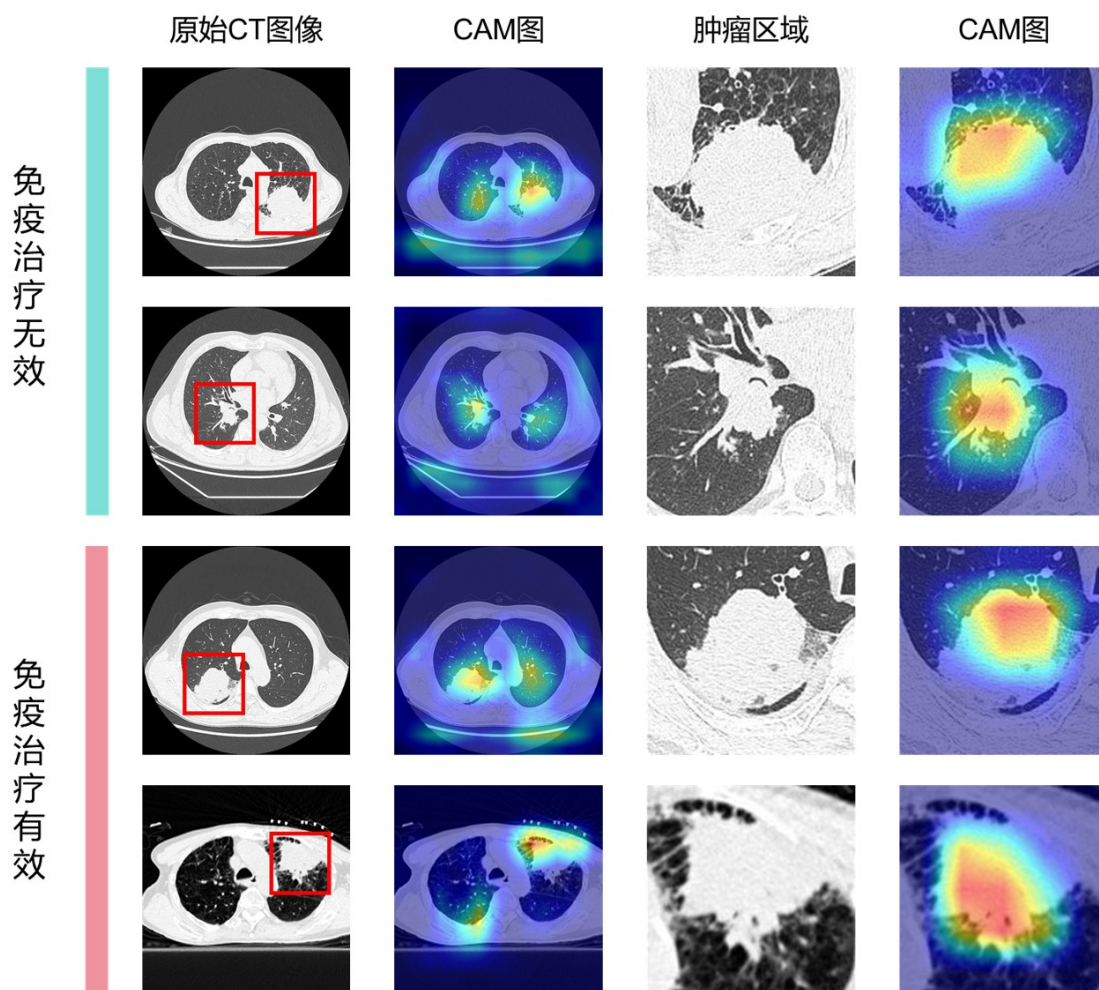


图 3-7 原始 CT 图像与肿瘤区域输入的可视化对比

为根据分割结果裁剪后得到的肿瘤区域图像；第四列为在该肿瘤区域图像基础上生成的 CAM 热力图。

CAM 热力图可以直观地展示模型在进行预测时主要关注图像中的哪些区域，其中红色区域表示模型关注度较高、权重较大的区域，而蓝色区域则表示模型关注度较低的区域。从图中可以观察到，当直接以原始 CT 图像作为输入时，模型的关注区域分布较为分散。除肿瘤区域外，一些正常组织或背景区域也被模型关注，这说明在复杂的肺部环境中，背景信息和健康组织可能对模型产生一定程度的干扰，使模型难以准确地捕捉关键的肿瘤特征，从而降低了预测效果。相比之下，当利用分割结果对原始 CT 图像进行裁剪，仅保留肿瘤区域作为模型输入时，模型的关注区域明显更加集中，热力图的高响应区域主要分布在肿瘤内部及其边界附近。这表明模型能够更加有效地聚焦于与肺癌免疫治疗相关的重要结构和纹理特征，同时减少背景噪声和无关组织的干扰。由此可以看出，基于分割结果对肿瘤区域进行裁剪，有助于引导模型更加关注关键病灶特征，从而提高模型的判别能力和预测性能。

(2) LIDC-IDRI 数据集

本章在 LIDC-IDRI 数据集上也进行了消融实验来探究模型中不同模块的作用。

表 3-6 不同模块在 LIDC-IDRI 数据集上的消融实验结果

Exp.	Base	ASAG	SFEM	GSNR	AUC	ACC	Sensitivity	Specificity
1	✓	×	×	×	0.7725	0.7093	0.7168	0.7073
2	✓	✓	×	×	0.7968	0.7323	0.7611	0.7244
3	✓	×	✓	×	0.8249	0.7476	0.7522	0.7463
4	✓	×	×	✓	0.8013	0.7266	0.7257	0.7268
5	✓	✓	✓	×	0.8565	0.7839	0.7876	0.7829
6	✓	✓	×	✓	0.8302	0.7418	0.7788	0.7317
7	✓	×	✓	✓	0.8498	0.8107	0.7965	0.8146
8	✓	✓	✓	✓	0.8789	0.8184	0.8053	0.8220

根据表3-6的实验数据，对比实验 1 与实验 2、3、4 可以观察到，在基础模型上单独添加任何一个改进模块，均能带来性能提升。SFEM 模块表现最为突出，加入 SFEM 后，AUC 从 0.7725 提升至 0.8249，提升了 5.24%，ACC 提升了 3.83%。这表明该模块在特征提取方面起到了核心作用，能有效捕捉判别性特征。ASAG 与 GSNR 模块表现稳健，两者分别将 AUC 提升了 2.43% 和 2.88%。其中，ASAG 在提升敏感性方面效果较好，而 GSNR 对整体准确率的增益较为均衡。

通过多模块组合实验（实验 5、6、7）可以发现，模块间的叠加并非简单的线性累加，而是存在显著的协同效应。当 ASAG 与 SFEM 同时使用时（实验 5），AUC 达到了 0.8565，显著优于两者单独存在时的表现。这说明 ASAG 提供的多尺度特征提取能够引导 SFEM 提取更具表达力的空间特征。对比实验 5 和实验 8（全量模型），加入 GSNR 后 AUC 从 0.8565 进一步优化至 0.8789。值得注意的是，在实验 7（SFEM+GSNR）中，特异性达到了 0.8146，说明 GSNR 模块在减少误诊（负样本识别）方面具有不可替代的价值。

3.4 本章小结

本章构建了一种面向肿瘤最大帧的深度学习模型 MaxFrameNet。针对肺部肿瘤尺度跨度大的问题，设计了基于多膨胀率卷积的多尺度特征提取结构，并引入轻量化门控子网络实现尺度权重的动态分配。面向肿瘤边缘结构与内部纹理异质性难以统一建模的问题，提出了空间域与频率域协同建模策略。针对肺实质背景及炎症等噪声干扰问题，构建了融合全局语义建模与空间一致性约束的噪声

抑制机制，从而增强模型对关键病灶区域的关注能力，提高特征判别性。在自建数据集 Xiangya-LIM 数据集上进行了免疫治疗疗效预测的实验，本章方法相较于现有主流算法在多项评价指标上均取得领先。在公开数据集 LIDC-IDRI 上进行了良恶性分类实验进一步验证了模型在提取肺部病理形态特征方面的有效性。消融实验进一步验证了所提架构中各功能模块对提升免疫疗效预测性能的独立贡献及其协同效应。

第4章 基于肿瘤多帧融合的肺癌免疫治疗疗效预测模型

4.1 引言

第三章提出了基于肿瘤区域面积最大帧 CT 图像来预测肺癌免疫治疗效果的模型 MaxFrameNet, 取得了不错的性能。然而, 这个方法仍然存在一些局限性。从临床诊疗实践来看, 肺癌免疫治疗疗效的临床表现复杂多样, 而且对应的影像学特征缺乏显著的特异性, 单帧 CT 图像仅能反映肿瘤某一层面的形态特征, 难以全面捕捉肿瘤的三维空间结构、生长趋势, 所以临床医生在制定诊疗方案时, 往往需要综合分析多幅连续 CT 切片, 通过整合序列影像信息构建肿瘤的三维视图, 以提升疗效评估的准确性与可靠性。这一临床需求也推动了研究方向逐步向 CT 序列影像的综合分析转变, 期望通过挖掘序列数据中的深层关联信息, 进一步提升预测模型的临床适配性。

尽管深度学习在三维图像分类、视频序列分析等领域已取得显著成果, 但将这些方法应用于肺癌免疫治疗疗效预测的 CT 序列影像分析时, 仍面临独特挑战: 医学影像采集的固有特性导致 CT 序列存在显著的空间相关性与信息冗余, 相邻 CT 切片的成像范围高度重叠, 使得序列数据中包含大量重复信息, 这不仅增加了模型的计算复杂度、提升了训练成本, 还可能导致模型过度拟合冗余信息。针对这一问题, 本章采用了等间距抽帧的方式, 从患者的 CT 序列中按照相等间距抽取四帧作为该患者的代表帧, 提取空间特征。

本章围绕肺癌免疫治疗的疗效预测问题展开研究, 在第三章的基础上, 构建了基于多帧 CT 影像融合的肺癌免疫治疗疗效预测模型 MultiFrameNet。针对全局特征提取不足的问题, 提出了 CNN-VMamba 双分支主干网络; 针对异构特征融合以及病灶尺度不确定性问题, 提出了跨尺度特征融合模块; 针对三维空间特征关联不足的问题, 提出了空间特征计算模块。该模型适用于肿瘤形态复杂、单一层面难以完整表征病灶信息的复杂病例, 可以在有充足算力资源的场景下提升预测的准确性。

4.2 算法模型

4.2.1 网络整体结构

本章设计了图像序列分类网络 MultiFrameNet, 如图4-1所示, 旨在实现肺癌免疫治疗疗效的精准预测。该网络以 CT 序列中等间距抽帧选取的 CT 图片集作为输入数据, 对于输入中的每一帧图像, 由 MaxFrameNet 与 VMamba^[86] 模型构成的双分支主干网络提取特征, 其中, MaxFrameNet 分支主要负责捕捉肿瘤的

局部特征信息，VMamba 分支则专注于挖掘全局特征关联，二者协同作用实现特征的深度解析。为了充分融合双分支主干提取的异构特征，同时解决肿瘤尺度不确定的问题，在 MaxFrameNet 与 VMamba 模型的前三个特征提取阶段之后均嵌入跨尺度特征融合模块（Cross-Scale Feature Fusion Module, CSFF），对两路分支输出的多尺度特征进行拼接与自适应融合，进一步增强特征的表达能力与鲁棒性。为了充分利用 CT 图像的空间关联性，网络中增设三维空间特征计算模块（Spatial Feature Calculation Module, SFCM），通过捕捉连续帧之间的空间关联信息，精准提取肿瘤的三维空间分布特征。最后，借助多层感知机（Multilayer Perceptron, MLP）对各帧图像的特征进行融合，实现肺癌免疫治疗疗效的分类预测任务。

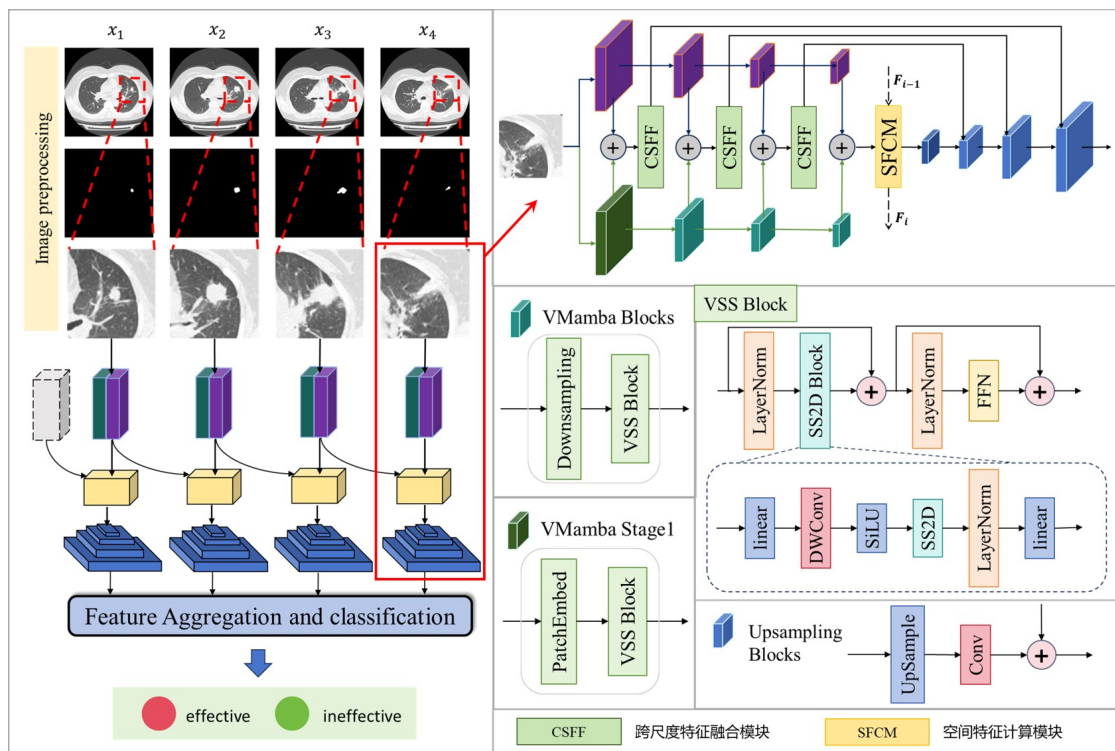


图 4-1 MultiFrameNet 网络结构图

4.2.2 双分支主干网络

卷积神经网络（Convolutional Neural Networks, CNNs）在局部特征提取方面表现出卓越的性能，能够精准捕捉肿瘤影像中的纹理细节、边缘特征及形态结构等局部信息，这与肺癌免疫治疗疗效评估中需关注肿瘤边缘收缩、密度变化等局部指标的临床需求高度契合。然而，CNN 固有的局限性在于难以建模特征间的长距离依赖关系，无法对复杂医学影像中的全局上下文信息进行整体整合，而全局信息（如肿瘤整体形态、与周围组织的空间关系）对疗效预测同样具有重要意

义。

为解决上述问题，本章提出一种创新的双分支融合架构，将 CNN 的局部特征提取能力与 VMamba 的全局序列建模能力进行协同融合，实现肺癌免疫治疗疗效的精准预测。该双分支架构相互补充、协同增效，既能够捕捉肿瘤的局部细粒度疗效相关特征，又能够整合全局上下文信息，有效提升网络的特征表征能力。

(1) CNN 分支设计

CNN 分支使用了第三章设计的 MaxFrameNet，如表4-1所示，以 ResNet34 为主干，由四个阶段组成，在第 2、3、4 阶段中引入自适应尺度感知门控模块 (Adaptive Scale-Aware Gating, ASAG)、空频双域特征增强模块 (Spatial-Frequency dual-domain feature Enhancement Module, SFEM)、全局语义引导的噪声抑制模块 (Global Semantic-guided Noise Refiner, GSNR)，网络详细的细节见第三章 3.2 节。

表 4-1 CNN 分支网络结构

阶段	网络结构	输出特征图形状
Stage1	7×7 Conv	64×56×56
	ResNet34 Layer1	
Stage2	ResNet34 Layer2	128×28×28
	ASAG	
Stage3	ResNet34 Layer3	256×14×14
	SFEM	
Stage4	ResNet34 Layer4	512×7×7
	GSNR	

(2) VMamba 分支设计

VMamba 是 2024 年提出的新一代视觉骨干网络，它的设计目标是突破传统卷积神经网络 (CNN) 与视觉 Transformer 模型在视觉特征建模中的固有局限，将自然语言处理领域取得突破性进展的选择性状态空间模型 (Selective State Space Model, S6) 成功迁移至视觉任务，构建了兼具线性计算复杂度、强全局长程依赖建模能力、硬件友好推理效率的统一视觉特征学习框架。

VMamba 的理论基础源于状态空间模型 (State Space Model, SSM)，这是一种实现线性复杂度序列建模的数学框架。连续时间域下的 SSM 定义为公式4-1、

4-2所示：

$$h'(t) = Ah(t) + Bx(t) \quad (4-1)$$

$$y(t) = Ch(t) + Dx(t) \quad (4-2)$$

其中， $h(t)$ 为连续时间隐状态向量， $x(t)$ 为模型输入序列， $y(t)$ 为模型输出序列， A 、 B 、 C 为固定参数矩阵， D 为直连投影矩阵。

由于深度学习模型在影像处理中处理的是离散化的特征数据，无法直接适配连续时间状态空间模型，因此采用零阶保持法（zero-order hold method）对连续时间 SSM 进行离散化处理，使其适配 CT 影像的离散特征建模需求。离散化后的状态空间模型表达式如公式4-3、4-4所示：

$$h_k = \bar{A}h_{k-1} + \bar{B}x_k \quad (4-3)$$

$$y_k = Ch_k + Dx_k \quad (4-4)$$

式中， \bar{A} 和 \bar{B} 如公式4-5和4-6所示：

$$\bar{A} = \exp(\Delta A) \quad (4-5)$$

$$\bar{B} = (\Delta A)^{-1}(\exp(\Delta A) - I) \cdot \Delta B \quad (4-6)$$

式中， Δ 为可学习的时间步长参数，由输入特征动态生成， \bar{A} 、 \bar{B} 为离散化后的状态转移矩阵与输入矩阵。

在此基础上，VMamba 将 A 、 D 设置为固定的可学习参数，将 B 、 C 、 Δ 改为由输入特征动态生成的结果，如公式4-7、4-8、4-9所示：

$$B = \text{linear}(x) \quad (4-7)$$

$$C = \text{linear}(x) \quad (4-8)$$

$$\Delta = \text{linear}(x) \quad (4-9)$$

改进后的状态空间模型，即选择性状态空间模型（Selective State Space Model, S6），如图4-2中的 S6 blocks 所示。这样模型能够根据视觉内容自适应调整隐状态更新权重，针对性地聚焦关键特征、抑制冗余信息，解决了传统 SSM 无法适配复杂视觉任务的问题。

VMamba 提出了 SS2D（2D State Space Model，二维状态空间模型）作为核心建模框架，其本质是通过多方向扫描将二维特征建模问题转化为多个一维 S6 blocks 建模问题，再通过融合还原二维特征。如图 4-2 所示，SS2D 主要包括三个部分：跨扫描 Cross-Scan、S6 模型，以及跨融合 Cross-Merge。

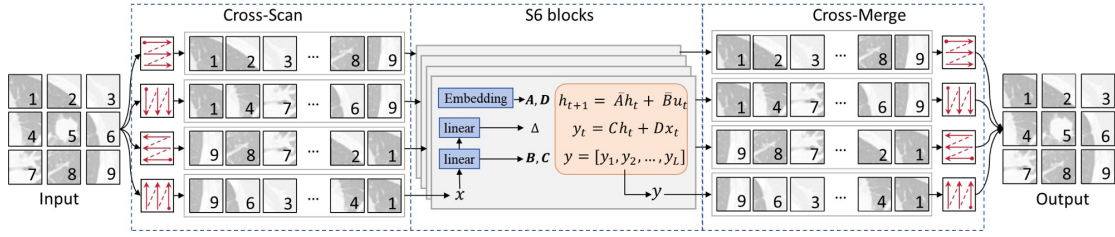


图 4-2 二维选择性扫描机制 (SS2D)

通过跨扫描操作，将输入特征图 X 沿水平、垂直、水平逆序、垂直逆序四个方向，将尺寸为 $H \times W \times C$ 的二维特征图展平为 4 条长度为 $L = H \times W$ 的一维序列，如公式4-10所示：

$$S = CrossScan(X), S \in R^{4 \times L \times C} \quad (4-10)$$

获得四组一维序列后，SS2D 模块针对每条扫描路径独立应用 S6 模块进行特征处理。S6 模块通过输入依赖的参数生成机制，实现了对各序列中不同位置特征的动态选择，对于每一组输入特征 $S_i, i \in 1, 2, 3, 4$ ，S6 模块通过特征选择机制生成输出 \tilde{S}_i ，如公式4-11所示：

$$\tilde{S}_i = S6(S_i) \quad (4-11)$$

这种位置感知的特征选择机制使得网络能够根据局部环境的不同灵活调整权重，突出重要信息，降低噪声影响，从而提升了特征表达的辨识能力。

S6 输出的序列需通过 Cross-Merge 跨融合操作恢复为二维空间特征，将四条建模后的序列按原扫描路径反向映射为二维特征图，再通过加权拼接与卷积融合多方向特征，最终输出与输入尺寸一致的二维特征。结合残差连接机制，VSS Block 的完整前向传播公式为：

$$Y = CrossMerge(\tilde{S}_i) + X \quad (4-12)$$

其中， X 为模块输入， Y 为模块输出，残差连接有效缓解了深度网络训练中的梯度消失问题，保证模型可稳定堆叠。

与 Transformer 的自注意力不同，S6 blocks 的计算复杂度仅为 $O(L)$ ，序列长度 L 增大时计算成本线性增长，在高分辨率视觉任务中具备绝对效率优势。

VSS 块的结构如图4-1所示，通过结合 SS2D 和 FFN 模块，将输入信号的空间特征和全局特征信息有效结合，增强模型的表现。FFN 和层归一化相结合，帮助提升模型的稳定性，避免梯度爆炸或消失，如公式4-13所示：

$$\hat{Y} = \max(0, W_2(LN(W_1Y + b_1)) + b_2) + Y \quad (4-13)$$

其中， \hat{Y} 是 VSS 块的输出特征， W_1 、 W_2 是权重矩阵， b_1 和 b_2 是偏置， LN 表示层归一化。

与 CNN 分支一样，VMamba 分支也可以分为四个阶段，如图4-1所示，第一阶段将输入图像分割成多个 patch，但没有将这些 patch 展平成一维序列，保留了图像的 2D 结构，再堆叠 VSS 块提取特征，第二至四阶段由下采样和多个 VSS 块组成，其中，一至四阶段中的 VSS 块的数量是 [2,2,9,2]。

多方向序列建模等机制为 VMamba 分支提供了全局感受野，能够有效捕捉 CT 影像中肿瘤与周围组织、肿瘤不同区域之间的长距离依赖关系，与 CNN 分支提取的局部特征形成互补，显著提升网络的整体特征表征能力，为后续疗效预测提供更全面的特征支撑。

4.2.3 跨尺度特征融合模块

在肺癌免疫治疗疗效预测任务中，为了充分融合 CNN 分支捕获的局部病灶纹理与 VMamba 分支捕获的全局上下文语义，本章设计了跨尺度特征融合模块 (Cross-Scale Feature Fusion Module, CSFF)，如图4-3所示。这个模块接收 CNN 分支和 VMamba 分支通道拼接后的异构特征作为输入，通过构建多尺度空间金字塔，配合共享权重的迭代卷积与特征筛选机制，实现特征的深度交互与自下而上的语义对齐，最终输出具有判别力的特征图，为免疫治疗的疗效预测提供可靠的表征。

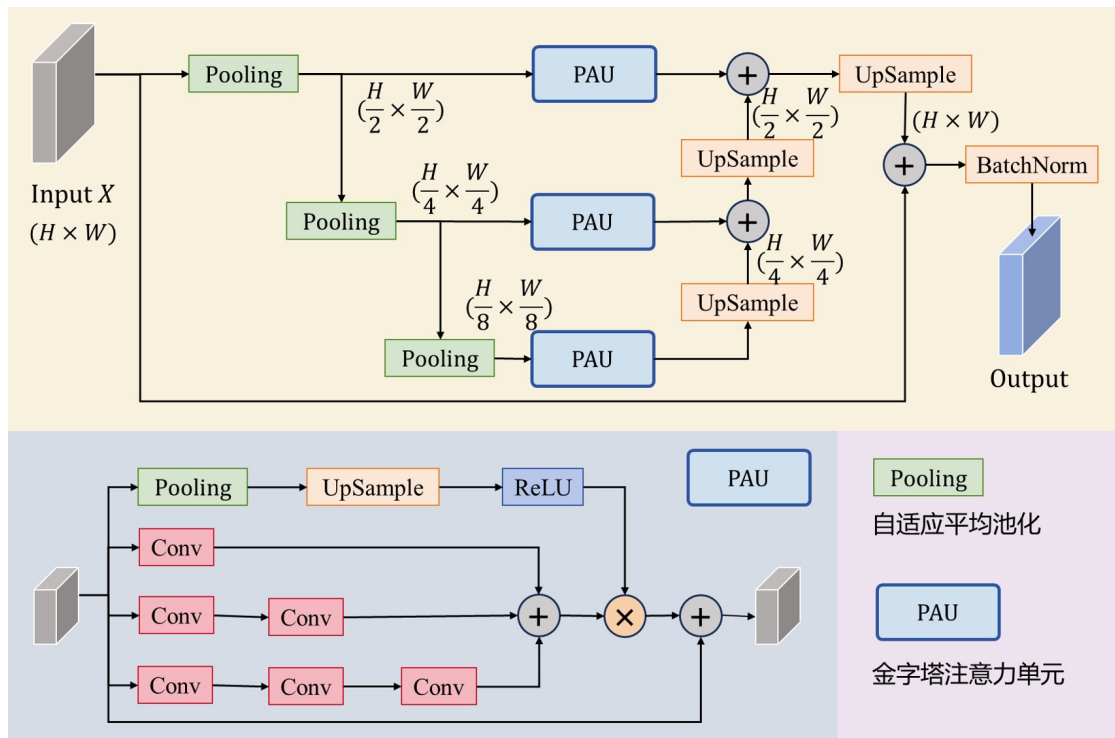


图 4-3 跨尺度特征融合模块网络结构图

跨尺度特征融合模块的核心构建块是金字塔注意力单元 (Pyramid Attention Unit, PAU), 如图4-3所示的 PAU 模块, 该单元通过递归的空间分解操作构建三级特征层次结构。

假设 PAU 模块的输入特征图为 $X' \in R^{B \times C \times H' \times W'}$, 其中 B 表示 Batch size, C 表示通道维度, H' 和 W' 分别表示空间高度与宽度, 通过三级串联卷积运算构建深度残差表征 R_1 、 R_2 、 R_3 , 如公式4-14、4-15、4-16所示:

$$R_1 = Conv(X') \quad (4-14)$$

$$R_2 = Conv(Conv(X')) \quad (4-15)$$

$$R_3 = Conv(Conv(Conv(X'))) \quad (4-16)$$

卷积操作对各尺度特征图进行精炼, 去除冗余信息, 将三级卷积获取的特征图相加得到多尺度特征 R_4 , 如公式4-17所示:

$$R_4 = R_1 + R_2 + R_3 \quad (4-17)$$

PAU 模块中设计了一个分支用于对多尺度特征 R 的通道特征进行全局特征筛选, 如公式4-18示:

$$\omega = Relu(Upsampling(pooling(X'))) \quad (4-18)$$

通过残差连接将这些筛选后的多尺度特征与原始输入特征进行融合, 既保留原始细节信息, 又整合多尺度抽象特征, 如公式4-19所示:

$$R = \omega \odot R_4 + X' \quad (4-19)$$

假设跨尺度特征融合模块中三个 PAU 模块输出的特征分别为 $P_0 \in R^{B \times C \times \frac{H}{2} \times \frac{W}{2}}$, $P_1 \in R^{B \times C \times \frac{H}{4} \times \frac{W}{4}}$, $P_2 \in R^{B \times C \times \frac{H}{8} \times \frac{W}{8}}$, 通过上采样和相加操作进行融合, 得到最终的输出 P , 如公式4-20所示:

$$P = X + UpSample(P_0 + UpSample(P_1 + UpSample(P_2))) \quad (4-20)$$

跨尺度特征融合模块最终的输出 Y 可以表示为公式4-21所示:

$$Y = BN(P) \quad (4-21)$$

其中, BN 表示批量归一化, 提高神经网络训练的效率和稳定性, 以及加速收敛, 从而使得网络训练更加高效, 避免梯度消失或爆炸问题。

4.2.4 空间特征计算模块

单帧肿瘤区域的 CT 影像仅能反映肿瘤某一截面的形态与纹理特征，而 CT 影像序列包含了肿瘤的三维空间结构信息及不同截面的特征关联，能够更全面地反映肿瘤的整体形态与空间分布，对肺癌免疫治疗疗效预测的准确性具有重要提升作用。因此，本章对连续 CT 影像帧进行深度分析，在捕捉每帧影像中肿瘤的空间结构与纹理特征的同时，挖掘序列中这些特征的序列演化规律（即不同帧之间的特征关联），实现肿瘤三维空间特征的精准提取。针对 CT 影像序列中抽取出来的每帧肿瘤影像，本章不仅提取其原始外观特征，还创新性地引入空间特征计算模块（Spatial Feature Calculation Module, SFCM），用于捕捉帧与帧之间的空间关联特征，进一步丰富肿瘤特征的表征维度。SFCM 的结构示意图如图4-4所示。

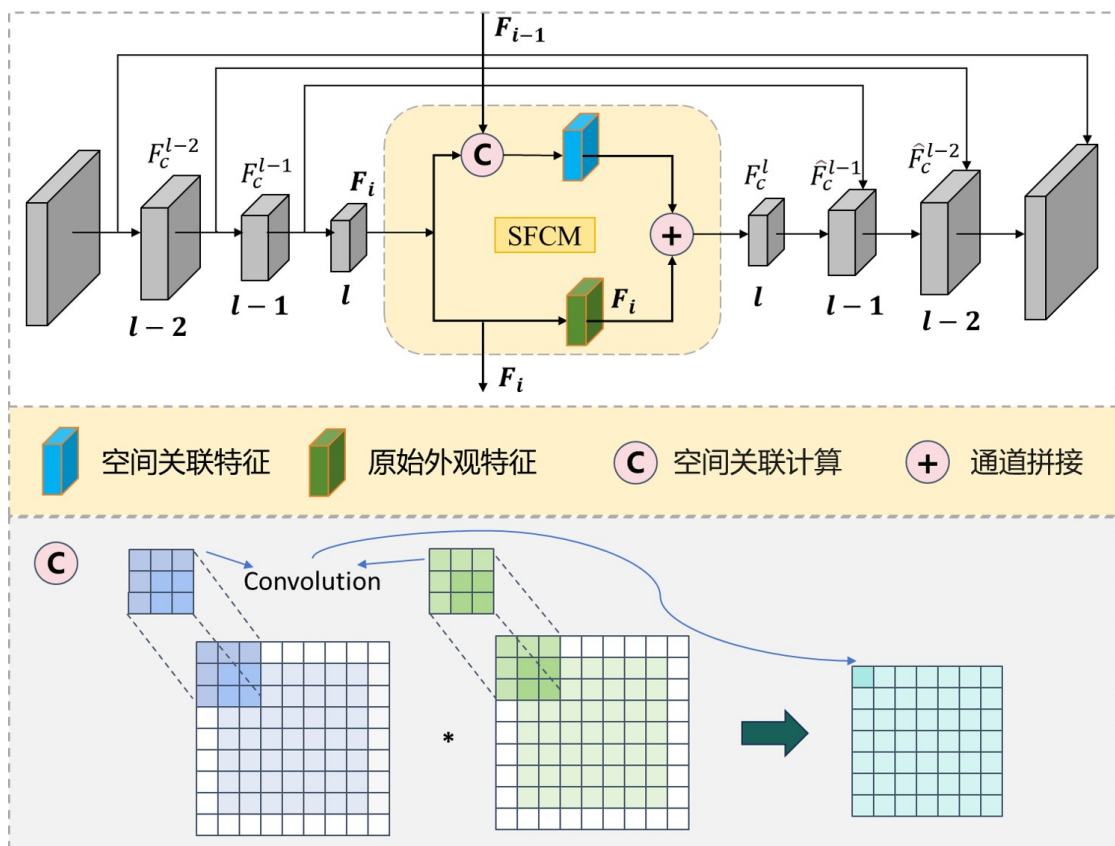


图 4-4 空间特征计算模块网络结构图

SFCM 的核心是计算相邻帧之间的空间关联特征，量化连续帧肿瘤特征的相似性与差异性，进而捕捉肿瘤的空间结构变化。本章将相邻两帧之间的空间关联关系定义为 C^l ，其数学表达式如公式4-22所示：

$$C^l(F_{i-1}, F_i) = conv(F_{i-1}^l, F_i^l) \tag{4-22}$$

其中， F_{i-1}^l 表示第 $i-1$ 帧 CT 图像的特征图经过金字塔网络结构的第 l 层输出

的结果, F_i^l 表示第 i 帧 CT 图像的特征图经过金字塔网络结构的第 l 层输出的结果, $conv$ 表示卷积操作, 用于计算两帧特征图之间的空间关联。这个卷积操作的实现流程如下: 首先对 F_{i-1}^l 和 F_i^l 进行边界填充 ($padding$ 设置为 1), 确保卷积后特征图尺寸与输入特征图一致; 随后采用 3×3 滑动窗口分别从 F_{i-1}^l 和 F_i^l 中提取子特征图; 将两个对应位置的子特征图进行逐元素相乘, 再对相乘后的结果求和, 最终生成相邻帧的空间关联特征图 C^l 。这个过程能够量化肿瘤在不同截面的形态变化, 捕捉上下帧肿瘤特征的空间对应关系, 为疗效预测提供空间结构依据。

为构建更全面、更具判别性的单帧肿瘤特征表示, 本文将通过公式4-22得到的空间关联特征与每帧的原始外观特征进行拼接融合, 同时通过 MLP 调整特征通道维度, 确保两者能够有效融合。融合过程的公式如4-23所示:

$$F_c^l = F_i^l + MLP^l (C^l (F_{i-1}^l, F_i^l)) \quad (4-23)$$

其中, F_i^l 表示第 i 帧 CT 图像在金字塔第 l 层的原始外观特征, C^l 表示第 i 帧 CT 图像与第 $i-1$ 帧 CT 图像在金字塔第 l 层的空间关联特征, MLP^l 表示第 l 层对应的多层感知机, 其核心作用是抽象出空间关联特征 C^l 中的高层特征, 同时对空间关联特征进行非线性变换与精炼, 提升模型的表达能力, 便于后续与原始外观特征的融合; F_c^l 表示融合空间关联特征后在金字塔第 l 层的特征图。

融合后的特征 F_c^l 将通过自上而下的金字塔结构进行特征传播, 金字塔结构能够有效捕捉肿瘤的多尺度纹理特征与空间结构特征, 对各层次的特征进行进一步精炼与增强, 确保模型能够全面感知从全局到局部的肿瘤信息, 提升特征的判别能力。金字塔结构的特征传播机制可表示为公式4-24:

$$\hat{F}_c^{l-1} = Conv (Upsample (F_c^l)) + F_c^{l-1} \quad (4-24)$$

其中, \hat{F}_c^{l-1} 表示金字塔第 $l-1$ 层经过特征传播后的增强特征图, $Upsample$ 表示上采样操作 (采用双线性插值法), 用于将第 l 层的融合特征 F_c^l 上采样至第 $l-1$ 层的特征图尺寸; $Conv$ 表示 1×1 卷积操作, 用于对于采样后的特征进行通道调整与精炼, 确保其与第 $l-1$ 层的原始融合特征 F_c^{l-1} 维度匹配; 通过残差连接, 将上采样精炼后的特征与第 $l-1$ 层的原始融合特征进行叠加, 实现多尺度特征的传播与融合, 进一步提升特征的代表能力。

CT 图像序列中的每一帧图像的特征提取完成后, 通过全连接层实现所有帧特征的全局融合。这一融合过程不仅增强了模型对肿瘤局部细粒度特征的理解, 还强化了模型对肿瘤整体结构与形态的识别能力, 为肺癌免疫治疗疗效预测提供更准确、更全面的特征支撑, 最终提升模型的预测精度与临床适用性。

4.3 实验结果与分析

4.3.1 数据集

与第三章一致，本章采用了自建数据集 Xiangya-LIM 与公开的 LIDC-IDRI 肺结节数据集。其中自建数据集用于模型的核心训练与肺癌免疫治疗疗效预测的任务，LIDC-IDRI 数据集主要用于验证 MultiFrameNet 网络的特征提取与分类能力。

(1) Xiangya-LIM 数据集

本章围绕患者 CT 序列开展研究，采用等间距抽帧的策略进行 CT 图像的选取，这种方式能够使选取的图像在整个扫描体积内均匀分布，从而完整覆盖目标病灶及周边解剖结构，有效减少局部信息缺失或特征冗余，同时具备结果确定、流程可重复的特点。与随机采样或者人工选择相比，等间距抽帧在保证数据整体代表性的前提下，无需复杂计算与额外开销，可为模型训练与性能优化提供稳定可靠的数据支撑。抽取方式如图4-5所示。若患者肿瘤区域的帧数不足四帧，则以最后一帧图像进行填充。

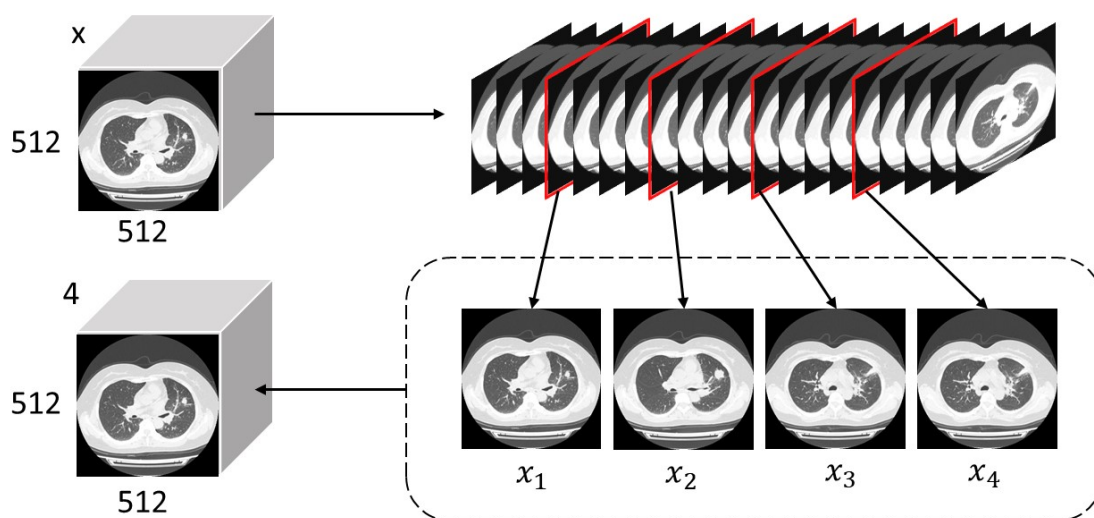


图 4-5 等间距抽帧示意图

帧数抽取完成后，在第二章肺部肿瘤分割的基础上，按照第三章数据的处理方法把每个患者的 CT 序列中的肿瘤区域裁剪出 150×150 的区域。如图4-6所示为 Xiangya-LIM 数据集中免疫治疗无效和有效患者的图像示例，第一组为免疫治疗无效患者的图像示例，第二组为免疫治疗有效患者的图像示例。

(2) LIDC-IDRI 数据集

LIDC-IDRI 数据集的抽帧方式和 Xiangya-LIM 数据集保持一致，均采用等间距抽帧的方式进行帧数的选取，确保数据处理的一致性与结果的可比性。

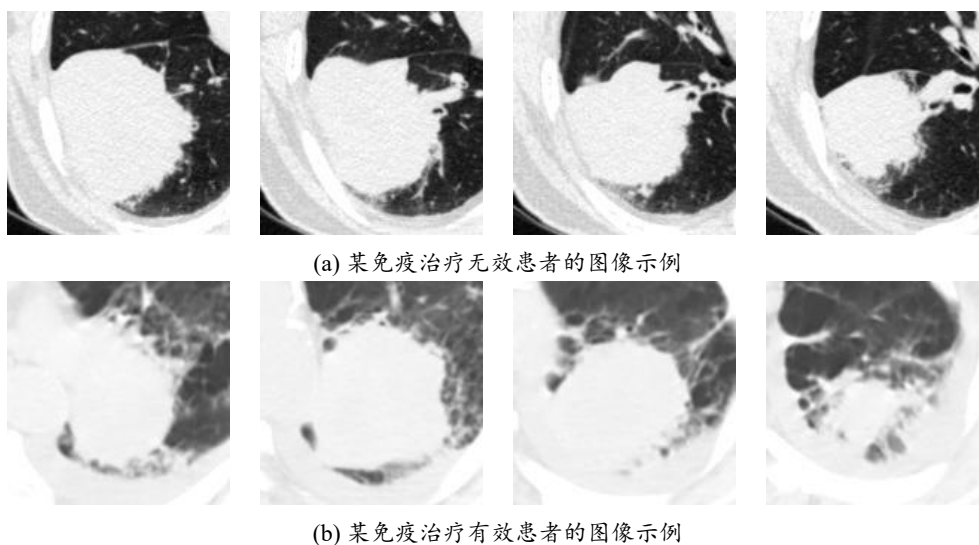


图 4-6 Xiangya-LIM 数据集中免疫治疗无效和有效患者的图像示例

抽帧完成后，以肿瘤区域外接矩形的中心点为中心，将 512×512 的 CT 图像裁剪出 50×50 的区域。

如图4-7所示为 LIDC-IDRI 数据集中良性结节和恶性结节抽帧后的图像示例。

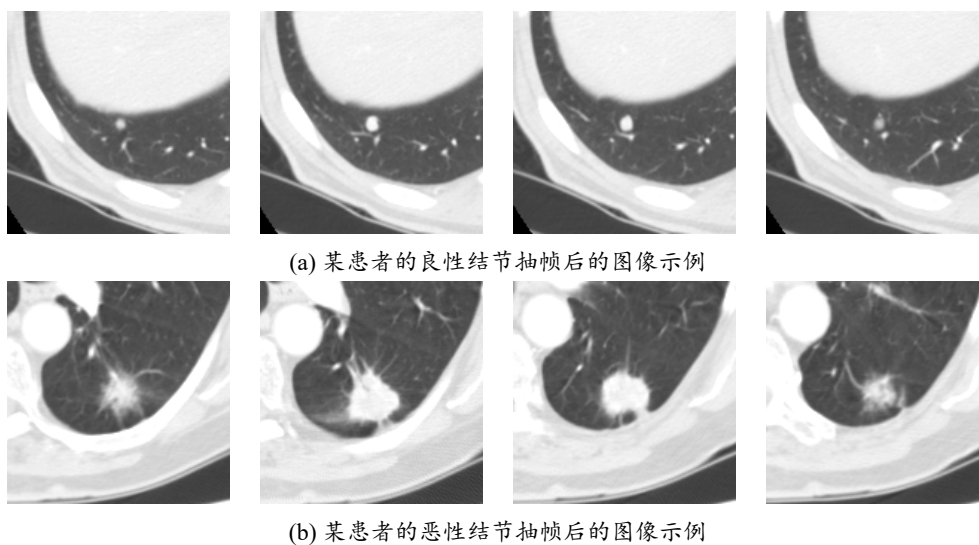


图 4-7 LIDC-IDRI 数据集中良性结节和恶性结节抽帧后的图像示例

4.3.2 评价指标

本节评价指标与 3.3.2 节一致。

4.3.3 实验设置

实验在 2.3.3 节所述的硬件平台和软件环境上进行，确保实验结果的可靠性和可重复性。

在超参数设置方面，训练周期设为 200 Epochs，批次大小 (Batch Size) 为 8。学习率调度采用余弦退火策略，初始学习率设为 10^{-3} ，并逐渐降至 10^{-7} 。优化器选用 Adam，以促进模型快速稳定收敛。

4.3.4 对比实验

为了全面评估 MultiFrameNet 的性能，本研究采用 Xiangya-LIM 数据集和 LIDC-IDRI 数据集进行验证。当前医学领域的大多数模型均基于深度学习领域的经典模型进行任务适配和改进，目前针对医学领域的肺癌免疫治疗的特定任务，尤其是基于三维图像的深度学习方法较为匮乏，大多是基于经典模型进行任务的适配和改进，所以本章从视频级分类的角度设计了对比实验，选取了基于三维图像的肺癌分类领域和近年来视频分类领域的代表性模型进行对比，全面验证模型的有效性和先进性。

(1) Xiangya-LIM 数据集：

本实验比较了多种模型在肺癌免疫治疗疗效预测任务中的表现，采用五折交叉验证，以 AUC (受试者曲线下面积)、ACC (准确率)、Sensitivity (灵敏度) 和 Specificity (特异性) 四项指标的平均值和方差对模型的性能进行综合评估。

详细的对比实验结果如表4-2所示，实验结果表明，本章所提的方法 MultiFrameNet 在受试者曲线面积、准确率、敏感性和特异性上均优于其它对比模型，AUC 达到 0.8068，准确率达到 0.7425，敏感性为 0.7655，特异性为 0.7171，表明 MultiFrameNet 在区分肺癌免疫治疗是否有效的能力上明显优于其他模型。

TAda2D 通过时间自适应卷积机制动态调整各 CT 切片的权重，AUC 达到 0.7780，ACC 达到 0.7239，敏感性和特异性分别达到 0.7314 和 0.7083，验证了自适应调制在识别治疗响应方面的有效性，但是其对肿瘤区域空间结构、尺度变化及纹理异质性的建模能力相对有限，相比之下，MultiFrameNet 以 MaxFrameNet 为基础，能够更全面地捕捉肿瘤的纹理特征等信息，显著提升了模型的预测性能。

DeepLung 是专门为肺部疾病设计的 3D 卷积神经网络，AUC 达到 0.7932，ACC 达到 0.7276，敏感性和特异性分别达到 0.7296 和 0.7150，表明其在多维特征提取方面具有优势，但其卷积架构在建模肿瘤的空间结构变化方面存在局限。相比之下，MultiFrameNet 通过空间特征计算模块，能够更全面地捕捉肿瘤的空间演变，从而提升分类性能。

表 4-2 各模型在 Xiangya-LIM 数据集上的性能对比结果

Model	AUC	ACC	Sensitivity	Specificity
DualPath Vit ^[87]	0.6835± 0.0000	0.6179± 0.0001	0.6114± 0.0015	0.6250± 0.0008
3DResnet ^[88]	0.6908± 0.0002	0.6617± 0.0002	0.6686± 0.0008	0.6544± 0.0004
Zhang 等 ^[45]	0.7158± 0.0015	0.6915± 0.0009	0.7048± 0.0002	0.6771± 0.0021
Video-Swin-T ^[89]	0.7194± 0.0006	0.6825± 0.0006	0.7486± 0.0001	0.6096± 0.0021
StreamFormer ^[90]	0.7343± 0.0002	0.7001± 0.0010	0.7333± 0.0007	0.6630± 0.0022
UniFormer ^[91]	0.7460± 0.0001	0.7090± 0.0001	0.7143± 0.0000	0.7031± 0.0002
Tada2D ^[92]	0.7780± 0.0018	0.7239± 0.0010	0.7314± 0.0000	0.7083± 0.0014
DeepLung ^[42]	0.7932± 0.0017	0.7276± 0.0001	0.7296± 0.0015	0.7150± 0.0029
MaxFrameNet(Ours)	0.7549± 0.0003	0.7079± 0.0009	0.7098± 0.0003	0.7067± 0.0014
MultiFrameNet(Ours)	0.8068± 0.0012	0.7425± 0.0009	0.7655± 0.0005	0.7171± 0.0016

UniFormer 借助多尺度关系聚合器统一处理局部卷积与全局自注意力，AUC 达到 0.7460 的展现了时空联合建模的优势，但其固定的浅层-深层分工在应对免疫治疗相关的复杂序列变化时灵活性略显不足。相比之下，MultiFrameNet 对深层和浅层特征均进行了跨尺度特征增强与融合，提升了模型对判别性特征的提取，从而提升了模型的性能。

相比之下，纯 Transformer 架构的 DualPath ViT 表现不佳，AUC 仅 0.6835，准确率 0.6179，敏感性和特异性分别为 0.6114 和 0.6250。这主要是由于参数量过大，在医学小样本场景下易产生过拟合，限制了模型对医学域数据的适应能力。StreamFormer 和 Video-Swin-T 虽然分别通过因果时间注意力和移位窗口机制实现了较高的敏感性，但特异性明显不足，分别为 0.6630 和 0.6096，假阳性率较

高，临床实用性受限，局部窗口机制可能过度关注炎症反应等干扰因素，导致对治疗响应的误判。相比之下，MultiFrameNet 通过双分支主干网络充分提取模型的全局和局部特征，并分层进行融合，提升特征表达能力，显著提升模型性能。

传统基线模型的性能普遍落后于现代架构，3D ResNet 和 Zhang 等人所提的方法，AUC 分别为 0.6908 和 0.7158，缺乏显式的序列或多尺度注意力机制，难以充分建模 CT 序列中帧与帧之间肿瘤的动态演变。相比之下，MultiFrameNet 提出了三维空间内的特征计算模块，充分融合序列帧的原始外观特征和空间结构变化，显著提升了模型在肺癌免疫治疗疗效预测上的能力。

值得关注的是，MaxFrameNet 为第三章中提出的基于单帧 CT 影像的肺癌免疫治疗疗效预测模型，AUC 为 0.7549，ACC 为 0.7079，敏感性和特异性分别为 0.7067 和 0.7098，展现了良好的性能，但与基于多帧 CT 影像的 MultiFrameNet 模型相比仍有差距。这一对比有力地证明了多帧信息整合对于提升预测准确性的关键作用。

基于上述分析，MultiFrameNet 通过双分支主干网络同时捕获序列图像数据中的局部细节信息与全局上下文信息，同时将异构特征分层跨尺度增强，充分融合肿瘤序列帧的原始外观特征与三维空间结构信息，显著提升了对肺癌患者免疫治疗疗效的预测能力，在敏感性和特异性之间取得了最佳平衡，既确保了潜在获益患者的检出，又避免了不必要的过度治疗，展现出最优的临床转化潜力。

(2) LIDC-IDRI 数据集

由于 Xiangya-LIM 数据集规模有限，且肺癌免疫治疗领域相关的公开数据集较为稀缺，为充分验证 MultiFrameNet 模型的有效性与泛化能力，本章进一步在公开数据集 LIDC-IDRI 上开展了肺结节的良恶性分类实验，以补充验证模型性能，数据的划分如表3-1所示。

LIDC-IDRI 数据集呈现明显的不平衡分布特征，测试集包含 410 例良性病例和 113 例恶性病例，阳性率仅为 21.6%，对模型的鉴别能力提出了较高要求。实验结果如表4-3所示。

实验结果表明，所提出的 MultiFrameNet 模型在 AUC 和准确率上均取得最优表现，其 AUC 达到 0.8910，准确率为 0.8279；敏感性和特异性达到了较好的平衡，敏感性为 0.8230，特异性为 0.8293，验证了 MultiFrameNet 在不平衡数据条件下的有效性。

MultiFrameNet 模型的 AUC 为 0.8910，较次优的 TAda2D (0.8849) 提升了 0.61%，较 DeepLung (0.8749) 提升了 1.61%，较 Video-Swin-T (0.8353) 提升 5.57%，较 3D ResNet (0.7912) 提升 9.98%。在准确率方面，MultiFrameNet (0.8279) 较 UniFormer (0.8203) 提升了 0.76%，较 MaxFrameNet (0.8184) 提升了 0.95%，较

表 4-3 各模型在 LIDC-IDRI 数据集上的性能对比结果

Model	AUC	ACC	Sensitivity	Specificity
DualPath Vit ^[87]	0.7720	0.7533	0.7080	0.7659
3DResnet ^[88]	0.7912	0.7801	0.7168	0.7976
Zhang 等 ^[45]	0.8064	0.7591	0.7699	0.7561
Video-Swin-T ^[89]	0.8353	0.7935	0.8230	0.7854
StreamFormer ^[90]	0.8217	0.8183	0.7522	0.8365
UniFormer ^[91]	0.8667	0.8203	0.7788	0.8317
Tada2D ^[92]	0.8849	0.8068	0.8230	0.8024
DeepLung ^[42]	0.8749	0.7992	0.8053	0.7976
MaxFrameNet(Ours)	0.8789	0.8184	0.8053	0.8220
MultiFrameNet(Ours)	0.8910	0.8279	0.8230	0.8293

StreamFormer (0.8183) 提升 0.96%。敏感性指标上, MultiFrameNet 与 TAda2D、Video-Swin-T 持平, 均为 82.30%, 较 DeepLung (0.8053) 提升 1.77%, 较 StreamFormer (0.7522) 提升 7.08%。特异性方面, MultiFrameNet (0.8293) 虽然较 StreamFormer (0.8365) 低 0.72%, 较 UniFormer (0.8317) 略低 0.24%, 但是 MultiFrameNet 在敏感性和但特异性的平衡上更好, 和其他模型如 TAda2D (0.8024) 相比, 提升了 2.69%, 较 Video-Swin-T (0.7854) 提升了 4.39%。

对比基于单帧 CT 影像的 MaxFrameNet 模型, MultiFrameNet 在 AUC 上提升了 1.21% (0.8910 vs 0.8789), 准确率提升 0.95% (0.8279 vs 0.8184), 敏感性提升了 1.77% (0.8230 vs 0.8053), 特异性提升了 0.73% (0.8293 vs 0.8220), 验证了多帧融合策略相较于单帧优化的有效性。

4.3.5 消融实验

(1) 模块消融

为了探究模型中不同模块的作用, 本节在 Xiangya-LIM 数据集上进行了消融实验。表4-4展示了消融实验的结果, 实验中我们系统性地移除模型的双分支结构、跨尺度特征融合模块 (CSFF) 及空间特征计算模块 (SFCM), 并采用五折交叉验证方法在数据集上评估模型性能, 表4-4中的结果为五折交叉验证的均值。

表4-4的结果表明, 对于等间距采样获取的四帧 CT 序列图像作为每个患者的输入数据, 单一 CNN 分支取得 0.7632 的 AUC 和 0.7082 的准确率, 敏感性和特异性分别为 0.7106 和 0.7069, 而单一 VMamba 分支表现略逊, AUC 为 0.7533, 准确率为 0.6998, 敏感性为 0.7057, 特异性为 0.6872, 这表明在孤立使用条件下, CNN 的局部特征提取能力优于 VMamba 的状态空间建模能力。将 CNN 与

表 4-4 各模块在 Xiangya-LIM 数据集上的消融实验结果

CNN branch	VMamba branch	CSFF	SFCM	AUC	ACC	Sensitivity	Specificity
✓	×	×	×	0.7632	0.7082	0.7106	0.7069
×	✓	×	×	0.7533	0.6998	0.7057	0.6872
✓	✓	×	×	0.7703	0.7154	0.7209	0.7078
✓	✓	✓	×	0.7854	0.7329	0.7511	0.7106
✓	✓	×	✓	0.7892	0.7385	0.7497	0.7153
✓	✓	✓	✓	0.8068	0.7425	0.7655	0.7171

VMamba 分支组合后，模型各项指标均提升了，AUC 达到 0.7703，准确率达到 0.7154，较最优单分支（CNN）分别提升 0.71% 和 0.72%，验证了双分支设计的互补性。

双分支主干网络引入 CSFF 模块后，模型 AUC 显著提升至 0.7854，较双分支基线提升 1.51%，准确率达到 0.7329，敏感性大幅提升至 0.7511，表明跨尺度特征融合有效整合了 CNN 的多尺度局部特征与 Mamba 的全局长程依赖，显著增强了模型对免疫治疗有效患者的识别能力。双分支主干网络单独引入 SFCM 模块同样带来性能增益，AUC 达 0.7892，准确率 0.7385，特异性提升至 0.7153，说明空间特征计算模块有效优化了特征的空间分布表达，提升了模型的分类性能。值得注意的是，SFCM 在特异性指标上表现优于 CSFF（0.7153 vs 0.7106），而 CSFF 在敏感性上更具优势（0.7511 vs 0.7497），揭示了两个模块的不同优化特性，CSFF 更擅长增强免疫治疗有效患者的检出，SFCM 更侧重于减少假阳性误判。

同时集成 CSFF 与 SFCM 的完整模型取得最优性能，AUC 达到 0.8068，较双分支基线（0.7703）提升了 3.65%，较仅含 CSFF 的模型（0.7854）提升了 2.14%，较仅含 SFCM 的模型（0.7892）提升了 1.76%；准确率达到 0.7425，敏感性 0.7655，特异性 0.7171。这一结果验证了 CSFF 与 SFCM 的协同效应：CSFF 通过跨尺度融合提供丰富的多粒度特征表征，SFCM 在此基础上进行精细的空间关系建模，二者相互促进。完整模型的敏感性较 CSFF 单模块提升了 1.44%（0.7655 vs 0.7511），特异性较 SFCM 单模块提升了 0.18%（0.7171 vs 0.7153），表明双模块联合使用实现了对两类错误率的同步压制。该消融实验结果可见，移除任意一个模块均会导致模型性能显著下降，充分证明了所提出各组件的有效性，以及模块间协同设计的必要性。

（2）帧数消融

由于患者的 CT 序列层面具有相邻性和相似性，本研究对 CT 层面序列采用

等距采样策略。为确定最优采样层面数量，本研究在 Xiangya-LIM 数据集上开展了针对性的消融实验，采用五折交叉验证的平均值作为各个指标的结果。

表 4-5 不同帧数下 MultiFrameNet 的性能表现

帧数 n	AUC	ACC	Sensitivity	Specificity
2	0.7698	0.7164	0.7353	0.7073
3	0.7863	0.7324	0.7468	0.7122
4	0.8068	0.7425	0.7655	0.7171
5	0.7992	0.7401	0.7552	0.7106
6	0.7816	0.7279	0.7409	0.7001
7	0.7459	0.7084	0.7124	0.6938
8	0.7138	0.6716	0.6800	0.6625

表4-5的实验结果表明，帧数与模型性能呈现明显的非单调关系，最优性能出现在 4 帧配置。

当抽取的帧数从 2 帧增加至 4 帧时，模型性能持续提升，2 帧配置下 AUC 为 0.7698，准确率为 0.7164，敏感性 0.7353，特异性 0.7073；3 帧时 AUC 为 0.7863，相比 2 帧提升了 1.65%，准确率 0.7324 (+1.6%)，敏感性 0.7468 (+1.15%)，特异性 0.7122 (+0.49%)；4 帧达到峰值，AUC 达到 0.8068，较 2 帧提升 3.7%，较 3 帧提升 2.05%，准确率 0.7425，敏感性 0.7655，特异性 0.7171，四项指标均为最优。这一趋势表明，适度增加帧数能够丰富肿瘤异质性的时空表征，提升模型对免疫治疗响应模式的捕获能力。

当抽取的帧数超过 4 帧后，模型性能呈现显著衰减趋势。5 帧配置下 AUC 为 0.7992，相比 4 帧时降低了 0.76%，敏感性降至 0.7552 (-1.03%)；6 帧时 AUC 进一步下滑至 0.7816 (-2.52%)，准确率 0.7279 (-1.46%)，敏感性 0.7409 (-2.46%)，特异性 0.7001 (-1.7%)；7 帧和 8 帧配置下性能恶化，AUC 分别为 0.7459 和 0.7138，8 帧准确率甚至跌破 0.70，仅有 0.6716，敏感性降至 0.6800。这一衰减现象揭示了帧数过多带来的负面影响：帧数过多导致相邻切片冗余度增高，引入重复信息的同时增加了噪声干扰。

4 帧配置取得了最佳的性能，该配置下模型以相对精简的输入实现了最高的敏感性，达到 0.7655，确保了对免疫治疗有效人群的高检出率，同时特异性 (0.7171) 维持在合理水平，有效控制了假阳性风险。当抽取帧数较少时，模型能够获取的空间上下文信息有限，难以充分刻画肿瘤在相邻切片之间的结构变化；当输入帧数逐渐增加时，多帧特征需要在同一嵌入空间中进行联合表达，由于相邻 CT 切片之间具有较高的结构相关性，过多帧输入会导致特征之间的冗余程度

显著提升，使得不同帧的特征在共享嵌入空间中发生一定程度的特征冲突与表示压缩现象，从而在特征融合阶段削弱关键判别信息的表达能力。相比之下，抽取4帧时既能够提供一定的空间上下文信息，又不会引入过多冗余切片，使得模型在特征表达能力与信息有效性之间达到较好的平衡。因此，在本实验设置下，4帧输入能够更充分地保留具有判别性的肿瘤结构信息，从而获得最佳的预测性能。

4.4 本章小结

为了解决单帧CT图像反映的特征有限，难以提供肺部肿瘤的空间结构信息，本章构建了以多帧CT图像序列为输入的分类网络MultiFrameNet，以实现肺癌免疫治疗疗效的精准预测。网络以等间距抽帧的CT图像序列为输入，通过第三章提出的MaxFrameNet分支提取肿瘤局部特征，同时利用VMamba分支挖掘全局特征关联，实现对局部-全局特征的深度解析。针对肿瘤尺度不确定性，引入跨尺度特征融合模块(CSFF)，进一步增强特征表达能力。为充分利用帧与之间的空间信息，设计了三维空间特征计算模块(SFCM)，精确捕捉肿瘤的三维空间分布特征。在自建数据集Xiangya-LIM数据集和公开数据集LIDC-IDRI上的对比实验表明，本章方法相较于现有主流算法在多项评价指标上取得领先。同时，消融实验进一步验证了所提架构中各功能模块对提升免疫疗效预测性能的独立贡献及其协同效应。

第 5 章 总结与展望

5.1 总结

肺癌是全球发病率和死亡率最高的恶性肿瘤之一，晚期患者的 5 年生存率长期处于较低的水平。目前，免疫治疗已成为肺癌治疗的重要手段，然而，并非所有患者都对免疫治疗敏感，有些甚至出现严重的不良反应。因此，治疗前有效预测肺癌患者接受免疫治疗的疗效具有重要意义。现阶段主要的预测手段高度依赖有创基因检测或病理分析，不仅侵入性强，而且检测费用昂贵。为了通过无创手段准确筛选出免疫治疗的潜在获益人群，本研究对肺癌患者的胸部 CT 影像进行特征挖掘，在实现自动化肺部肿瘤区域精准分割的基础上，进一步构建了基于肿瘤面积最大帧和基于多帧 CT 影像的免疫治疗疗效预测模型。本文的主要研究内容、贡献及核心创新点如下：

(1) 针对临床中高质量标注数据获取难度大、标注成本高的问题，以及肺癌免疫治疗的疗效预测对肿瘤区域特征聚焦的需求，提出了一种基于双学生交叉指导与不确定性校正的半监督分割框架。为了打破单一学生网络在无标签数据上容易陷入的预测同质化困境，设计了双学生并行结构，通过引入分歧损失和交叉指导机制，迫使两个学生网络在不同的特征空间进行探索，增加模型对未标注样本的表征多样性。针对病灶边界模糊导致的噪声干扰，创新了一种融合不确定性筛选与特征原型双重优化策略的高质量伪标签构建方法。通过多次随机推断捕捉像素级的预测波动性，构建不确定性掩码，并进一步提取特征空间的全局类别原型，利用语义一致性约束校正伪标签，提高肺部肿瘤的分割精度。该模型不仅为免疫治疗的疗效预测提供了肿瘤区域的聚焦，而且为能够显著提升临床工作效率，减轻放射科医师与临床医生在大量 CT 影像上手动勾画病灶的繁重负担。

(2) 在肺部肿瘤自动化分割的基础上提出了一种基于最大帧多维度特征提取的肺癌免疫治疗疗效预测模型 MaxFrameNet。针对肺部肿瘤尺度跨度大的问题，设计了基于多膨胀率卷积的多尺度特征提取结构，并引入轻量化门控子网络实现尺度权重的动态分配；面向肿瘤边缘结构与内部纹理异质性难以统一建模的问题，提出了空间域与频率域协同建模策略；针对肺实质背景及炎症等噪声干扰问题，构建了融合全局语义建模与空间一致性约束的噪声抑制机制，从而增强模型对关键病灶区域的关注能力，提高特征判别性。该模型可应用于临床的辅助评估，在无创条件下为影像医生提供初步的免疫治疗疗效预测提示，尤其适用于基层医院或资源有限环境中肺癌免疫治疗潜在获益人群的初筛。

(3) 为了进一步提高肺癌免疫治疗疗效预测的精度，提出了基于多帧影像信

息融合的疗效预测模型 MultiFrameNet。针对全局特征提取不足的问题，通过上述的 MaxFrameNet 分支提取肿瘤局部特征，同时利用 VMamba 分支挖掘全局特征关联，实现对局部-全局特征的深度解析。针对肿瘤尺度不确定性，引入跨尺度特征融合模块，充分融合全局与局部特征，进一步增强特征表达能力。为充分利用帧与之间的空间信息，设计了三维空间特征计算模块，精确捕捉肿瘤的三维空间分布特征，显著提高了模型的免疫疗效预测能力。该模型适用于肿瘤形态复杂、单一层面难以完整表征病灶信息的复杂病例，可以在有充足算力资源的场景下提升预测的准确性。

本文提出的肺部肿瘤分割方法和免疫疗效预测方法，在现有技术框架的基础上进行了有效的改进与创新，为肺癌免疫治疗疗效的早期评估与精确化治疗提供了有价值的理论依据和技术支持，为肺癌个体化免疫治疗提供科学、有效的智能化临床决策支持，并推动肺癌精准诊疗水平的提升。

5.2 工作展望

尽管本文提出的方法在肺部肿瘤分割与疗效预测中取得了一定的成果，但仍存在一些挑战和不足，未来的工作将集中在以下几个方面：

(1) 不同不确定性估计方法的系统对比与融合优化：本研究在分割算法中采用 MC-Dropout 方法实现不确定性估计，该方法部署简便、运算成本低，适配医学影像大规模训练，但属于近似贝叶斯推断，不确定性校准效果尚有不足。贝叶斯神经网络虽理论严谨、可解释性强，却存在结构复杂、算力消耗大的问题。后续将多维度对比各类估计算法，探究方法融合思路，进一步提升伪标签构建的精准度与可靠性。

(2) 更大规模的多中心数据集验证：本研究采用的免疫治疗疗效预测的数据集样本量相对有限，且均来源于单一医疗中心，在一定程度上可能影响模型的泛化性能。未来工作可进一步扩大数据集规模，并引入多中心数据开展外部验证，从而更全面、可靠地检验模型的鲁棒性与临床适用性。

(3) 与临床指标的联合建模：本研究仅基于 CT 影像特征构建预测模型，尚未纳入患者层面的多维临床信息。未来可进一步开展多模态融合建模，将肿瘤分割结果、影像组学特征与患者年龄、性别、吸烟史、肿瘤分期、病理类型等临床特征，以及基因突变、PD-L1 表达、肿瘤突变负荷等基因组学及免疫相关指标进行深度融合，构建影像-临床-基因的联合预测模型，从而更全面地刻画肿瘤异质性与患者个体特征，进一步提升免疫治疗疗效预测的精度与临床实用价值。

(4) 模型的可解释性与临床转化应用：尽管基于深度学习的预测模型在性能上展现出显著优势，但现有方法普遍存在“黑箱”特性，决策过程缺乏直观可解

释性，难以被临床医师理解与采信，在一定程度上制约了模型在真实临床场景中的落地应用。未来研究可重点围绕模型可解释性展开深入探索，通过可视化病灶关键区域、量化特征贡献度、构建可解释决策规则等方式，清晰揭示模型判断依据，增强临床医生对模型预测结果的认可度与信任度，进而推动智能预测模型逐步从科研验证走向临床辅助决策，实现真正意义上的临床转化与落地应用。

参考文献

- [1] Siegel R L, Miller K D, Goding Sauer A, et al. Colorectal cancer statistics, 2020 [J]. *CA: a cancer journal for clinicians*, 2020, 70(3): 145-164.
- [2] Guo L, Zhu C, Cai L, et al. Global burden of lung cancer in 2022 and projected burden in 2050 [J]. *Chinese Medical Journal*, 2024, 137(21): 2577-2582.
- [3] Liu L, Yang L, Li H, et al. The tumor microenvironment in lung cancer: Heterogeneity, therapeutic resistance and emerging treatment strategies [J]. *International Journal of Oncology*, 2025, 68(1): 11.
- [4] 周彩存, 王洁, 王宝成, 等. 中国非小细胞肺癌免疫检查点抑制剂治疗专家共识 (2020 年版) [J]. *中国肺癌杂志*, 2021, 24(4): 217-235.
- [5] Pu X, Zhou Y, Wang J, et al. Immune checkpoint inhibitor-based therapy as the first-line treatment for advanced non-small cell lung cancer: efficacy, challenges, and future perspectives [J]. *Thoracic Cancer*, 2025, 16(12): e70113.
- [6] Oh M S, Abascal J, Rennels A K, et al. Tumor heterogeneity and the immune response in non-small cell lung cancer: emerging insights and implications for immunotherapy [J]. *Cancers*, 2025, 17(6): 1027.
- [7] He Q, Guo X b, Chen Y r, et al. The progression patterns and subsequent treatments of first-line immunotherapy in advanced non-small cell lung cancer: A retrospective cohort study [J]. *Thoracic Cancer*, 2025, 16(20): e70173.
- [8] Akkurt E S, Duvenci Birben O, Deliktas Onur I, et al. Determinants of severity in immune checkpoint inhibitor-related pneumonitis: clinical, radiologic, and laboratory insights from real-life clinical data [J]. *BMC cancer*, 2026, 26(1): 243.
- [9] Mahumud R A, Chen Y, Dahal P K, et al. Economic value, affordability, and scale-up of adjuvant immunotherapies in lung cancer treatment: From cost-effectiveness decision to budget impact analysis [J]. *Journal of Cancer Policy*, 2026, 47: 100718.
- [10] 胡嘉捷, 石磊. 晚期非小细胞肺癌免疫治疗生物标志物及疗效预测的研究进展 [J]. *国际医学放射学杂志*, 2026, 49(2): 190-196.
- [11] Rakaee M, Tafavvoghi M, Ricciuti B, et al. Deep learning model for predicting immunotherapy response in advanced non-small cell lung cancer [J]. *JAMA oncology*, 2025, 11(2): 109-118.
- [12] Fomin V, So W V, Barbieri R A, et al. Machine learning identifies clinical tumor

- mutation landscape pathways of resistance to checkpoint inhibitor therapy in nsclc [J]. *Journal for Immunotherapy of Cancer*, 2025, 13(3): e009092.
- [13] Xia J, He T, Hu Y, et al. Development and validation of a circulating tumor dna-based machine learning model for predicting immunotherapy response in non-small cell lung cancer [J]. *Translational Cancer Research*, 2025, 14(12): 8778-8791.
- [14] Pan Y, Yang T, Xu T, et al. Construction and validation of a ct-based radiomics-deep learning signature for non-invasive prediction of pd-11 expression and immunotherapy outcomes in non-small cell lung cancer [J]. *Translational Lung Cancer Research*, 2026, 15(1): 18.
- [15] Zhu J, Shi H, Liang Z, et al. Enhanced ct-based deep learning radiomics and biological correlations for predicting immunotherapy efficacy in advanced non-small cell lung cancer [J]. *Translational Cancer Research*, 2026, 15(2): 81.
- [16] Yang J, Qu F, Wang Q, et al. Tumor rim-specific computed tomography radiomics improves prediction of pathological complete response to neoadjuvant chemoimmunotherapy in non-small cell lung cancer [J]. *Translational Lung Cancer Research*, 2025, 14(9): 3468-3485.
- [17] Shen L, Fu H, Tao G, et al. Pre-immunotherapy contrast-enhanced ct texture-based classification: a useful approach to non-small cell lung cancer immunotherapy efficacy prediction [J]. *Frontiers in Oncology*, 2021, 11: 591106.
- [18] Benzekry S, Grangeon M, Karlsen M, et al. Machine learning for prediction of immunotherapy efficacy in non-small cell lung cancer from simple clinical and biological data [J]. *Cancers*, 2021, 13(24): 6210.
- [19] Wei F, Azuma K, Nakahara Y, et al. Machine learning for prediction of immunotherapeutic outcome in non-small-cell lung cancer based on circulating cytokine signatures [J]. *Journal for immunotherapy of cancer*, 2023, 11(7): e006788.
- [20] Li Z, Pei S, Wang Y, et al. Advancing predictive markers in lung adenocarcinoma: A machine learning-based immunotherapy prognostic prediction signature [J]. *Environmental Toxicology*, 2024, 39(10): 4581-4593.
- [21] Liu Z, Lin G, Yan Z, et al. Predictive mutation signature of immunotherapy benefits in nsclc based on machine learning algorithms [J]. *Frontiers in Immunology*, 2022, 13: 989275.
- [22] Li B, Yang L, Zhang H, et al. Outcome-supervised deep learning on pathologic whole slide images for survival prediction of immunotherapy in patients with non-

- small cell lung cancer [J]. *Modern Pathology*, 2023, 36(8): 100208.
- [23] Yang Y, Yang J, Liang Y, et al. Identification and validation of efficacy of immunological therapy for lung cancer from histopathological images based on deep learning [J]. *Frontiers in Genetics*, 2021, 12: 642981.
- [24] Patkar S, Chen A, Basnet A, et al. Predicting the tumor microenvironment composition and immunotherapy response in non-small cell lung cancer from digital histopathology images [J]. *NPJ Precision Oncology*, 2024, 8(1): 280.
- [25] Wang Z, Liu X, Han K, et al. Multimodal deep learning for immunotherapy response prediction and biomarker discovery in non-small cell lung cancer [J]. *Journal of the American Medical Informatics Association*, 2025, 32(11): 1641-1653.
- [26] Yang Y, Yang J, Shen L, et al. A multi-omics-based serial deep learning approach to predict clinical outcomes of single-agent anti-pd-1/pd-l1 immunotherapy in advanced stage non-small-cell lung cancer [J]. *American journal of translational research*, 2021, 13(2): 743.
- [27] He B X, Zhong Y F, Zhu Y B, et al. Deep learning for predicting immunotherapeutic efficacy in advanced non-small cell lung cancer patients: a retrospective study combining progression-free survival risk and overall survival risk [J]. *Translational Lung Cancer Research*, 2022, 11(4): 670.
- [28] Liao C Y, Chen Y M, Wu Y T, et al. Personalized prediction of immunotherapy response in lung cancer patients using advanced radiomics and deep learning [J]. *Cancer Imaging*, 2024, 24(1): 129.
- [29] Chiu H Y, Wang T W, Hsu M S, et al. Progress in serial imaging for prognostic stratification of lung cancer patients receiving immunotherapy: a systematic review and meta-analysis [J]. *Cancers*, 2024, 16(3): 615.
- [30] Saad M B, Hong L, Aminu M, et al. Predicting benefit from immune checkpoint inhibitors in patients with non-small-cell lung cancer by ct-based ensemble deep learning: a retrospective study [J]. *The Lancet Digital Health*, 2023, 5(7): e404-e420.
- [31] Masson-Grehaigne C, Lafon M, Palussière J, et al. Enhancing immunotherapy response prediction in metastatic lung adenocarcinoma: Leveraging shallow and deep learning with ct-based radiomics across single and multiple tumor sites [J]. *Cancers*, 2024, 16(13): 2491.
- [32] Kanakarajan H, Zhou J, Gomes A L, et al. Predicting overall survival of nslc patients with clinical, radiomics and deep learning features [J]. *Journal of Imaging*

- Informatics in Medicine, 2026, 39: 1-18.
- [33] She Y, He B, Wang F, et al. Deep learning for predicting major pathological response to neoadjuvant chemoimmunotherapy in non-small cell lung cancer: A multicentre study [J]. *EBioMedicine*, 2022, 86: 104364.
- [34] Wu Q, Wang J, Sun Z, et al. Immunotherapy efficacy prediction for non-small cell lung cancer using multi-view adaptive weighted graph convolutional networks [J]. *IEEE journal of biomedical and health informatics*, 2023, 27(11): 5564-5575.
- [35] Yan X, Pang J, Qi H, et al. Classification of lung nodule malignancy risk on computed tomography images using convolutional neural network: A comparison between 2d and 3d strategies [C]//*Asian Conference on Computer Vision*. Springer, 2016: 91-101.
- [36] Xu Y, Hosny A, Zeleznik R, et al. Deep learning predicts lung cancer treatment response from serial medical imaging [J]. *Clinical cancer research*, 2019, 25(11): 3266-3275.
- [37] He B, Dong D, She Y, et al. Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker [J]. *Journal for immunotherapy of cancer*, 2020, 8(2): e000550.
- [38] Xu H, Li C, Zhang L, et al. Immunotherapy efficacy prediction through a feature re-calibrated 2.5 d neural network [J]. *Computer Methods and Programs in Biomedicine*, 2024, 249: 108135.
- [39] Marentakis P, Karaiskos P, Kouloulis V, et al. Lung cancer histology classification from ct images based on radiomics and deep learning models [J]. *Medical & biological engineering & computing*, 2021, 59(1): 215-226.
- [40] Yan K, Bagheri M, Summers R M. 3d context enhanced region-based convolutional neural network for end-to-end lesion detection [C]//*International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018: 511-519.
- [41] Dou Q, Chen H, Jin Y, et al. Automated pulmonary nodule detection via 3d convnets with online sample filtering and hybrid-loss residual learning [C]//*International conference on medical image computing and computer-assisted intervention*. Springer, 2017: 630-638.
- [42] Zhu W, Liu C, Fan W, et al. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification [C]//*2018 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2018: 673-681.

- [43] Tekade R, Rajeswari K. Lung cancer detection and classification using deep learning [C]//2018 fourth international conference on computing communication control and automation (ICCUBEA). IEEE, 2018: 1-5.
- [44] Guo Y, Song Q, Jiang M, et al. Histological subtypes classification of lung cancers on ct images using 3d deep learning and radiomics [J]. *Academic radiology*, 2021, 28(9): e258-e266.
- [45] Zhang Q, Wang Y, Qiu S, et al. 3d-pulcnn: Pulmonary cancer classification from hyperspectral images using convolution combination unit based cnn [J]. *Journal of Biophotonics*, 2021, 14(12): e202100142.
- [46] 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. Springer, 2015: 234-241.
- [47] Agnes S A, Solomon A A, Karthick K. Wavelet u-net++ for accurate lung nodule segmentation in ct scans: Improving early detection and diagnosis of lung cancer [J]. *Biomedical Signal Processing and Control*, 2024, 87: 105509.
- [48] Wu Z, Li X, Zuo J. Rad-unet: Research on an improved lung nodule semantic segmentation algorithm based on deep learning [J]. *Frontiers in oncology*, 2023, 13: 1084096.
- [49] Zhang X, Kong S, Han Y, et al. Lung nodule ct image segmentation model based on multiscale dense residual neural network [J]. *Mathematics*, 2023, 11(6): 1363.
- [50] Hou J, Yan C, Li R, et al. Lung nodule segmentation algorithm with smr-unet [J]. *Ieee Access*, 2023, 11: 34319-34331.
- [51] Zhao J, Dang M, Chen Z, et al. Dsu-net: Distraction-sensitive u-net for 3d lung tumor segmentation [J]. *Engineering Applications of Artificial Intelligence*, 2022, 109: 104649.
- [52] Banu S F, Sarker M M K, Abdel-Nasser M, et al. Aweu-net: an attention-aware weight excitation u-net for lung nodule segmentation [J]. *Applied Sciences*, 2021, 11(21): 10132.
- [53] Ming R, Wang F, Zheng T, et al. Glance: continuous global-local exchange with consensus fusion for robust nodule segmentation [J]. *npj Digital Medicine*, 2026, 9(1): 69.
- [54] Yan W, Xu Y, Yan S. Promptable segmentation of ct lung lesions based on improved u-net and segment anything model (sam) [J]. *Journal of X-Ray Science and Technology*, 2025, 33(6): 1015-1026.

- [55] Zhang X, Wang J, Wei J, et al. A review of non-fully supervised deep learning for medical image segmentation [J]. *Information*, 2025, 16(6): 433.
- [56] Jin D, Xu Z, Tang Y, et al. Ct-realistic lung nodule simulation from 3d conditional generative adversarial networks for robust lung segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018: 732-740.
- [57] Qin Y, Zheng H, Huang X, et al. Pulmonary nodule segmentation with ct sample synthesis using adversarial networks [J]. *Medical physics*, 2019, 46(3): 1218-1229.
- [58] Liu W, Zhang L, Li X, et al. A semisupervised knowledge distillation model for lung nodule segmentation [J]. *Scientific Reports*, 2025, 15(1): 10562.
- [59] Fredriksen V, Sevre S O M, Pedersen A, et al. Teacher-student approach for lung tumor segmentation from mixed-supervised datasets [J]. *Plos one*, 2022, 17(4): e0266147.
- [60] Lu S, Yan Z, Chen W, et al. Dual consistency regularization with subjective logic for semi-supervised medical image segmentation [J]. *Computers in Biology and Medicine*, 2024, 170: 107991.
- [61] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results [C]//Advances in Neural Information Processing Systems. 2017: 1195-1204.
- [62] Wang W, Li Z, Zhang X, et al. Pe-mt: A perturbation-enhanced mean teacher for semi-supervised image segmentation [J]. *Bioengineering*, 2025, 12(5): 453.
- [63] Su J, Sun W, Adamyk B. Dsmt-net: Dual-student mean teacher network with pixel-level pseudo-label optimization for semi-supervised medical image segmentation [J]. *Computational Biology and Chemistry*, 2025, 119: 108579.
- [64] Xi Y, Xu C, Ye F, et al. Coreformer high fidelity pulmonary nodule segmentation with structural core priors and geodesic implicit fields [J]. *npj Digital Medicine*, 2026, 9(1): 48.
- [65] Liu H, Ren P, Yuan Y, et al. Uncertainty global contrastive learning framework for semi-supervised medical image segmentation [J]. *IEEE Journal of Biomedical and Health Informatics*, 2024, 29(1): 433-442.
- [66] Lu W, Lei J, Qiu P, et al. Upcol: Uncertainty-informed prototype consistency learning for semi-supervised medical image segmentation [C]//International conference on medical image computing and computer-assisted intervention.

- Springer, 2023: 662-672.
- [67] Yu L, Wang S, Li X, et al. Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation [C]//International conference on medical image computing and computer-assisted intervention. Springer, 2019: 605-613.
- [68] Luo X, Chen J, Song T, et al. Semi-supervised medical image segmentation through dual-task consistency [C]//Proceedings of the AAAI conference on artificial intelligence: volume 35. 2021: 8801-8809.
- [69] Bai Y, Chen D, Li Q, et al. Bidirectional copy-paste for semi-supervised medical image segmentation [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 11514-11524.
- [70] Wu S, Wei X, Chen X, et al. Cross-view mutual learning for semi-supervised medical image segmentation [C]//Proceedings of the 32nd ACM International Conference on Multimedia. 2024: 9253-9261.
- [71] Assefa M, Naseer M, Ganapathi I I, et al. Dycon: Dynamic uncertainty-aware consistency and contrastive learning for semi-supervised medical image segmentation [C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 30850-30860.
- [72] Liu Y, Wu M, Zhang Y, et al. Imaging biomarkers to predict and evaluate the effectiveness of immunotherapy in advanced non-small-cell lung cancer [J]. *Frontiers in oncology*, 2021, 11: 657615.
- [73] Wu M, Zhang Y, Zhang J, et al. A combined-radiomics approach of ct images to predict response to anti-pd-1 immunotherapy in nsclc: a retrospective multicenter study [J]. *Frontiers in Oncology*, 2022, 11: 688679.
- [74] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [75] Armato III S G, McLennan G, Bidaut L, et al. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans [J]. *Medical physics*, 2011, 38(2): 915-931.
- [76] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
- [77] Liu Z, Mao H, Wu C Y, et al. A convnet for the 2020s [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 11976-

- 11986.
- [78] Manzari O N, Ahmadabadi H, Kashiani H, et al. Medvit: a robust vision transformer for generalized medical image classification [J]. *Computers in biology and medicine*, 2023, 157: 106791.
- [79] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 2818-2826.
- [80] Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3 [C]//*Proceedings of the IEEE/CVF international conference on computer vision*. 2019: 1314-1324.
- [81] Yue Y, Li Z. Medmamba: Vision mamba for medical image classification [J]. *arXiv preprint arXiv:2403.03849*, 2024.
- [82] Ma X, Dai X, Bai Y, et al. Rewrite the stars [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2024: 5694-5703.
- [83] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]//*International conference on machine learning*. PMLR, 2019: 6105-6114.
- [84] Lou M, Yu Y. Overlock: An overview-first-look-closely-next convnet with context-mixing dynamic kernels [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2025: 128-138.
- [85] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017: 4700-4708.
- [86] Liu Y, Tian Y, Zhao Y, et al. Vmamba: Visual state space model [C]//*Advances in Neural Information Processing Systems*. 2024: 103031-103063.
- [87] Park J, Lee J, Sohn K. Dual-path adaptation from image to video transformers [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: 2203-2213.
- [88] Hara K, Kataoka H, Satoh Y. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? [C]//*Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2018: 6546-6555.
- [89] Liu Z, Ning J, Cao Y, et al. Video swin transformer [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022: 3202-3211.
- [90] Yan Y, Xu J, Di S, et al. Learning streaming video representation via multitask

- training [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2025: 9900-9912.
- [91] Li K, Wang Y, Zhang J, et al. Uniformer: Unifying convolution and self-attention for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(10): 12581-12600.
- [92] Huang Z, Zhang S, Pan L, et al. Tada! temporally-adaptive convolutions for video understanding [C]//International Conference on Learning Representations. 2022.

攻读学位期间主要的研究成果

一、学术论文

[1] Xia Z. Convolutional Neural Networks Combine with VMamba for Predicting Immunotherapy Efficacy in Lung Cancer based on CT sequences[C]//Proceedings of the 2025 6th International Symposium on Artificial Intelligence for Medical Sciences. 2025: 413-417.

二、主持和参与的科研项目

- (1) 企、事业单位委托项目：肺癌免疫治疗疗效预测，2024.5-至今，参与.
- (2) 企、事业单位委托项目：基于智能感知的交通资产管理与监控视频分析系统，2023.8-2024.8，参与.

三、个人获奖情况

- (1) 国家级，中国研究生数学建模竞赛，国家级三等奖，2024.
- (2) 国家级，MathorCup 数学应用挑战赛，国家级三等奖，2024.
- (3) 校级，中南大学校级一等奖学金，2023.
- (4) 校级，中南大学校级二等奖学金，2024.

致 谢

时光荏苒，行文至此，本论文的研究工作即将画上句点，数年的求学之路也将落下帷幕。回首漫漫求学岁月，一路走来，有探索科研的迷茫与艰辛，亦有收获与成长。借此机会，谨向所有给予我指导、帮助与陪伴的师长、同窗与家人，致以最诚挚的谢意。

读研三载，幸遇良师。由衷感谢我的导师谢斌教授。从研究方向的敲定、实验方案的设计，到论文的构思、修改，每一个关键环节都离不开导师的悉心指导与耐心点拨。科研道路上，导师严谨务实的治学态度、精益求精的科研作风、开阔前沿的学术视野，时刻指引着我脚踏实地、潜心钻研。生活中，导师温和宽厚、待人谦和，给予我诸多关怀与包容，言传身教的处世之道，将使我受益终身。

同窗之谊，历久弥新。感谢实验室朝夕相伴的各位同门与挚友。感谢浩然师兄提供的无私帮助与指引，慷慨分享论文与方法；感谢海涛师兄耐心解答科研中遇到的诸多困惑。感谢德坤、硕清、思航，我们并肩作战，共同跨越了求学路上的重重关卡。感谢实验室的其他伙伴们，我们一起攻克 PPT 难题，一起应对科研中的各种挑战。科研路上，我们相互探讨、彼此交流，在代码调试、实验分析、论文撰写的过程中互帮互助；生活之中，我们并肩同行、彼此慰藉，一同消解科研压力，见证彼此的成长与进步。朝夕相处的温暖时光，让枯燥的科研生活变得充实而鲜活，这段珍贵的同窗情谊，是我求学路上最宝贵的财富。

亲恩似海，润物无声。感谢我的家人。多年来，你们始终是最坚实的后盾与最温暖的港湾。无条件的理解、默默无声的付出、毫无保留的包容与支持，让我能够心无旁骛地投入学习，勇敢面对求学途中的所有挑战。你们的牵挂与期许，是我不断前行、奋勇拼搏的底气与动力。

山水一程，感恩相遇。岁月绵长，步履不停。前路漫漫，我将带着求学期间的收获与感动，心怀感恩、脚踏实地，保持热爱与初心，在未来的人生道路上继续砥砺前行，奔赴下一场山海。

2026年5月20日

于湖南长沙