

硕士学位论文

基于 CT 影像的嗜铬细胞瘤基因亚型识别
研究

Research on the Identification of Pheochromocytoma
Gene Subtypes Based on CT Images

专业学位类别

电子信息

专业领域

人工智能

作者姓名

马浩然

指导教师

谢斌 副教授

2025 年 5 月

中图分类号 TP391.4

UDC 004.9

学校代码 10533

学位类别 专业学位

硕士学位论文

基于 CT 影像的嗜铬细胞瘤基因亚型识别研究

Research on the Identification of Pheochromocytoma Gene Subtypes Based on CT Images

作者姓名	马浩然
专业学位类别	电子信息
专业领域	人工智能
研究方向	智慧医疗
二级培养单位	自动化学院
指导教师	谢斌 副教授

论文答辩日期 2025年5月24日

答辩委员会主席 王

中南大学

2025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：马浩然 日期：2025年5月24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：马浩然

日期：2025年5月24日

（填写阿拉伯数字）

指导教师签名：谭斌

日期：2025年5月24日

基于 CT 影像的嗜铬细胞瘤基因亚型识别研究

摘要：嗜铬细胞瘤及其副神经节瘤（PPGL）是罕见神经内分泌肿瘤，其基因亚型与肿瘤恶性风险、治疗方案及患者预后密切相关。研究显示缺氧通路激活基因突变驱动的病例转移风险较高，其余病例预后相对较好。传统基因检测因创伤性高、成本昂贵等问题难以普及，临床迫切需要基于 CT 影像的无创识别方法。本研究提出双路径研究范式，并行构建基于肿瘤最大截面积的单帧模型与基于多帧 CT 的序列模型，系统比较两类深度学习框架在基因亚型识别中的效能差异，为临床无创分型提供多维度技术选择。本文主要研究内容包括：

（1）构建 PPGL 影像-基因关联数据集。本研究以湘雅医院临床数据为核心，构建包含 425 例患者样本的湘雅嗜铬细胞瘤影像基因数据集（Xiangya-PPGL）。目前 PPGL 领域尚无公开数据集支持基因亚型相关研究，本研究构建的数据集涵盖高质量 CT 影像及完整基因亚型信息，为后续研究提供有力支持。数据分析表明不同患者肿瘤区域大小差异显著，多数患者肿瘤区域在图像中占比较低，这些发现为模型优化提供关键依据。

（2）基于单帧影像的 PPGL 基因亚型识别。本研究提出了一种基于最大肿瘤区域帧的分类模型（TAMax-Net）。针对 PPGL 患者间肿瘤尺寸变化较大以及切片图像中肿瘤占比低的挑战，TAMax-Net 设计了局部细节优化模块，通过多层次特征提取和细粒度调整，增强网络对局部细节的感知能力和多尺度特征融合能力。其次，研究引入了跨层级特征融合增强模块，融合不同分辨率的特征图，以减少低肿瘤占比图像在特征传递过程中的信息损失。最后，因细节信息与语义信息混合会导致模型难以聚焦关键特征，在模型末端设计末端注意力模块聚焦重要特征。实验表明，TAMax-Net 在 Xiangya-PPGL 数据集上的受试者工作特征曲线下面积（Area Under the ROC Curve, AUC）为 0.9318，准确率为 0.8831，特异性为 0.8067，敏感度为 0.9325，显著优于现有深度学习方法和人类医生，为基于单帧影像的 PPGL 基因亚型识别提供了可靠的方案。

（3）基于多帧融合的 PPGL 基因亚型识别。针对单帧方法难以充分利用空间信息、存在帧序列冗余及肿瘤整体形态特征缺失的问题，

本研究开发了 MFF-Net 模型。该模型的特征聚焦模块能自动分析影像，减少背景干扰，借助 ROI-Align 技术稳定提取肿瘤特征，更好保留了影像细节。空间特征提取模块则通过学习肿瘤形态和分布规律，弥补单帧分析的不足。为实现多帧特征深度融合，序列感知多帧融合模块能整合各帧信息，生成更具辨识度的肿瘤特征。实验结果表明，MFF-Net 在 Xiangya-PPGL 数据集上，受试者工作特征曲线下面积为 0.959，准确率为 0.9267，特异性为 0.825，敏感度为 0.9583，相比同类多帧分类方法优势明显，为多帧影像识别 PPGL 基因亚型提供了可靠途径。

本研究采用双路径深度学习范式，构建专属数据集并设计特定模块优化模型，解决肿瘤尺寸差异大、信息利用难的问题，相比传统方法大幅提升模型性能，为临床提供无创高效的基因亚型识别手段，推动罕见病个性化诊疗从理论走向实际应用。

图23幅，表13个，参考文献74篇

关键词：嗜铬细胞瘤及其副神经节瘤；放射基因组学；基因亚型识别；CT影像序列

分类号： TP391.4

Research on the Identification of Pheochromocytoma Gene Subtypes Based on CT Imaging

Abstract: Pheochromocytoma and its paraganglioma (PPGL) are rare neuroendocrine tumors, and their genetic subtypes are closely related to the risk of tumor malignancy, treatment options, and patient prognosis. Studies have shown that cases driven by hypoxia pathway activation gene mutations have a higher risk of metastasis, while the prognosis of other cases is relatively good. Traditional genetic testing is difficult to popularize due to its high trauma and high cost. Clinically, non-invasive identification methods based on CT images are urgently needed. This study proposes a dual-pathway research paradigm, and constructs a single-frame model based on the maximum cross-sectional area of the tumor and a sequence model based on multi-frame CT in parallel. The performance differences of the two types of deep learning frameworks in genetic subtype identification are systematically compared, providing multi-dimensional technical options for clinical non-invasive classification. The main research contents of this article include:

(1) Construction of PPGL image-gene association dataset. This study is based on the clinical data of Xiangya Hospital and constructs the Xiangya pheochromocytoma image-gene dataset (Xiangya-PPGL) containing 425 patient samples. At present, there is no public dataset in the PPGL field to support genetic subtype-related research. The dataset constructed in this study covers high-quality CT images and complete genetic subtype information, providing strong support for subsequent research. Data analysis showed that the size of tumor regions varied significantly among patients, and the tumor regions of most patients accounted for a low proportion of the image. These findings provide a key basis for model optimization.

(2) Identification of PPGL gene subtypes based on single-frame images. This study proposed a classification model based on the maximum tumor region frame (TAMax-Net). In response to the challenges of large tumor size variations among PPGL patients and low tumor proportion in slice images, TAMax-Net designed a local detail optimization module to enhance the network's perception of local details and multi-scale feature fusion capabilities through multi-level feature extraction and fine-grained adjustment. Secondly, the study introduced a cross-level feature fusion

enhancement module to fuse feature maps of different resolutions to reduce information loss in the feature transmission process of low tumor proportion images. Finally, because the mixture of detail information and semantic information makes it difficult for the model to focus on key features, an end-point attention module is designed at the end of the model to focus on important features. Experiments show that the area under the ROC curve (AUC) of TAMax-Net on the Xiangya-PPGL dataset is 0.9318, the accuracy is 0.8831, the specificity is 0.8067, and the sensitivity is 0.9325, which is significantly better than the existing deep learning methods and human doctors, and provides a reliable solution for the identification of PPGL gene subtypes based on single-frame images.

(3) PPGL gene subtype identification based on multi-frame fusion. In view of the problems that the single-frame method is difficult to fully utilize spatial information, there is frame sequence redundancy, and the overall morphological characteristics of the tumor are missing, this study developed the MFF-Net model. The feature focusing module of this model can automatically analyze the image, reduce background interference, and stably extract tumor features with the help of ROI-Align technology, better retaining image details. The spatial feature extraction module makes up for the shortcomings of single-frame analysis by learning tumor morphology and distribution patterns. To achieve deep fusion of multi-frame features, the sequence-aware multi-frame fusion module can integrate the information of each frame to generate more recognizable tumor features. The experimental results show that the area under the receiver operating characteristic curve of MFF-Net on the Xiangya-PPGL dataset is 0.959, the accuracy is 0.9267, the specificity is 0.825, and the sensitivity is 0.9583. Compared with similar multi-frame classification methods, it has obvious advantages and provides a reliable way to identify PPGL gene subtypes in multi-frame images.

This study adopts a dual-path deep learning paradigm, constructs a dedicated dataset and designs a specific module optimization model to solve the problem of large tumor size differences and difficult information utilization. Compared with traditional methods, the model performance is greatly improved, providing a non-invasive and efficient means of identifying gene subtypes for clinical use, and promoting personalized diagnosis and treatment of rare diseases from theory to practical application.

Keywords: Pheochromocytoma and Paraganglioma; Radiogenomics;
Genetic Subtype Identification; CT Image Sequence

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.....	3
1.2.1 基于传统影像组学方法识别肿瘤基因亚型的现状.....	3
1.2.2 基于深度学习方法识别肿瘤基因亚型的现状.....	4
1.3 研究目的与内容.....	6
1.4 本文组织结构.....	8
第 2 章 数据集建立与模型评估.....	9
2.1 计算机断层扫描技术.....	9
2.1.1 CT 技术的基本原理.....	9
2.1.2 常见的 CT 成像方式及其特点.....	9
2.1.3 NECT 在 PPGL 研究中的应用价值.....	10
2.2 湘雅嗜铬细胞瘤影像基因数据集 Xiangya-PPGL.....	10
2.2.1 数据集的获取.....	11
2.2.2 数据集的预处理.....	12
2.2.3 数据集的分析.....	15
2.3 模型测试框架与性能度量.....	17
2.3.1 损失函数.....	17
2.3.2 评价指标.....	17
2.3.3 实验配置.....	18
2.4 本章小结.....	19
第 3 章 TAMax-Net: 基于肿瘤区域面积最大帧的分类模型.....	20
3.1 概述.....	20
3.1.1 总体思路.....	20
3.1.2 模型总体结构.....	21
3.2 模型主要模块.....	22
3.2.1 EfficientNet 介绍.....	22
3.2.2 局部细节优化模块.....	23
3.2.3 跨层级特征融合增强模块.....	25

3.2.4 末端注意力模块	29
3.3 实验结果与分析	31
3.3.1 实验配置	31
3.3.2 对比实验	31
3.3.3 消融实验	34
3.3.4 可视化分析	37
3.4 本章小结	38
第 4 章 MFNet: 基于肿瘤多帧融合网络的分类模型	39
4.1 概述	39
4.1.1 总体思路	39
4.1.2 模型总体结构	40
4.2 模型主要模块	41
4.2.1 特征聚焦模块	41
4.2.2 空间特征提取模块	43
4.2.3 序列感知多帧融合模块	45
4.3 实验结果与分析	47
4.3.1 实验配置	47
4.3.2 对比实验	47
4.3.3 消融实验	49
4.4 本章小结	51
第 5 章 总结与展望	52
5.1 论文总结	52
5.2 工作展望	53
5.2.1 研究局限性	53
5.2.2 未来研究方向	53
参考文献	55
攻读学位期间主要的研究成果	62
致 谢	63

第1章 绪论

嗜铬细胞瘤及其副神经节瘤（Pheochromocytoma and Paraganglioma, PPGL）是发病率 0.46/10 万的罕见神经内分泌肿瘤^[1]，其基因亚型关乎肿瘤恶性风险、治疗方案选择和患者预后。研究显示 SDHB 突变型患者恶性转化风险达 50%，VHL 突变型患者可伴发肾细胞癌、胰腺神经内分泌肿瘤等多器官病变。临床依靠基因检测确定患者基因亚型，但 RNA 测序和穿刺活检等常用方法存在应用难题。穿刺活检有创会给患者带来痛苦和风险，RNA 测序成本高、技术设备要求严，在基层医院难以普及。这些因素制约了技术推广。

本研究联合湘雅医院等多家医疗机构收集了 425 例患者数据，通过两条并行技术路线探索 CT 影像在 PPGL 基因分型中的应用：（1）基于肿瘤区域面积最大帧构建单帧分类模型，验证二维图像特征对基因亚型的判别能力；（2）设计基于 CT 序列的多帧分类模型，评估空间信息对性能的提升效果，研究成果为精准医疗提供了新技术支持。

本章将系统阐述嗜铬细胞瘤基因亚型识别的研究背景与意义，综述肿瘤基因亚型识别相关研究现状与进展，说明研究内容及各章节组织结构。

1.1 研究背景与意义

PPGL 是起源于神经内分泌系统的罕见肿瘤，其中嗜铬细胞瘤起源于肾上腺髓质嗜铬细胞，副神经节瘤起源于肾上腺外交感或副交感神经节嗜铬细胞^[2]，如图 1-1 所示。其核心病理特征为嗜铬细胞异常增生，这些细胞持续或间歇性的释放大量的儿茶酚胺进入血液循环，影响心血管系统及多器官功能，会导致血压剧烈波动、代谢紊乱等典型临床表现，严重时可引发高血压危象、脑出血或急性心肌梗死，危及患者生命。

世界卫生组织（World Health Organization, WHO）指出 PPGL 是高度异质性疾病^[3]，异质性由复杂的遗传和表观遗传事件驱动。研究显示约 30%-35% 的 PPGL 病例存在种系突变，40% 有易感基因体细胞突变^[4]。依据分子生物学特征，PPGL 分为由缺氧通路激活基因突变驱动的真性缺氧簇（Cluster 1）和非真性缺氧簇（Non-Cluster 1）。Cluster 1 中 SDHB 突变型患者因缺氧诱导因子 2 α （HIF-2 α ）异常激活及基因组高甲基化，转移风险高^[5-6]，需更积极治疗策略。转移性 PPGL 因缺乏有效治疗死亡率高，转移患者 5 年总生存率 12%-92%^[7]，无转移患者为 89.3%^[8]。因此基因突变检测与基因亚型确定对 PPGL 患者个体化治疗和预后评估至关重要^[9]。

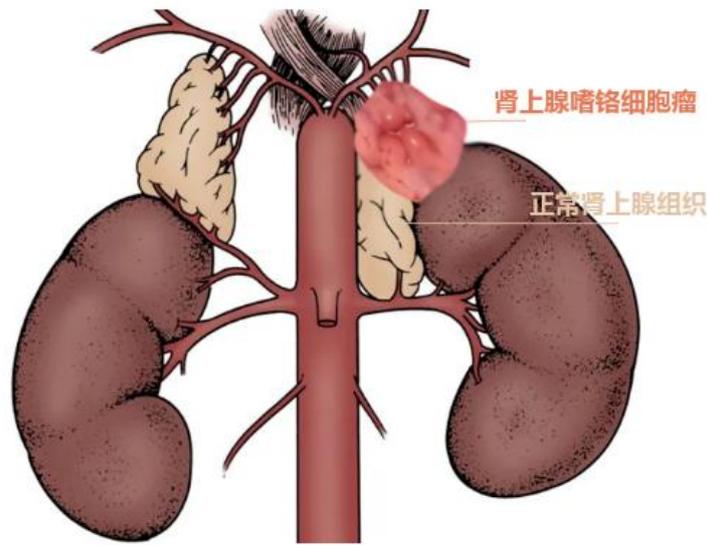


图 1-1 嗜铬细胞瘤示意图

临床上 PPGL 患者基因突变检测主要依靠下一代测序技术 (Next-Generation Sequencing, NGS) [10], 检测术前外周血样本和术后肿瘤组织, 但该方法存在局限: 体细胞突变可能仅存于肿瘤细胞外周血检测难以发现; 肿瘤高度异质性使单一组织样本无法完整呈现基因特征; NGS 处理大片段基因组缺失或未知显著变体时敏感度不足[11]; 肿瘤位置特殊或患者身体条件差时术前组织取样易引发严重并发症, 这些因素制约了其临床应用, 因而无创精准的基因亚型识别方法成为 PPGL 临床诊疗迫切需求。

深度学习技术发展推动医学影像分析变革, 基于医学影像无创识别肿瘤基因亚型成为肿瘤精准诊疗研究重点。CT 影像作为 PPGL 常规诊断方式, 既能提供解剖学信息也能反映潜在病理特征。如图 1-2 所示, 某 PPGL 患者腹部横断面 CT 影像可辅助医生评估肿瘤大小、位置及脏器血管状况以制定诊疗方案。虽然已有研究证实 CT 成像可识别 PPGL 异质性并辅助临床 PPGL 的诊断[12-15], 但其与基因亚型的联系尚未深入探究。

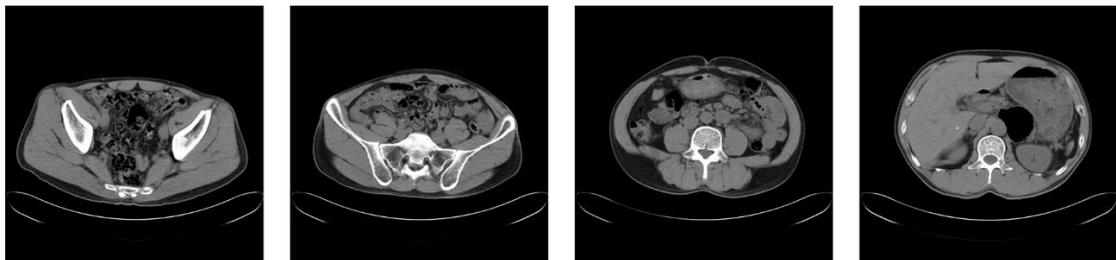


图 1-2 某 PPGL 患者的 CT 切层图像

本研究联合湘雅医院, 运用深度学习技术对患者肿瘤区域的非增强 CT

(Non-Enhanced CT, NECT) 影像进行特征提取与智能分析, 实现 PPGL 基因亚型术前无创识别, 致力于为临床提供无创、低成本且高效的基因亚型识别工具, 优化 PPGL 患者个体化诊疗策略。

1.2 国内外研究现状

PPGL 作为罕见病, 其基因亚型识别研究尚处于初步阶段, 尚未形成系统性研究。本文将对医学肿瘤领域基因亚型识别相关研究进行综述探讨。目前肿瘤基因亚型临床检测多依靠活检、手术或免疫组织化学等传统方法, 不过这些方法存在明显缺陷。比如活检属于侵入性操作, 在老年或体质虚弱患者中应用受限, 更关键的是受肿瘤组织异质性的影响, 病理穿刺活检可能无法准确采集肿瘤细胞, 造成诊断失误。

针对这些问题, 近年来研究者开始关注传统影像组学与深度学习技术。这种新方法通过定量分析和提取医学图像特征, 为肿瘤特征识别和临床预后预测提供非侵入性方案, 能为患者提供安全高效的识别手段, 且研究显示其在辅助临床决策和预后预测方面优势显著^[16-18]。基于此, 本节将从传统影像组学和深度学习两条技术路径出发, 系统综述相关研究进展。

1.2.1 基于传统影像组学方法识别肿瘤基因亚型的现状

传统影像组学运用统计学和机器学习方法定量分析医学影像, 为临床诊疗提供支持, 不仅能量化评估肿瘤形状、密度、纹理等多维度特征, 还能挖掘影像中潜在生物学与疾病特征。相较于传统影像诊断, 肿瘤放射组学借助高通量计算^[19], 从标准医学成像提取大量定量图像特征, 获取人眼无法识别的肿瘤异质性信息^[20], 既提升诊断客观性, 也为疾病机制研究、疗效评估和预后预测开拓新视角^[21]。

在影像组学的临床应用方面, Zhu 等人^[22]的代表性研究从 129 名患者 CT 影像手动勾勒肿瘤区域, 提取强度、形状、纹理和小波特征等 485 个特征, 经 LASSO 逻辑回归模型筛选出 5 个关键特征构建 Rad-score 评分系统, 该系统在验证队列中区分肺腺癌和鳞状细胞癌的 AUC 值达 0.905, 证实影像组学特征可作为非侵入性诊断工具辅助临床治疗决策。

Zhou 等人^[23]在 PPGL 风险分层研究中纳入 201 例患者数据, 将数据集分为训练集 (125 例)、外部验证集 (45 例) 和外部测试集 (31 例), 利用肾上腺嗜铬细胞瘤和副神经节瘤分期系统将患者分为低危组和高危组, 通过支持向量机 (SVM)、k 近邻、随机森林和多层感知器等模型对 CT 影像提取的 13 个放射组学特征进行分类, 发现 SVM 模型表现最佳 (训练集、外部验证集、外部测试集 AUC 分别为 0.915、0.846、0.857), 表明 CT 特征可辅助 PPGL 风险分层。

Yang 等人^[24]将放射影像组学应用于非小细胞肺癌诊断, 从分割后的肿瘤区

域提取 788 个放射组学特征，通过最小冗余最大相关性特征选择、序贯前向选择和 LASSO 等方法筛选特征，最终采用逻辑回归、支持向量机和随机森林分类器评估，为肺肿瘤亚型分类提供了新方法学参考。

Chen 等人^[25]纳入 60 例无功能肾上腺皮质腺瘤和 91 例醛固酮腺瘤患者，按 8:2 比例随机分为训练集和测试集，结果显示基于 CT 影像的影像组学模型可有效区分两类腺瘤(诊断敏感度 81.2%、特异性 100%、准确率 87.5%，AUC 为 0.98)，该发现与 Maggio 等人^[26]的研究相互印证，后者利用 CT 影像提取的 314 个纹理特征建立的预测模型诊断醛固酮腺瘤的 AUC 达 0.982、敏感度为 93.5%、特异性为 100%。

Krishan 等人^[27]提出基于 CT 图像的多类肝癌疾病分类方法为影像组学应用提供新思路，该方法采用对比受限自适应直方图均衡化算法增强 CT 图像，从增强后感兴趣区域提取包括一阶统计量、灰度差异矩阵和灰度共生矩阵在内的纹理特征，经遗传算法自动选择特征后使用支持向量机分类器区分原发性肝癌(HCC)和继发性肝癌(MET)。实验显示此方法在多类肿瘤分类任务中准确率达 97%、敏感性为 94.3%、特异性为 100%，且自动特征选择减少特征数量后分类准确率仅下降 10.5%，为肝癌精准诊疗提供技术支撑。

Lin 等人^[28]对比研究发现，样本量充足时 CNN 自动提取特征比传统影像组学手工提取特征性能更优，数据集规模较小时则相反。该发现揭示深度学习在特征提取方面的潜力，表明随着数据规模和计算能力提升，基于深度学习的医学影像分析方法有望取得更好性能，开拓了新研究方向。

1.2.2 基于深度学习方法识别肿瘤基因亚型的现状

尽管在嗜铬细胞瘤领域关于影像特征与基因表型的关联研究较少，但在神经内分泌肿瘤及其他实体瘤研究中，医学影像驱动的分型预测已取得显著进展。现行研究方法体系可根据数据建模维度划分为基于单帧和基于多帧两大类，下面将对这两类方法进行综述。

(1) 基于单帧影像进行基因亚型分类

在单帧影像分析领域已有多种创新方法，如 Yang 等人^[29]结合 CNN 与 Transformer 架构用于透明细胞肾癌 Fuhrman 核分级的非侵入性预测(CNN 提取图像局部特征，Transformer 捕捉全局信息)，实验显示该模型在分类准确率(87.1%)、敏感度(91.3%)、特异性(85.3%)和 AUC 值(0.903)等指标上显著优于传统方法。

在肾上腺肿瘤诊断方面，Kusunoki 等人^[30]在两位经验丰富的放射科医师(分别具有 4 年和 23 年经验)指导下，选取最大病变层面的轴位 CT 图像(包括未

增强图像、动脉期图像、延迟期图像及相对洗脱率图像等多种输入)训练深度卷积神经网络模型,发现使用未增强 CT 图像的模型表现最优(AUC 值达 0.99)。

在临床应用方面, Yasaka 等人^[31]采用由六层卷积层、三层最大池化层和三层全连接层组成的卷积神经网络对动态增强 CT 图像中的肝脏肿块进行分类,针对肝脏肿块异常区域面积最大帧分析并通过数据增强将训练数据扩充至 55536 幅图像,模型在 100 幅测试图像上准确率达 84%、区分恶性与良性肿块的 AUC 值为 0.92,为临床诊断提供了可靠工具。

Santra 等人^[32]开发 PPGL-Transformer 方法,利用对比增强 CT 图像识别 PPGL 遗传簇,采用双分支视觉转换器确定 PPGL 解剖位置和遗传类型,通过监督对比学习策略优化模型。五折交叉验证显示,该方法在 289 名患者数据集上实现 0.63 ± 0.08 的准确度和 0.46 ± 0.08 的 F1-score,为 PPGL 精准医疗开辟新方向。

Sori 等人^[33]受相关研究启发在肺癌检测领域取得了进展,提出基于单帧 CT 图像的肺癌检测方法 DFD-Net。该方法以结节最大横截面切片为输入,通过不同大小感受野提取局部和全局特征,结合判别相关性分析进行特征融合,在测试集上实现 87.8%的准确率、87.4%的敏感度和 89.1%的特异性,性能优异。

Li 等人^[34]提出基于 DCA-Xception 网络的肺结节分类方法,结合条件 Wasserstein 生成对抗网络与多种数据增强技术解决样本不足和类别不平衡问题,引入自适应双通道特征提取模块与卷积块注意力模块提升肺结节良恶性分类准确率。但该方法仅分析单个 CT 切片病变区域,忽视相邻切片像素关联,或影响复杂病例分类性能。

(2) 基于多帧影像进行基因亚型分类

与单帧分析相比,多帧影像分析可捕捉更丰富时空信息,Lei 等人^[35]通过基于三维 Transformer 框架的深度学习解决方法解决冠状动脉粥样硬化斑块分类问题,该方法通过桥接通信模块融合卷积局部特征提取能力与 Transformer 全局建模能力,在双分支信息交汇处引入移位注意力模块,并采用 Ghost Plus 模块平衡计算需求,最终在测试集上实现优于当时主流单帧分类模型的性能。

在此基础上,Zhu 等人^[36]进一步优化多切片特征提取策略,研究纳入 250 例患者并按 7:3 比例分为训练集和测试集,从中央切片上下、前后方向提取相邻切片数据集,同时构建基于单一切片输入的 2D 模型,使用预训练的 ResNet-101 提取深度学习特征,所有特征均通过支持向量机评估,采用五重交叉验证测量模型准确度和 AUC 值以检验诊断性能。

与此同时, Li 等人^[37]在肺癌亚型鉴别领域取得了重要突破。他们选取病变面积最大的切片和相邻切片,建立 $3\times 128\times 128$ 的 3 通道图像块。开发了一种基于 ResNet-18 架构的注意力引导残差融合网络,在肺癌患者脑转移瘤病例亚型上

进行分类，分类小细胞肺癌和非小细胞肺癌的 AUC 为 0.751。

Saihood 等人^[38]提出基于多方向局部纹理特征 (TFDs) 和引导注意力融合模块的深度学习方法用于肺结节分类，为多帧分析提供新思路。研究采用 LIDC-IDRI 数据集的 1018 例患者 CT 扫描图像，提取含肺结节的全部切片并构建 33 层体积数据，通过计算多方向共现模式提取单张 CT 切片 TFDs，经非局部注意力机制融合后得到最终特征用于分类，模型在肺结节分类任务中实现 95.2% 的准确率和 0.99 的 AUC 值。

针对模型复杂度提升带来的计算效率问题，Tian 等人^[39]提出 Transformer 和 CNN 结合的轻量化网络架构解决肺结核病例分类难题。研究使用兰州大学第一医院提供的 CT 图像数据集，对单张 CT 图像提取特征后，经融合模块整合局部与全局特征用于分类，在测试集上达成 97.23% 准确率和 98.41% 的敏感性。

Suman 等人^[40]提出基于注意力机制的两阶段 CNN-LSTM 网络，第一阶段 CNN 分类器提取 CT 切片特征，第二阶段双向 LSTM 和全连接层序列模型学习切片间远距离依赖关系，经注意力模块加权聚合生成图像级和患者级预测结果，注意力机制助力网络定位肿瘤关键切片并赋予较高权重。Shen 等人^[41]提出基于 3D 卷积神经网络的方法，通过多裁剪池化策略从 3D 胸部 CT 图像直接提取多尺度特征，省去传统结节分割和手工特征提取步骤，提升肺结节恶性肿瘤可疑性的分类性能，这些研究为后续临床应用奠定基础。

通过对现有研究方法的深入分析发现，当前基于传统影像组学的方法和单帧深度学习方法主要依赖于肿瘤区域面积最大的单帧图像进行基因亚型判断，基于深度学习方法的多数研究也局限于对经典深度学习模型的简单适配和改进。

基于此，本研究创新性地从单帧特征提取和多帧空间建模两个维度构建深度学习模型，通过系统探究单帧特征提取与多帧空间建模两种技术路径的性能差异，不仅能够确定最适合 PPGL 基因亚型识别的技术方案，更重要的是为这一领域建立了一套全新的技术范式和研究方法论，为实现精准医疗提供了重要的技术支撑。

1.3 研究目的与内容

本研究通过两条并行技术路线构建基于单帧和多帧的识别模型，系统探索 CT 影像在 PPGL 基因亚型分类中的临床应用价值。研究首先从单帧影像分析切入，基于罕见病数据稀缺的现实背景，单帧模型因参数量小、计算复杂度低，更适合数据量有限的罕见病场景，且肿瘤区域面积最大帧包含一定基因亚型信息。例如 Yang 等人^[29]在肾癌研究中证实最大截面切片往往含最显著形态学特征，Lin 等人^[28]研究也表明单帧分析在数据有限时仍可实现高效模型训练与优化，因此基于肿瘤区域面积最大帧的单帧分类模型为 PPGL 研究提供了可行技术路径。

然而,单帧分析方法因无法全面反映肿瘤空间异质性和三维生长模式存在固有局限,如 Li 等人^[34]研究表明仅依赖单一切片可能丢失重要空间上下文信息,尤其嗜铬细胞瘤影像中肿瘤形态复杂性和空间结构常无法通过单一切片充分捕捉。为此本研究提出基于多切片关联增强的识别方法,该模型不同于传统 3D 方法,不依赖完整三维卷积,而是通过增强切片间空间关系,在有限数据条件下有效提升模型对肿瘤空间信息的建模能力。

本研究的意义在于,通过深度学习技术改善 PPGL 患者基因亚型识别方面对于无创技术的需求,帮助优化 PPGL 患者的个体化诊疗策略。同时,本研究通过并行设计独立的单帧与多帧深度学习模型,系统比较二者性能差异,为临床选择更高效的影像组学方案提供依据。

本文的主要研究思路和工作如下:

(1) 构建嗜铬细胞瘤影像-基因关联数据集

本研究构建并整理一套完整的嗜铬细胞瘤影像-基因关联数据集,通过系统检索 2010 年 7 月至 2022 年 10 月期间以湘雅医院为核心、长沙及周边医院为补充的多中心医疗数据库,筛选经外科病理确诊的 PPGL 患者病例,对患者 CT 数据清洗整理后最终获得 425 例高质量患者数据。数据预处理阶段对肿瘤区域的分析发现关键特征:不同患者肿瘤区域大小差异显著,且大多数患者肿瘤区域面积占整张图像比例较低,这些发现为后续模型改进提供重要依据。

(2) 基于单帧图像的分类模型改进

为提高肿瘤切片图像分类的精度,本研究提出了**基于肿瘤区域面积最大帧的分类模型 (Tumor Area Maximum Network, TAMax-Net)**,针对不同患者肿瘤尺寸变化大及低肿瘤占比切片图像的特点,从三方面提升模型表征能力:首先设计局部细节优化模块,通过多层次特征提取和细粒度调整增强网络对局部细节的感知能力及多尺度特征融合能力,提升模型对不同尺寸肿瘤区域的适应性;其次引入特征融合编码器融合不同分辨率特征图,避免低肿瘤面积占比图像在特征传递过程中的信息损失;最后针对融合后特征图兼具浅层细节信息和深层语义信息的特点,设计末端注意力模块,通过注意力机制使模型更聚焦重要特征以进一步提升分类性能。

(3) 基于多帧融合的分类模型设计

针对传统单帧分类模型在空间信息利用和肿瘤形态特征提取方面的不足,本研究提出**基于肿瘤多帧融合的分类模型 (Multi-frame Fusion Network, MFF-Net)**。传统方法通常仅基于肿瘤区域面积最大的一帧进行分析,而最大帧未必能完整展示肿瘤全部特征,这限制了模型精度和鲁棒性提升。为此本文提出多帧融合分类模型,旨在解决帧序列信息冗余及肿瘤整体形态与空间分布特征缺

失问题以提升分类性能。具体设计包括：首先通过特征聚焦模块生成精确包围肿瘤区域的最小外接矩形框，并对框内图像进行 ROI-Align 操作，提升模型细节保留能力；其次引入空间特征提取模块，利用可学习向量和自注意力机制传播帧间上下文信息，通过交叉注意力机制聚焦肿瘤特征以有效捕捉整体形态和空间分布；最后提出序列感知多帧融合模块，采用双路并行架构分别建模局部序列依赖和全局动态变化，实现多帧特征深度融合。上述模块协同作用使多帧融合分类模型在全局和局部层面有效提取肿瘤空间特征，避免信息冗余和细节丢失，显著提升复杂肿瘤切片上的分类表现。

1.4 本文组织结构

第一章介绍嗜铬细胞瘤基因亚型诊断的研究背景与意义，从传统影像组学和深度学习路径综述国内外相关研究进展，引出研究方法思路，提出研究目的并概括论文整体结构。

第二章阐述本研究数据集的构建过程及模型评估方法，说明选择 NECT 作为数据模态的原因，详细介绍数据收集、清洗与分析过程，以及后续实验所用的损失函数、模型评价指标和实验环境配置。

第三章探索基于肿瘤区域面积最大帧的分类模型的有效性，针对不同患者肿瘤尺寸变化大及低肿瘤占比问题，设计局部细节优化模块通过多层次特征提取与细粒度调整提高多尺度肿瘤特征提取能力，引入跨层级特征融合模块处理不同分辨率特征图以减少低肿瘤占比图像信息损失并提升网络适应性，结合末端注意力模块强化模型对关键特征的关注，显著提升复杂肿瘤切片的分类性能。

第四章针对单帧模型在空间信息利用和肿瘤形态特征提取上的不足，设计基于肿瘤多帧融合网络的分类模型，针对多帧引入肿瘤区域面积较小切片可能导致的帧退化问题，在模型前端设计特征聚焦模块通过特征驱动的自适应图像分析定位肿瘤区域并利用 ROI-Align 操作放大该区域以提升细节保留能力，引入空间特征提取模块利用可学习向量表示肿瘤信息、基于自注意力机制传播帧间上下文信息并通过交叉注意力机制聚焦肿瘤特征以有效捕捉肿瘤整体形态和空间分布，提出序列感知多帧融合模块采用双路并行架构分别建模局部序列依赖与全局动态变化以实现多帧特征深度融合，各模块协同作用显著提升模型在复杂肿瘤切片上的分类精度与鲁棒性。

第五章总结全文三大主要工作，并对后续研究方向进行展望。

第 2 章 数据集建立与模型评估

本章介绍 CT 影像生成机制，说明其在肿瘤特征表征中的优势并阐明以 NECT 为主要研究模态的理论依据。针对罕见病数据稀缺问题，本研究以湘雅医院临床数据为核心构建了包含 425 例患者样本的湘雅嗜铬细胞瘤影像基因数据集（Xiangya Paraganglioma-Pheochromocytoma Imaging Genomics Dataset, Xiangya-PPGL），该数据集涵盖了 CT 影像与基因测序数据。本文对数据集进行系统数据预处理与分析，为后续研究奠定基础。

2.1 计算机断层扫描技术

计算机断层扫描（Computed Tomography, CT）是利用 X 射线对人体内部成像的工具，其通过计算机处理多角度 X 射线拍摄的图像数据生成高分辨率患者横断面图像，推动了医学影像学发展，使临床医生能直观观察人体内部结构，在诊断肿瘤、内脏器官疾病等方面具有重要意义。

2.1.1 CT 技术的基本原理

CT 技术基于 X 射线穿透特性工作，X 射线作为波长较短的电磁波具有强穿透能力，其光束通过人体时不同组织对 X 射线吸收和散射程度各异，可在图像上以深浅不同的视觉差异展现人体组织结构差异，CT 通过多角度多方位 X 射线扫描形成详细横断面图像，扫描过程通常由旋转的 X 射线管和固定探测器组成，X 射线管与探测器阵列围绕患者身体旋转以采集不同角度的 X 射线数据，每个 X 射线束穿过人体时的衰减程度遵循比尔-朗伯定律如公式 2-1 所示：

$$I = I_0 e^{-\mu x} \quad (2-1)$$

其中， I_0 是入射 X 射线的强度， I 是穿透物体后的 X 射线强度， μ 是物体的线性衰减系数， x 是 X 射线穿透物体的路径长度。通过计算穿透人体的 X 射线的衰减系数，就可以推算出人体内部结构的二维投影。然后，通过计算机的图像重建技术，将二维的投影数据转换为多层次的横截面图像，从而生成一个三维数据。这样医生就能够从不同的角度观察人体内部的结构。

CT 技术最重要的价值就是能够提供高分辨率的横断面图像，帮助医生清晰观察人体内部结构。

2.1.2 常见的 CT 成像方式及其特点

CT 成像技术根据其扫描方式和应用场景的不同，可以分为多种类型，每种类型都有其独特的优势和局限性。以下是几种常见的 CT 成像方式：

非增强 CT (Non-Enhanced CT, NECT) 作为 CT 成像的基础模态, 无需依赖造影剂即可完成检查, 是临床常规检查的重要组成部分。其显著优势就是安全性高, 可以有效规避造影剂可能引发的过敏反应及肾功能损伤等风险。同时, NECT 在钙化灶检测及出血性病变识别方面有着优异的诊断效能, 常作为病变初步评估的首选方法。绝大多数确诊 PPGL 的病人都会有 NECT 数据。

增强 CT (Contrast-Enhanced CT, CECT) 通过注射含碘造影剂增加图像对比度使血管、器官及病变组织更清晰, 但其成本高且存在使用风险, 需考虑患者对造影剂耐受性及潜在过敏反应, 肾功能不全患者使用含碘造影剂可能加重肾脏负担甚至引发急性肾损伤。

螺旋 CT 因快速连续扫描能在短时间完成大面积扫描、减少呼吸或心跳引起的运动伪影, 但其辐射剂量较高限制了在长期随访中的应用。

除常见 CT 成像方式外还有电子束 CT、高分辨率 CT、低剂量 CT 及 CT 血管成像等特殊用途 CT 技术, 这些技术各有独特优势和应用场景, 而本研究针对嗜铬细胞瘤选择合适成像方式至关重要。

2.1.3 NECT 在 PPGL 研究中的应用价值

在 PPGL 的分类和评估中, NECT 因其独特的优势, 成为一种理想的选择。

(1) **安全性高, 适用性广。** NECT 无需使用造影剂, 避免了因造影剂可能引发的过敏反应和肾功能损害。NECT 的广泛应用性和较低的并发症风险使其成为一种安全有效的工具。并且绝大多数确诊 PPGL 的患者都会有 NECT 数据。

(2) **有效识别钙化和出血。** NECT 在检测钙化灶和出血等病变方面具有显著优势。钙化和出血是 PPGL 的重要影像学特征, NECT 能够清晰显示这些特征, 从而为 PPGL 的分类提供重要依据。

(3) **通用性强, 局限性小。** 在 PPGL 的诊断流程中, NECT 通常是大多数患者的第一步检查。因此, 基于 NECT 数据的 PPGL 基因亚型识别方法具有广泛的适用性, 能够覆盖大多数患者群体。

在 PPGL 基因亚型识别研究中, NECT 因其安全性高、局限性小以及能够有效识别钙化和出血等特征, 成为一种理想的选择。

2.2 湘雅嗜铬细胞瘤影像基因数据集 Xiangya-PPGL

人工智能在 PPGL 领域研究面临重大挑战, 缺乏高质量数据集是制约其应用的关键瓶颈。目前国内外尚未公开专门针对嗜铬细胞瘤的研究数据集, 研究人员只能依赖自建数据集, 且样本收集渠道有限、研究成本高昂、伦理审批严格等因素限制了数据集规模, 导致人工智能技术在嗜铬细胞瘤诊断、分型和预后评估等关键环节的潜力难以充分发挥。鉴于此, 本研究与长沙市湘雅医院紧密合作, 双

方共同组建涵盖放射科医师、泌尿科医师及信息技术工程师等专业人才的多学科交叉研究团队，系统检索 2010 年 7 月至 2022 年 10 月期间以湘雅医院为核心、长沙及周边医院为补充的多中心医疗数据库，筛选出所有经外科病理确诊的 PPGL 患者病例，最终构建湘雅嗜铬细胞瘤影像基因数据集（Xiangya-PPGL）。

2.2.1 数据集的获取

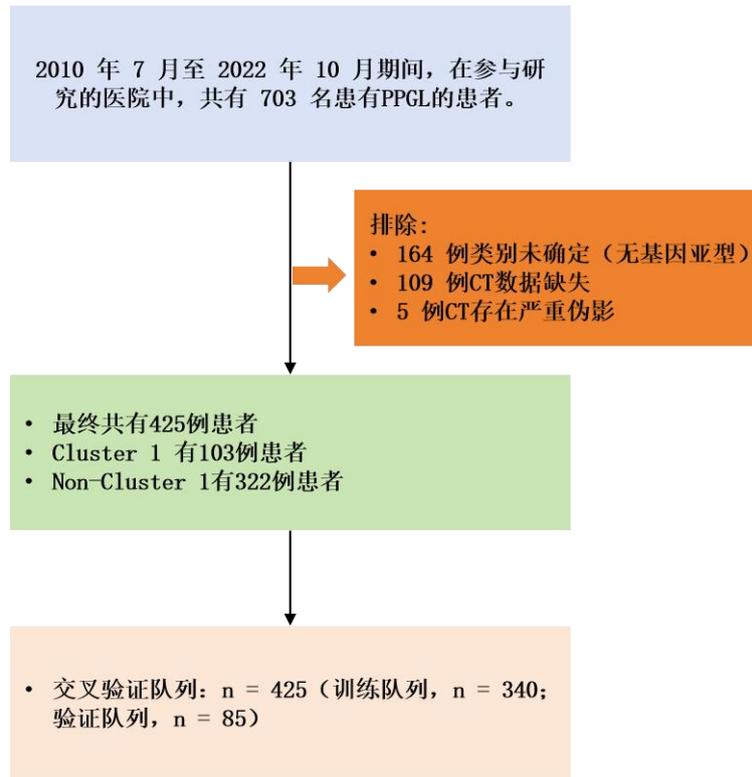


图 2-1 研究队列的识别和选择以及排除标准

本研究已获得各参与医院的机构审查委员会（IRB 编号：2021101051）的批准，遗传数据已在中国临床试验注册中心登记（ChiCTR2100050937）。鉴于本研究为回顾性研究，故未获取患者的知情同意。本研究检索了 2010 年 7 月至 2022 年 10 月期间参与医院外科病理确诊的 PPGL 患者的数据库。研究队列的纳入、排除标准及选择流程见图 2-1。该队列的临床信息和 CT 扫描从医疗记录中检索，并由课题组导师匿名。这些数据随后由两位研究人员进行了审查。研究人员间的任何分歧均由第三名高级临床研究员裁决。本研究中所有患者的病理切片由两名泌尿病理学专家（分别具有 10 年和 5 年经验）重新评估。当这两位病理学家之间有差异时，由第三位有经验的病理学家（有 26 年的泌尿病理学经验）来确定最终的病理。基因突变测试如文献^[42]所述。将缺氧通路激活基因突变驱动的病例被归类为 **Cluster 1 型 PPGL（假性缺氧亚型）**，排除这些基因中未知的显著变异。没有或伴有其他 PPGL 相关基因突变的患者被定义为 **Non-Cluster 1 型 PPGL（非**

假性缺氧亚型)。

所有肿瘤区域影像标注由湘雅医院医学影像中心认证且从业 18 年的资深放射科医师使用专业医学影像标注软件 3D Slicer 完成, 该软件支持 DICOM 格式数据三维可视化浏览, 可通过调节窗宽窗位清晰呈现肿瘤与周围组织边界并提供画笔、橡皮擦、区域生长等工具实现精细化勾勒, 医师在轴位、冠状位、矢状位三平面视图下沿肿瘤边缘逐帧勾勒感兴趣区域并结合 CT 值及三维重建形态对边界模糊区域综合判断。初步标注完成后研究采用双盲复核机制确保标注准确性, 即由另一位同等资质放射科医师对标注结果进行交叉审核, 针对分歧区域通过多平面重建及三维容积分析讨论修正后由主任医师终审, 经检验标注结果组内相关系数 (ICC) 达 0.92, 有效保障标注的一致性与可靠性。

2.2.2 数据集的预处理



图 2-2 Xiangya-PPGL 数据集预处理流程

本节详细介绍 Xiangya-PPGL 数据集预处理步骤, 该过程是保障影像数据质量与一致性, 后续分析奠定基础的重要环节, 预处理核心步骤见图 2-2。

(1) 格式转换

医学 CT 数据多以 DICOM 格式存储, 因后续有重采样需求, 需将其转换为常用的 NIFTI 格式^[43], 具体是把切片数量作为第三个维度, 将每位患者多个 DICOM 格式数据批量转成 nii.gz 的 3D 格式。

(2) 图像重采样

重采样作为医学图像处理关键技术, 借助插值技术把原始数据体素尺寸和空间分辨率调整到特定大小, 目标图像每个体素值经坐标映射与插值计算从原始图像获取。CT 三维数据里像素间距以毫米为单位体现相邻像素物理距离, 其大小直接决定 CT 图像空间几何形状和尺寸。本研究数据源于不同医院, 各医院或设备生成的 CT 数据分辨率存在差异, 不统一会影响数据一致性。本研究利用重采样使所有数据体素保持一致, 实现多中心数据整合, 消除设备或协议差异对结果的影响, 将全部数据像素间距都转化为 [1.0, 1.0, 1.0] 的形式。

(3) CT 值转换

医学 CT 扫描原始 CT 值范围一般在 [-1000HU, 3000HU], 普通自然图像灰度分辨率范围仅为 [0, 255], 直接将原始 CT 影像转为灰度图会丢失重要生物学信息, 所以医学影像分析常用窗口化处理技术, 通过设定窗宽和窗位把不同 CT 值

区间映射到图像灰度范围，突出特定组织结构细节特征，既保留关键诊断信息又增强组织间对比度，便于医生准确区分解剖结构，为临床诊断和治疗决策提供可靠影像依据。窗口化技术的两个关键要素是窗宽和窗位，窗宽决定图像灰度级别范围大小，取值为正整数表示用于映射像素值的灰度范围，窗位确定图像灰度级别范围中心位置，取值为正整数控制像素值中心灰度级别，如公式(2-2)(2-3)所示，其中 a 和 b 分别指定需放大的上下边界值进而确定最终放大范围。

$$WW = b - a \quad (2-2)$$

$$WL = (a + b) / 2 \quad (2-3)$$

采用窗宽窗位调整技术，可以将原始数据的CT值与灰度值映射到一个函数(2-4)，从而更好地描述数据在处理过程中的像素强度变化：

$$j = f(i, a, b) = \begin{cases} 0, & i < a \\ \Delta \times \frac{i - a}{b - a}, & a \leq i \leq b \\ \Delta, & b < i \end{cases} \quad (2-4)$$

其中， $\Delta=255$ ，表示最大灰度值， i 代表原始像素点的CT值，而 j 则指经过映射处理之后图像的像素灰度值。使用窗口化技术改变图像宽高可实现以下效果：

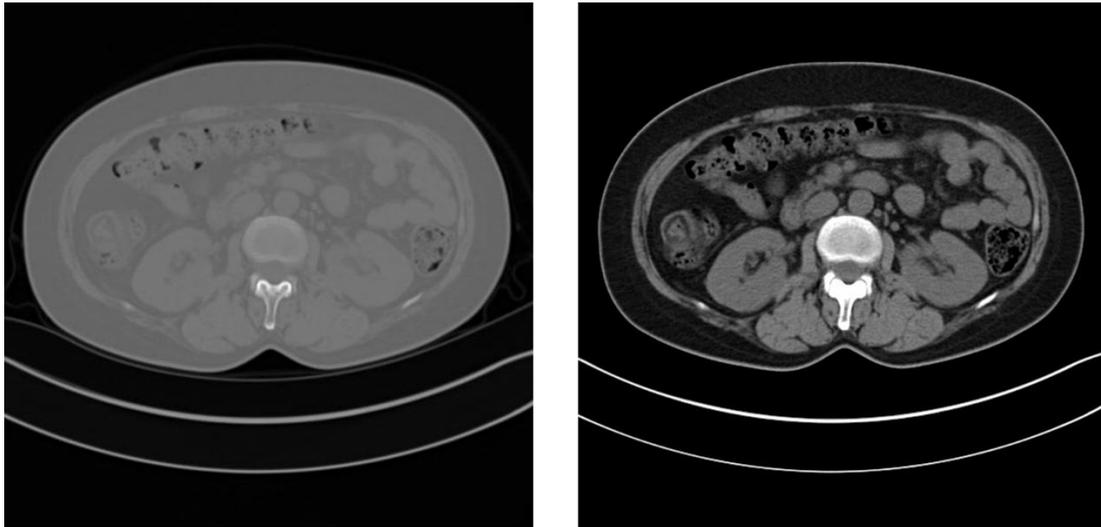
(1) 窗宽影响图像对比度，大窗宽可显示更多组织结构但降低微小病变清晰度，小窗宽提升图像对比度便于发现细微病变却可能增强图像噪声；

(2) 窗位确定CT值显示范围中心点，提高窗位侧重显示高密度组织，降低窗位侧重显示低密度组织，临床中放射科医生依据诊断需求调整窗宽窗位值，常见人体组织标准窗宽窗位设置见表2-1。

表 2-1 人体常见组织的窗宽和窗位值^[44]

部位	窗宽	窗位
肝脾	100~200	30~45
头颅	80~100	30~40
胸腔	300~500	30~50
腹部	300~500	30~50
肾脏	200~300	25~35
胰腺	300~500	30~50

结合表2-1建议值与湘雅医院放射科医生的实践尝试，本研究将窗宽设为260、窗位设为32，图2-3对比显示未经优化的图像(图2-3-a)较为模糊，调整后的图像(图2-3-b)清晰度和细节表现力更高。



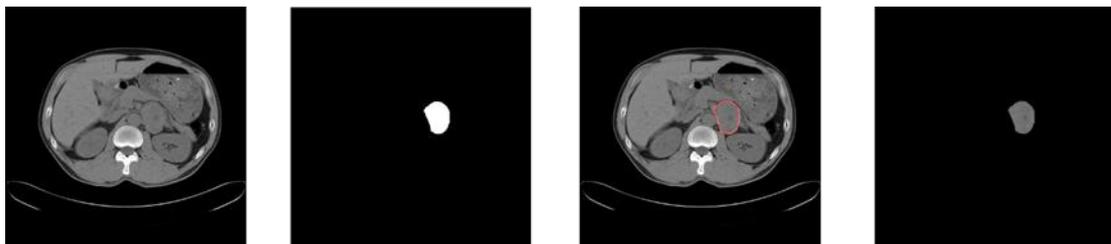
(a)原始 CT 图像

(b)CT 值转化后图像

图 2-3 CT 值转化对比图

(4) 肿瘤区域提取

将处理后的 CT 图像与医生标注的肿瘤区域配准重叠,提取出仅含肿瘤区域、背景全黑的图像数据,此过程确保肿瘤区域边界与标注完全一致,防止错误提取正常组织,全黑背景消除非肿瘤区域干扰信号让模型专注分析肿瘤特征,结果如图 2-4 所示,该方法能精准分离肿瘤区域并保留原始形态和灰度信息,为后续分析提供高质量肿瘤区域图像,显著提升模型训练的准确性与可靠性。



(a)CT 值转化后图像

(b)肿瘤区域标注结果

(c)标注结果可视化

(d)肿瘤区域

图 2-4 CT 图处理后结果及标注结果

(5) 裁剪图像与填充

原始体素间距各异的患者数据经重采样和 CT 值转换后会出现图像尺寸不统一的情况,而神经网络对输入图像的尺寸调整会扰乱不同患者肿瘤尺寸相对关系,因此需通过裁剪和填充统一图像尺寸:对大尺寸图像裁剪提取关键区域,小尺寸图像用黑色像素值填充边缘,最终将所有图像调整 224×224 尺寸,该操作可避免因图像尺寸不一致导致的模型性能下降。裁剪能显著提升肿瘤区域在图像中的

占比，如图 2-5 所示，裁剪前约 90% 的 CT 切片肿瘤区域面积占比低于 5%，裁剪后该比例降至 48%，使肿瘤区域更突出，助力模型精准捕捉肿瘤特征提升分类性能，统一的输入尺寸也增强了模型对不同来源数据的泛化能力。

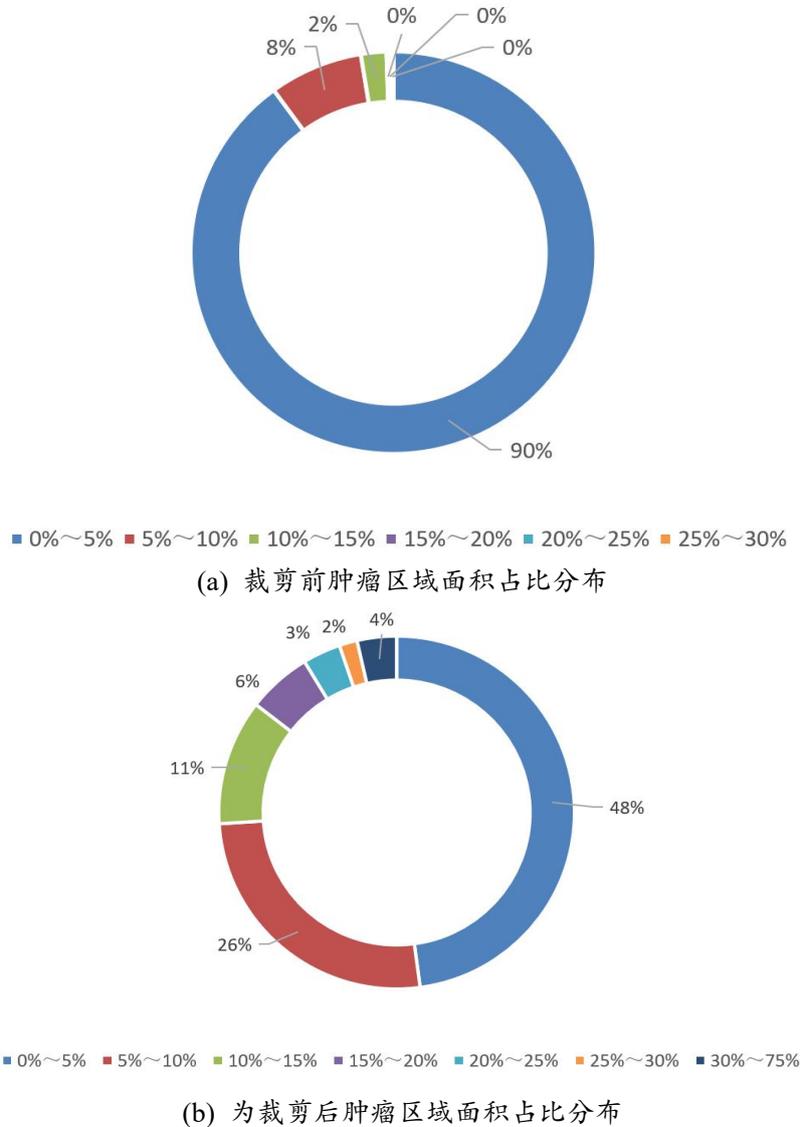


图 2-5 裁剪前后肿瘤区域面积占比分布图

2.2.3 数据集的分析

数据预处理后本数据集含 425 名患者 20741 张 CT 切片影像数据，其中 322 名 **Non-Cluster 1** 类（非假性缺氧亚型）患者、103 名 **Cluster 1** 类（假性缺氧亚型）患者。表 2-2 显示 **Non-Cluster 1** 类患者占数据集约 75.8%、**Cluster 1** 类约 24.2%，这种不均衡分布或使模型对 **Cluster 1** 类患者识别能力弱于 **Non-Cluster 1** 类。鉴于本研究数据集样本有限且医学图像具有临床特殊性与复杂性，为规避采样及数据增强引入偏差决定保留原始数据分布，后续将采用**加权损失函数**提升模型对

Cluster 1 类患者的识别能力。

表 2-2 数据集类别分布

类别	Non-Cluster 1	Cluster 1
患者数	322	103
CT 切片数	15843	4898

部分样例图像预处理后展示于图 2-6，可见不同患者肿瘤面积及同一患者不同帧肿瘤面积均存在明显差异，表明肿瘤在不同切片中表现出异质性。从视觉角度分析，Non-Cluster 1 类与 Cluster 1 类患者肿瘤形态均呈现类圆形、多结节状等特征，肿瘤区域密度变化均匀、灰度值接近，未见明显坏死区域或不同组织显著分化，说明不同基因亚型肿瘤在影像学特征上具有相似性。本研究对所有 CT 切片肿瘤区域面积的统计分析结果如图 2-5-b 所示，嗜铬细胞瘤患者 CT 影像呈现不同患者肿瘤尺寸变化大、肿瘤区域面积占比小的特点，前者导致影像中边界清晰度、内部结构完整性及微小病灶辨识度等关键细节信息表达不一致，影响神经网络提取稳定有效的形态学特征，后者使模型难以提取有效信息，尽管通过裁剪提升肿瘤区域占比，仍有 48% 的切片肿瘤区域面积低于整张图的 5%，即便将非肿瘤区域设置为全黑以减少背景干扰，神经网络仍难以从大量无效信息中准确捕捉关键肿瘤特征，因此临床上通常采用肿瘤区域最大的一帧进行特征提取^[45-48]，该帧可较清晰反映肿瘤形态特征，但仅选择最大帧会忽略肿瘤动态变化、生长模式等空间信息。

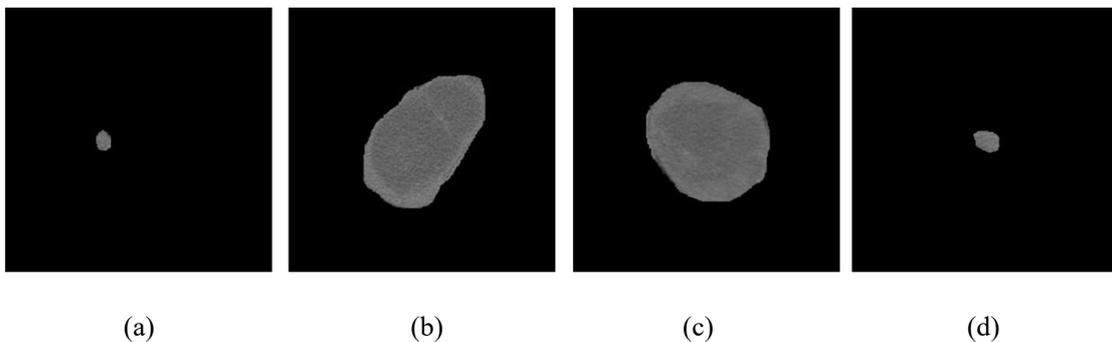


图 2-6 Xiangya-PPGL 数据集部分样本示例：(a)、(b) 为某位 Non-Cluster 1 类患者 CT 图像；
(c)、(d) 为某位 Cluster 1 类患者 CT 图像

故本研究考虑到临床诊断中普遍采用肿瘤区域面积最大帧进行分析的实际情况，构建了基于最大帧的诊断模型。选用最大帧不仅符合临床常规，也为后续研究提供了可对比的基准。其次，针对单帧分析在反映肿瘤空间特征方面的局限性，开发基于多帧序列的增强模型。为了弥补临床常规方法的不足，通过多帧信息来提取肿瘤的空间特征。这种研究设计既尊重了临床实践的现实需求，又突破

了传统诊断方法的局限，为肿瘤的精准诊断提供了更全面的解决方案。

2.3 模型测试框架与性能度量

2.3.1 损失函数

数据分析结果表明，该数据集存在明显的类别分布不平衡问题。针对这一问题，本研究采用加权损失函数的方法对模型进行优化，通过给 Cluster 1 型样本分配较大的权重，为 Non-Cluster 1 型样本分配相对较小的权重，从损失梯度的角度对样本的贡献进行再平衡，缓解数据不平衡带来的负面影响。加权交叉熵损失函数公式如 2-5 所示：

$$L = -\frac{1}{N} \sum_i^N [w_{y_i} y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (2-5)$$

其中， N 为样本总数， y_i 表示样本 i 的标签，正类为 1，负类为 0。 p_i 表示第 i 个样本被预测为正类的概率， w_{y_i} 为第 i 个样本所属类别的权重。权重的计算公式如公式 2-6 所示。

$$w_{y_i} = \frac{1}{\text{count}(y_i)} \quad (2-6)$$

2.3.2 评价指标

评价指标是衡量深度学习模型性能的重要工具。以下是对一些评估指标的简要说明。

以本研究的 PPGL 基因亚型分类为例，假设将 Non-Cluster 1 型 PPGL 作为正类，Cluster 1 型 PPGL 作为负类。当模型正确识别出 Non-Cluster 1 型 PPGL 样本时，记为真正例 (TP)；当模型将 Cluster 1 型 PPGL 样本误判为 Non-Cluster 1 时，记为假正例 (FP)；当模型正确识别出 Cluster 1 型 PPGL 样本时，记为真反例 (TN)；当模型将 Non-Cluster 1 型 PPGL 误判为 Cluster 1 型 PPGL 时，记为假反例 (FN)。

准确率 (Accuracy) 衡量模型在所有样本中正确预测的比例，计算公式如 2-7 所示。

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2-7)$$

敏感度 (Sensitivity) 反映了模型正确识别 Non-Cluster 1 型 PPGL 的能力，计算方法如公式 2-8 所示。较高的敏感度表明模型能够有效识别大多数的 Non-Cluster 1 型。

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2-8)$$

特异性 (Specificity) 则衡量了模型正确识别 Cluster 1 型 PPGL 的能力, 计算方法如公式 2-9 所示。较高的特异性表明该模型能够有效识别大多数的 Cluster 1 型。

$$Specificity = \frac{TN}{TN + FP} \quad (2-9)$$

ROC 曲线 (Receiver Operating Characteristic Curve) 是用于评估二分类模型性能的工具。该方法通过调整分类阈值, 计算并绘制模型在不同决策阈值下的真阳性率 (TPR) 与假阳性率 (FPR) 的变化关系。在 ROC 曲线图中, 横轴表示 FPR (实际为负例但被误判为正例的比例), 纵轴代表 TPR (实际正例被正确识别的比例), 曲线上的每个坐标点都对应一个特定分类阈值下的模型性能表现。TPR 和 FPR 的具体计算公式详见式 (2-10) 和式 (2-11)。这种分析方法能够全面反映模型在不同判别标准下的分类能力, 为模型性能评估提供重要依据。

$$TPR = \frac{TP}{TP + FN} \quad (2-10)$$

$$FPR = \frac{FP}{FP + TN} \quad (2-11)$$

AUC (Area Under the ROC Curve) 是 ROC 曲线与坐标轴围成的区域面积, 可以量化评估 ROC 曲线的优劣, 是衡量分类模型综合性能的核心指标。当 AUC 值趋近于 1 时, 表明模型具有优异的判别能力, 当 AUC 值趋近于 0.5 时, 表明模型分类能力近似于随机猜测。在医学影像分析中, 因为 AUC 指标对类别不平衡数据具有鲁棒性被广泛应用。

2.3.3 实验配置

表 2-3 实验的软件与硬件环境

类别	具体描述
CPU	Intel(R) Xeon(R) Platinum 8336C CPU @ 2.30GHz
GPU	NVIDIA GeForce RTX 4090
内存	128GB
操作系统	Ubuntu 20.04.6 LTS
编程语言	Python 3.9.12
深度学习框架	PyTorch 2.2.2

本研究所有实验均在统一的计算环境下完成, 具体配置如下: 软件平台采用 Ubuntu 20.04.6 操作系统, 并基于 PyTorch 2.2.2 深度学习框架进行开发; 硬件方面搭载 Intel Xeon Platinum 8336C 处理器(主频 2.3GHz)和 NVIDIA GeForce RTX 4090 显卡。详细的软硬件环境配置参见表 2-3。

本研究的实验实现基于 Torchvision 工具库完成。在模型训练过程中, 主要

采用以下超参数配置：训练周期（Epoch）设为 300 次迭代，优化器选用 Adam 算法，基础学习率设置为 0.0001，并引入 0.01 的权重衰减系数（L2 正则化）。为进一步提升模型泛化性能，在网络各层输出后均实施了 10% 比例的 Dropout 操作。完整的实验超参数配置详见表格 2-4。

表 2-4 实验超参数设置

名称	具体设置
批次大小	32
优化器	Adam
初始学习率	0.0001
权重衰减参数	0.01
训练轮次	300

2.4 本章小结

本章解释了采用 NECT 作为主要研究模态的原因。接着，整理出一个全新的，带有 PPGL 基因测序标签和肿瘤区域标注的湘雅嗜铬细胞瘤影像基因数据集（Xiangya-PPGL），收集了以湘雅医院临床数据为核心的 425 例真实就诊患者，包含每个患者的 NECT 序列信息。本章具体介绍了数据集的获取、清洗、标注、预处理的实现。最后介绍了后续研究使用的损失函数，模型评价指标和实验环境。

第 3 章 TAMax-Net: 基于肿瘤区域面积最大帧的分类模型

上一章节中本研究依托中南大学湘雅医院丰富临床资源通过多中心协作研究, 系统性收集、整理并构建嗜铬细胞瘤基因亚型数据集 Xiangya-PPGL 为后续研究提供数据支持, 通过第二章对数据集的分析发现嗜铬细胞瘤患者 CT 影像呈现不同患者肿瘤尺寸变化大、肿瘤区域面积占比小的特点, 该特性会严重影响神经网络在肿瘤基因亚型上的分类表现。

本章提出基于肿瘤区域面积最大帧的深度学习分类方法——**最大帧分类网络 (Tumor Area Maximum Network, TAMax-Net)**。选最大帧作分类依据是因其表达肿瘤形态学信息上有优势, 可提供较完整肿瘤结构信息帮助网络全面捕捉关键特征, 还能清晰呈现肿瘤边界轮廓与内部结构细节, 为模型提供优质训练样本。基于此构建以肿瘤区域面积最大帧为核心的深度学习模型, 着力解决嗜铬细胞瘤基因亚型分类中肿瘤尺寸差异大、小面积样本占比高两大关键问题, 旨在为临床提供更精准辅助诊断工具支持个性化治疗决策。

3.1 概述

3.1.1 总体思路

深度学习技术通过模拟人脑神经结构并经大量数据训练在自然图像处理领域成果显著, 但应用于 PPGL 医学影像面临独特挑战, 最突出的是不同患者肿瘤图像尺寸差异大、病变区域面积占比低。

患者间肿瘤尺寸变化大造成同一病变特征在不同图像中形态结构表现不一, 深度学习模型难以捕捉具有诊断价值的特征, 易引发误诊漏诊, 如同病不同亚型肿瘤在影像中可能仅呈现微小形态或信号强度差异, 而这些差异对区分亚型、评估病情和指导个性化治疗至关重要。

如 2.2.3 节数据集分析显示, 医学 CT 影像肿瘤区域面积占比低, 即便在最大 CT 切片中肿瘤区域也只占图像极小部分, 深度学习模型提取特征时易忽视微小肿瘤区域而关注大面积背景, 且小肿瘤区域信噪比低、形态多样进一步加大特征提取难度, 影响模型准确捕捉肿瘤关键特征和诊断准确性。

针对肿瘤尺寸变化大与肿瘤区域面积占比低的双重挑战, 本文展开详细研究, 通过局部细节优化模块提升网络局部感知和多尺度特征融合能力, 利用跨层级特征融合增强模块融合不同分辨率特征图保留更多细节, 还设计末端注意力模块在网络末端强化模型对重要特征的关注, 防止因细节与语义信息混杂导致模型难以聚焦关键特征。

3.1.2 模型总体结构

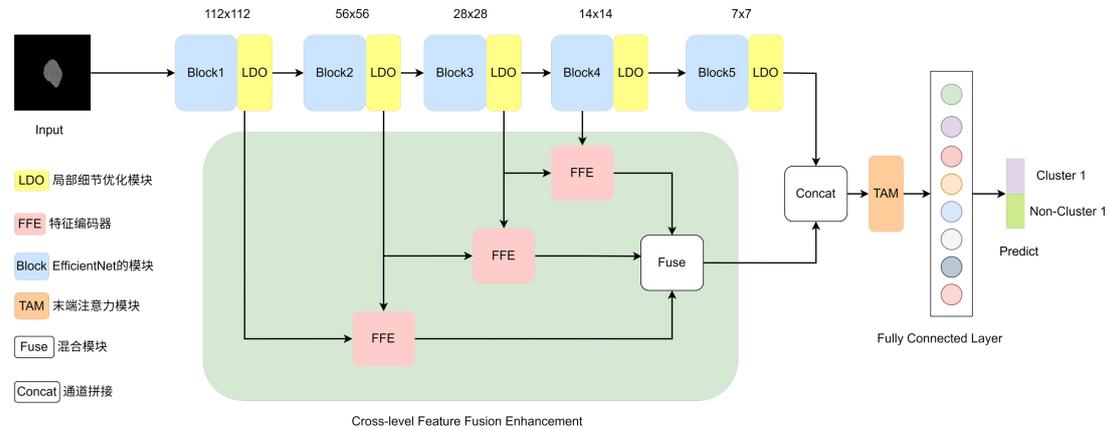


图 3-1 TAMax-Net 网络结构图

整体网络流程结构如图 3-1 所示，具体设计实现如下：

(1) 输入数据初始化处理：先将输入图像数据预处理，把原始图像尺寸统一调整为 224×224 像素，再将预处理后的图像转为张量形式输入主干网络 TAMax-Net，保证输入数据格式一致并为后续特征提取奠定基础。

(2) 主干网络结构优化：针对 EfficientNet-B0 架构进行优化以提升网络提取图像细节和全局特征的能力，在 EfficientNet-B0 的 Block1 至 Block5 层后端嵌入局部细节优化模块（LDO）增强网络对图像局部细节的感知。

(3) 特征融合与编码：为充分利用多层次特征信息，网络采用跨层级特征融合增强模块（CFE）融合主干网络多层输出，Block1 与 Block2、Block2 与 Block3、Block3 与 Block4 的输出特征图分别经特征融合编码模块（FFE）两两融合，既保留更多细节信息又减少特征传递中的信息丢失，为分类提供更丰富特征表示，所有经 FFE 处理的特征图再通过 Fuse 模块最终融合生成更具判别性的特征表示。

(4) 最终分类结果生成：特征融合阶段 Fuse 模块输出与主干网络第五层输出经通道拼接形成高维特征表示，Concat 后引入末端注意力模块（TAM），该模块通过感知全局信息评估特征图不同区域重要性并动态调整权重，增强模型对重要特征的学习能力，避免细节与语义信息混合导致模型难以聚焦关键特征，提升网络整体性能，该特征表示经全连接层处理后通过 Softmax 函数生成最终分类概率，确保模型基于多层次多尺度特征信息实现高精度分类。

3.2 模型主要模块

3.2.1 EfficientNet 介绍

谷歌团队开发的 EfficientNet 是一种高效 CNN 架构, 采用多维复合缩放方法, 在计算资源有限时可实现更高准确率与计算效率。其核心是统一缩放网络深度 (*depth*)、宽度 (*width*) 及输入图像分辨率 (*resolution*) 以优化神经网络性能, 相比传统单一维度缩放方法, EfficientNet 在三个维度协调缩放, 避免单一维度扩展时的性能瓶颈, 通过复合缩放公式在深度、宽度和分辨率间找到最优平衡, 解决传统卷积神经网络单一维度缩放易致性能快速饱和的问题。

复合缩放的公式如下所示:

$$\begin{aligned}
 \text{depth: } d &= \alpha^\phi \\
 \text{width: } w &= \beta^\phi \\
 \text{resolution: } r &= \gamma^\phi \\
 \text{s.t. } \alpha \cdot \beta^2 \cdot \gamma^2 &\approx 2 \\
 \alpha \geq 1, \beta \geq 1, \gamma \geq 1
 \end{aligned} \tag{3-1}$$

其中, d 、 w 和 r 分别表示网络的深度、宽度和分辨率, α 、 β 、 γ 是常数系数, ϕ 是复合缩放系数。通过选择合适的 α 、 β 、 γ 值, EfficientNet 能够在保持计算效率的同时, 提升模型的准确性。

EfficientNet-B0 作为 EfficientNet 系列基础模型经复合缩放策略优化, 其网络架构由 16 个 MBConv 模块构成, 每个模块含深度可分离卷积与逐点卷积等操作, 且 MBConv 模块引入 Squeeze-and-Excitation (SE) 模块增强特征表达能力。

MBConv 模块的具体结构如图 3-2 所示:

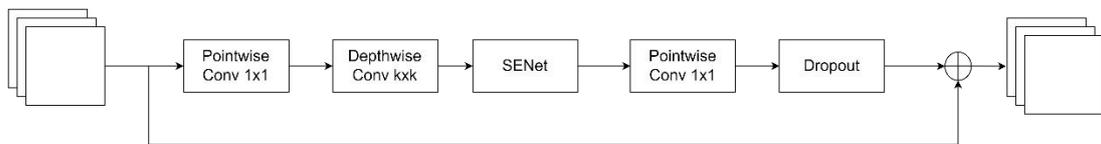


图 3-2 MBConv 结构图

通过 1×1 逐点卷积按系数 1 或 6 扩展输入特征图通道数, 对扩展后的特征图应用批归一化和 Swish 激活函数, 利用深度可分离卷积提取特征并引入 SE 模块提升网络特征表达能力, 再通过 1×1 逐点卷积恢复特征图通道数, 最后经跳跃连接和 Dropout 机制实现特征融合。

3.2.2 局部细节优化模块

本章核心任务是基于患者 CT 影像序列中肿瘤区域面积最大帧提取关键特征以判断嗜铬细胞瘤基因亚型。临床数据中不同患者肿瘤区域大小差异显著，Xiangya-PPGL 数据集肿瘤最大帧面积比例分布如图 3-3 所示，48%患者肿瘤区域占比 0%~5%，27%患者肿瘤区域占比 5%~10%，仅 25%患者肿瘤区域占比超 10%，这种多尺度特性使传统单一尺度卷积操作难以兼顾不同尺度特征信息，导致特征提取不充分影响分类性能。

CT 影像中小肿瘤边界模糊、信噪比低，传统方法特征提取易丢关键信息，深层网络更难捕捉微小目标细节，且肿瘤特征易被复杂背景、伪影及无关组织信息掩盖，增加有效特征提取难度。多尺度特征提取虽能从不同层次捕获局部与全局信息，但传统固定权重融合方法缺乏灵活性，无法动态调整各尺度特征贡献，难以适应肿瘤大小、形状和位置的高度差异，限制分类模型性能提升与应用。

为此本研究设计局部细节优化模块，由局部特征提取与动态门控融合组成（结构见图 3-4），采用多尺度卷积与动态权重分配机制，通过并行卷积分支提取不同感受野局部特征，结合动态门控网络自适应融合多尺度特征。

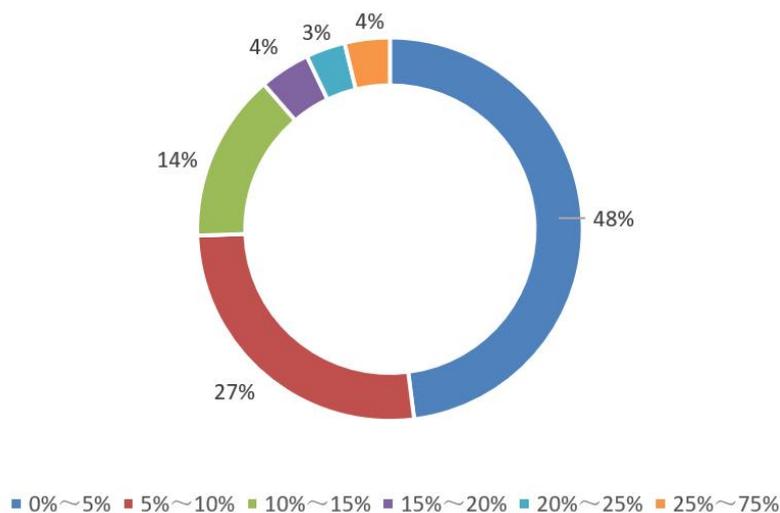


图 3-3 最大帧的肿瘤区域面积占比分布图

为了解决这些问题，本研究创新性地设计了局部细节优化模块。该模块主要由局部特征提取与动态门控融合组成，模块结构如图 3-4 所示。该模块采用多尺度卷积与动态权重分配机制，通过并行卷积分支提取不同感受野下的局部特征，并结合动态门控网络自适应地融合多尺度特征。

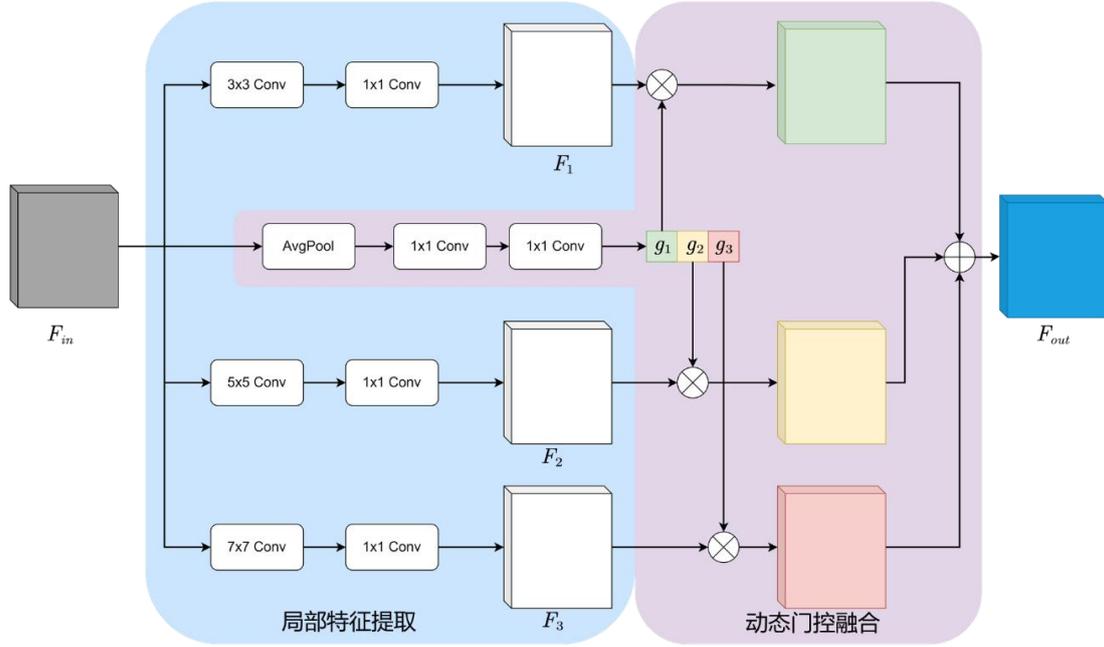


图 3-4 局部细节优化模块结构图

模块运用 3×3 、 5×5 和 7×7 卷积核分别获取小范围、中等范围和大范围信息，借助 1×1 卷积与批归一化操作强化特征表达能力。动态门控网络经全局平均池化和 **Softmax** 归一化生成各分支权重，实现多尺度特征加权融合。该设计大幅提升肿瘤区域特征密度，有效保留肿瘤形态学与空间结构特征，为基因亚型分类提供高质量的特征表示。

该模块主要由三个不同尺度的卷积路径和一个用于自适应权重计算的动态门控网络组成。每个卷积路径负责提取特定尺度的信息，动态门控网络根据输入数据自适应地调整这些路径输出的重要性权重，实现有效的特征融合。假设，输入的特征为 $F_{in} \in R^{B \times C \times H \times W}$ ，其中， B 表示批次大小， C 表示特征通道数， H 表示特征图的高， W 表示特征图的宽。三个不同尺度的卷积路径分别为 $[3 \times 3, 1 \times 1]$ 、 $[5 \times 5, 1 \times 1]$ 和 $[7 \times 7, 1 \times 1]$ 卷积融合运算。该进程可以描述如下公式所示：

$$F_1 = Conv_{1 \times 1}(Conv_{3 \times 3}(F_{in})) \quad (3-2)$$

$$F_2 = Conv_{1 \times 1}(Conv_{5 \times 5}(F_{in})) \quad (3-3)$$

$$F_3 = Conv_{1 \times 1}(Conv_{7 \times 7}(F_{in})) \quad (3-4)$$

其中， F_1 ， F_2 ， F_3 分别代表第一、二、三分支的输出特征，每个分支的输出与输入具有相同的空间维度 $H \times W$ 和通道数 C 。接着通过动态门控网络来对三个分支自适应的生成权重进行融合。首先，应用全局平均池化（Global Average Pooling, GAP）将输入特征图压缩为 $[B, C, 1, 1]$ 的形式。然后，通过一个 1×1

卷积层将通道数减少到 $C/4$ 。接着应用 ReLU 激活函数，再通过另一个 1×1 卷积层将通道数映射到 3，表示三个卷积路径的权重。最后，应用 Softmax 函数对这三个权重进行归一化处理。该进程可以描述如公式 3-5 所示：

$$g = \text{Softmax}(\text{Conv}_{1 \times 1}(\text{Conv}_{1 \times 1}(\text{GAP}(F_m)))) \quad (3-5)$$

其中， g 是形状为 $[B, 3, 1, 1]$ 的张量，表示即将与三个卷积路径相乘的权重。根据动态门控网络生成的权重，对三个卷积路径的输出进行加权求和，得到最终的特征表示，过程如公式 3-6 所示。

$$F_{out} = g_1 \times F_1 + g_2 \times F_2 + g_3 \times F_3 \quad (3-6)$$

其中， g_1 ， g_2 ， g_3 分别是动态门控网络得到的权重。

本模块结合多尺度卷积操作与动态门控机制，通过捕捉不同尺度细节信息并依据输入数据自适应调整各尺度特征重要性，显著提升输入特征图表达能力，进而提高模型泛化能力与表现力。

3.2.3 跨层级特征融合增强模块

在医学影像分析领域，基于 CT 影像的肿瘤基因亚型识别面临多尺度特征判别性表征失衡的核心难题，小尺度肿瘤病灶采用传统单层次特征深度学习方法效果往往较差，如 EfficientNet 等深层卷积神经网络提取高层特征时，因感受野指数级扩展和空间下采样操作导致特征图分辨率大幅降低，小尺度肿瘤区域在高层特征中仅占极少数像素甚至被背景噪声淹没，使模型难以捕捉小肿瘤区域关键生物学特性，尤其在肿瘤边界模糊或形态异质性显著时分类性能明显下降。目前主流方法多采用端到端分类范式，主要依赖网络末端高层语义特征进行分类决策，深层特征虽能反映肿瘤宏观形态学特性，但空间信息丢失导致对微小结节表征能力不足，而浅层网络输出的特征图包含丰富局部细节信息（如微钙化、边缘锐利度及纹理异质性），对嗜铬细胞瘤等形态多样的肿瘤基因亚型鉴别至关重要，现有研究利用浅层特征存在明显不足：（1）浅层特征语义抽象层级较低，难以与高层特征直接对齐；（2）缺乏有效的跨层级特征交互机制，导致多尺度特征互补优势未能充分发挥，这种单层次特征利用策略面对复杂形态肿瘤时分类精度存在明显理论瓶颈。

针对上述问题本研究提出跨层级特征融合增强模块（Cross-level Feature Fusion Enhancement, CFFE），该模块核心在于构建浅层细节特征与深层语义特征的互补表达，以此显著提升模型对小尺度肿瘤及形态异质性病例的鉴别能力，为解决医学影像分析中肿瘤基因亚型识别难题提供新思路与新方法。

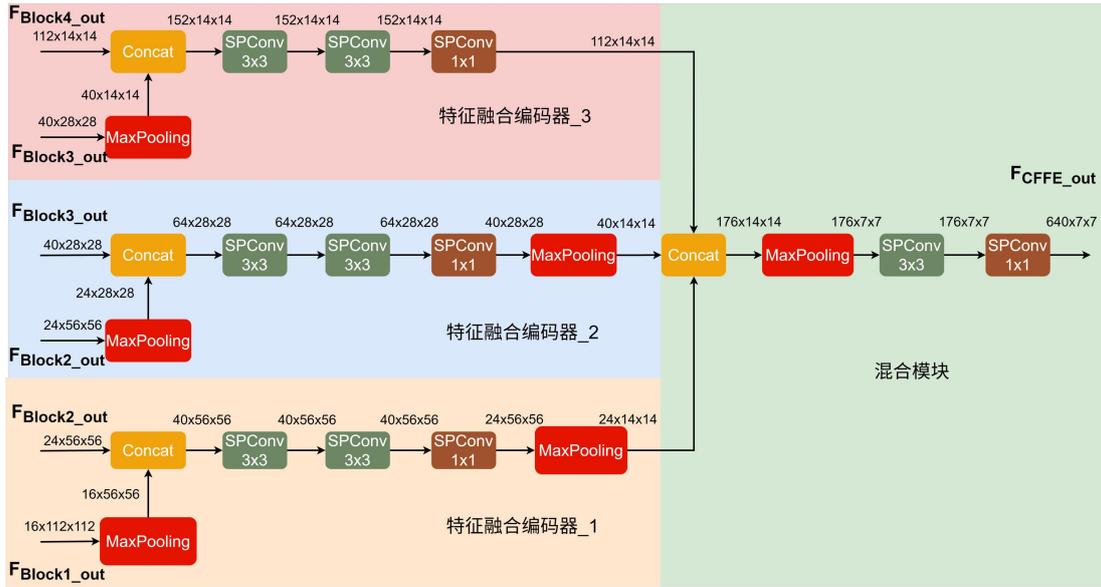


图 3-5 跨层级特征融合增强模块结构图

跨层级特征融合增强模块结构如图 3-5 所示。首先，对 EfficientNet-B0 相邻两个 Block 输出的特征图进行两两融合，通过特征融合编码器（Feature Fusion Encoder）生成三个中间特征图。这一设计能够有效结合相邻 Block 的特征信息，既保留了浅层 Block 提取的肿瘤纹理和边缘等细节特征，又融入了深层 Block 的提取到的全局语义特征。最后，通过设计的混合模块对三个中间特征图进行特征融合，生成具有更强表征能力的最终输出特征图。

特征融合编码器结构图如图 3-6 所示。假设特征融合编码器输入的浅层特征为 $F_1 \in R^{B \times C_1 \times H \times W}$ ，深层特征为 $F_2 \in R^{B \times C_2 \times \frac{H}{2} \times \frac{W}{2}}$ 。首先是通过最大池化（MaxPooling）进行下采样，将浅层特征 F_1 的空间尺寸统一到与深层特征 F_2 相同的尺寸。接下来，将下采样后的浅层特征与深层特征进行通道拼接，来获得融合后的特征图 F_{concat} ，具体操作如公式 3-7 所示：

$$F_{concat} = Concat(F_2, MaxPool(F_1)) \quad (3-7)$$

其中， $F_{concat} \in R^{B \times (C_1 + C_2) \times \frac{H}{2} \times \frac{W}{2}}$ ，Concat 表示通道拼接操作。接下来，对拼接后的特征图 F_{concat} 利用 SPConv（Sparsity-aware Convolution）^[49] 进行卷积处理，以进一步提取和融合特征。

SPConv 是面向稀疏数据设计的卷积神经网络结构，可高效利用数据稀疏性，通过增强模型鲁棒性、减少特征图模式冗余显著提升网络泛化能力，相比传统卷积操作，其以更智能计算方式优化相似内容，避免无效计算，增强对输入数据噪声和变化的容忍度。

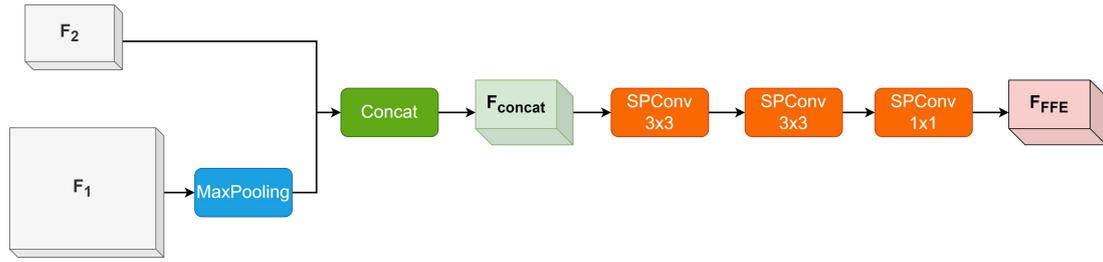


图 3-6 特征编码器结构图

在标准的卷积操作中，输入特征图 X 和卷积核 W 的输出 Y 通常是通过遍历整个特征图进行计算的，公式如 3-8 所示：

$$Y_{i,j} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X_{i+m,j+n} W_{m,n} \quad (3-8)$$

其中， i, j 表示卷积核的索引， m, n 表示特征图的坐标索引。这种方法虽然直观，但对于稀疏输入数据，它并未有效利用其中的大量零元素，导致了计算资源的浪费。SPConv 通过引入稀疏感知机制，只对输入特征图中的非零元素进行计算，从而大幅提高了计算效率。其稀疏卷积操作如公式 3-9 所示：

$$Y_{i,j} = \sum_{(m,n) \in \Omega(X)} X_{i+m,j+n} W_{m,n} \quad (3-9)$$

其中， i, j 表示卷积核的索引， m, n 表示特征图的坐标索引， $\Omega(X)$ 是输入特征图中非零元素的索引集合。这种仅计算非零元素的方式，不仅加快了运算速度，也减少了内存的占用。

SPConv 是针对稀疏数据设计的卷积结构，可高效利用数据稀疏性，大幅提升模型计算效率与泛化能力。本研究中肿瘤区域面积小、特征稀疏，传统卷积处理此类数据易出现冗余计算，SPConv 则通过智能优化相似内容避免无效计算，增强对噪声和变化的鲁棒性。本研究利用 SPConv 融合神经网络临近特征图，因相邻特征图相似性高，SPConv 通过特征共享复用减少模式冗余，进一步提升计算效率。该机制既增强模型应对肿瘤形态变化的稳定性，也为医学影像分析提供高效鲁棒方案，使其在处理小肿瘤区域和稀疏特征时表现优异。

在完成特征图的通道拼接后，将应用 SPConv 卷积操作以进一步提炼特征，并优化特征表示。先是对拼接后的特征图 F_{concat} 进行两次 3×3 大小的 SPConv 卷积，旨在捕捉更加精细的空间细节，并强化特征间的关联，然后通过 1×1 大小的 SPConv 卷积整合特征信息，并调整最终输出的通道数。具体过程如下，计算公式如 3-10 所示：

$$F_{FFE} = SPConv_{1 \times 1} (SPConv_{3 \times 3} (SPConv_{3 \times 3} (F_{concat}))) \quad (3-10)$$

这一系列卷积操作通过减少特征图特征冗余、增强模型对输入数据噪声和变化的不敏感性，提高鲁棒性与泛化能力。鉴于 EfficientNet 相邻阶段特征图携带信息高度相似，且层内特征图具有相似但不完全相同的特征，其冗余难以直接识别和去除，故采用 SPConv 作为主要卷积操作，以有效识别处理特征图冗余特征、捕捉更本质特征，通过降低冗余特征干扰，提升模型对输入数据噪声和变化的不敏感性，进而优化模型性能。

经过三路特征融合编码器得到的特征图分别为 $F_{FFE1} \in R^{112 \times 14 \times 14}$ ， $F_{FFE2} \in R^{40 \times 14 \times 14}$ ， $F_{FFE3} \in R^{24 \times 14 \times 14}$ ，这些特征图将被传送至混合模块进行特征融合，以提取包含多层次信息的高级特征图 $F_{CFE} \in R^{640 \times 7 \times 7}$ 。混合模块结构如图 3-7 所示。首先，将经过特征融合编码器输出的三个特征图进行通道拼接操作，接着将其通过最大池化层进行下采样，将特征图的空间维度由 14×14 变为 7×7 ，然后通过 3×3 的 SPConv 卷积操作进行特征融合，最后再通过 1×1 的 SPConv 卷积层进行通道调整。最终，通过这一系列卷积操作，得到输出特征图 $F_{CFE} \in R^{640 \times 7 \times 7}$ ，为后续的任务提供了融合后的多层次特征信息。

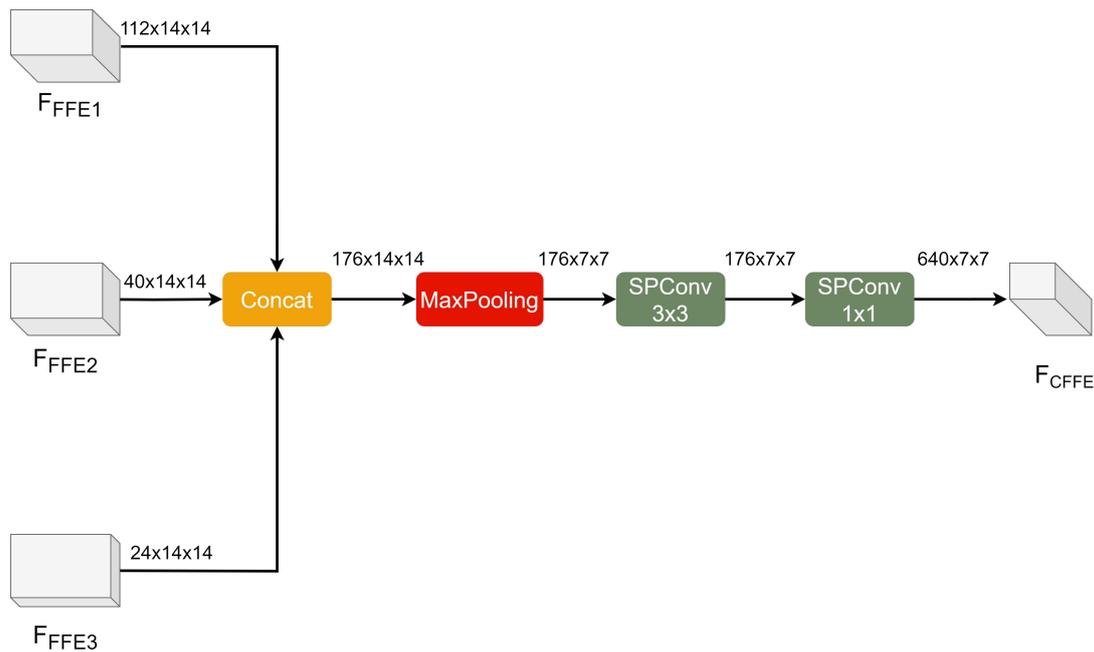


图 3-7 混合模块结构图

本模块借助跨层级特征融合机制整合浅层细节特征与深层语义特征，生成兼具局部细节和全局语义的判别性特征向量，同时通过混合模块对多尺度特征进行自适应融合，进一步增强特征的鲁棒性与表征能力。

3.2.4 末端注意力模块

图像分析领域提升模型性能的关键在于使其聚焦关键信息, 针对肿瘤影像这类结构复杂、特征多尺度的数据, 传统卷积神经网络存在局限性。其固定的权重和感受野无法依据图像不同区域重要性动态调整特征提取策略, 本研究中特征图融合低维空间与高维语义信息更增加了模型判别特征重要性的复杂性, 传统卷积结构和固定卷积核难以有效识别强化关键区域特征, 导致模型表现欠佳。

标准卷积依赖固定卷积核和权重分配, 不能根据输入图像实际需求调整计算重点, 处理带噪声或含细小结构图像时灵活性和鲁棒性不足。在肿瘤 CT 影像中, 标准卷积难以突出关键区域, 易忽略肿瘤区域微小重要细节, 而肿瘤形态位置多变、边界模糊的特点, 更使传统卷积网络特征提取面临巨大挑战。

为此本研究在网络末端引入基于小波变换^[50]大核卷积的注意力模块, 通过大核卷积实现全局信息感知并动态调整特征图权重分布, 提升模型应对噪声和形态变化的鲁棒性。小波变换具备多分辨率分析能力, 可在不增加参数复杂度的情况下获得更广泛感受野, 相比传统大核卷积能更灵活捕捉不同尺度特征信息, 处理复杂结构影像时更有效识别关键区域细节, 且其平移不变性和局部化特性可在保证特征表达丰富性的同时降低计算开销和过拟合风险, 该模块结构见图 3-8。

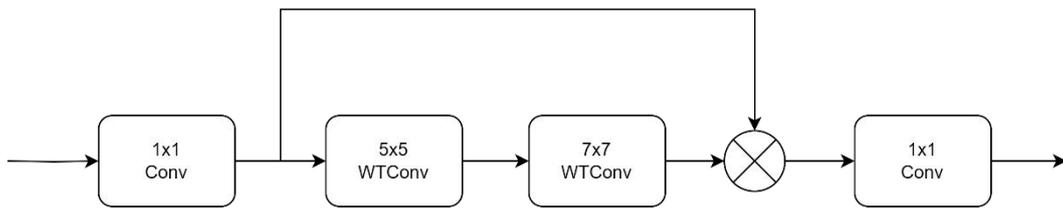


图 3-8 末端注意力模块结构

模块的第一层为一个 1×1 标准卷积层。该层的作用是通过每个像素点的特征进行线性组合, 调整输入特征图的通道数, 同时保持图像的空间尺寸不变。假设输入为 $X \in R^{H \times W \times C_m}$, H 和 W 分别表示图像的高度和宽度, C_m 为输入特征图的通道数。输出为 $X_1 \in R^{H \times W \times C_1}$, 其中, C_1 是卷积后的通道数。模块的第一层是一个 1×1 卷积层, 该层的计算公式如 3-11 示:

$$X_1 = Conv_{1 \times 1}(X) \quad (3-11)$$

其中, $Conv_{1 \times 1}$ 表示 1×1 卷积操作。第二层为 5×5 的小波变换卷积 (WTConv)。在 WTConv 的结构中, 输入特征图 X 经由小波变换 W 分解为不同频率分量。假设小波变换操作对输入图像 X 进行二维小波变换得到低频和低频分量, 计算公式如下所示:

$$X_{low} = W_{low}(X) \quad (3-12)$$

$$X_{high} = W_{high}(X) \quad (3-13)$$

其中, X_{low} 表示低频分量, 通常包含图像的全局信息和大范围特征, 而 X_{high} 包含图像中的高频信息, 如纹理、边缘等细节特征。通过这些分量输入到卷积网络中, WTConv 能够在不同频段上对图像特征进行进一步的提取与融合。对于每个频率分量, WTConv 利用传统的卷积核 W_{conv} 进行处理, 公式表示如下所示:

$$Y_{low} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X_{low, i+m, j+n} W_{conv, m, n} \quad (3-14)$$

$$Y_{high} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X_{high, i+m, j+n} W_{conv, m, n} \quad (3-15)$$

其中, i, j 表示卷积核的索引, m, n 表示特征图的坐标索引。最终, WTConv 将这些多尺度的特征信息融合, 生成最终的输出 Y , 作为肿瘤分类任务的输入特征。通过这种方式, WTConv 能够同时处理图像的细节与全局信息, 增强了对多尺度肿瘤特征的敏感性, 尤其是在面对肿瘤区域的大小、形态以及位置变化时, 能够有效提升模型的分类性能。

第二层为 5×5 的小波变换卷积通过小波变换进行多尺度特征提取, 能够有效捕捉局部特征以及全局信息。小波变换卷积在不增加过多参数的情况下, 能够扩展网络的感受野, 并提高对细节特征的捕捉能力。假设输入为 $X_1 \in R^{H \times W \times C_1}$, 输出为 $X_2 \in R^{H \times W \times C_2}$, 计算公式如 3-16 所示:

$$X_2 = WTConv_{5 \times 5}(X_1) \quad (3-16)$$

其中, $WTConv_{5 \times 5}$ 表示应用 5×5 小波变换卷积操作。第三层为 7×7 的小波变换卷积, 其功能与第二层类似, 但采用更大的卷积核, 进一步扩展感受野以捕捉更加广泛的上下文信息。假设输入为 $X_2 \in R^{H \times W \times C_2}$, 输出为 $X_3 \in R^{1 \times 1 \times C_3}$, 计算公式如 3-17 所示:

$$X_3 = WTConv_{7 \times 7}(X_2) \quad (3-17)$$

其中, $WTConv_{7 \times 7}$ 表示应用 7×7 小波变换卷积操作。最后一层是一个 1×1 的卷积层, 用于对经过前几层处理的特征图进行整合和通道数压缩。假设输出为 $Y \in R^{H \times W \times C_{out}}$, 其中, C_{out} 是输出的通道数。计算公式如 3-18 所示:

$$Y = Conv_{1 \times 1}(X_3 \cdot X_1) \quad (3-18)$$

本模块是针对融合后特征图兼具浅层细节信息与高层语义信息的特点设计

末端注意力模块,通过注意力机制动态调整特征权重使模型更聚焦重要特征以提升分类性能。模块结合标准卷积与小波变换卷积,利用不同尺寸卷积核扩展感受野,其中小波变换卷积可在不显著增加计算复杂度的情况下有效捕捉肿瘤区域关键细节特征。此外模型最后一层引入的注意力机制进一步提升了其在低对比度、多噪声影像中的鲁棒性,显著提高图像分类任务的精度与稳定性。

3.3 实验结果与分析

3.3.1 实验配置

实验配置与 2.3.3 节一致。

3.3.2 对比实验

实验定量评估结果见表 3-1、3-2,各指标最优结果已加粗,表中“-”表示无对应值。实验采用综合性指标 AUC、准确率 (Accuracy) 及单项评价指标敏感度 (Sensitivity)、特异性 (Specificity) 比较方法性能,各指标含义与计算方法在 2.3.2 节已说明。需说明的是,敏感度和特异性虽起源于疾病筛查领域,但其数学定义本质为与类别无关的统计度量,本章沿用机器学习领域通用方法,将非假性缺氧型 (Non-Cluster 1) 视为阳性类计算敏感度,将假性缺氧型 (Cluster 1) 视为阴性类计算特异性,以此分别量化不同方法对两类基因亚型的鉴别能力。

表 3-1 Xiangya-PPGL 数据集上各深度学习模型的实验结果

Model	AUC	Accuracy	Specificity	Sensitivity
ConvNeXt-T ^[51]	0.6087	0.6163	0.6032	0.6146
ViTamin-S ^[52]	0.6029	0.5723	0.5786	0.5703
BiFormer-T ^[53]	0.8476	0.7675	0.7067	0.8063
MobileNetV3 ^[54]	0.7442	0.6516	0.6129	0.6625
ShuffleNetV2 ^[55]	0.6961	0.6334	0.6114	0.6375
EfficientNet-B0 ^[56]	0.7978	0.7187	0.7476	0.7000
DenseNet-121 ^[57]	0.8653	0.7939	0.7067	0.8500
Ms-Mamba-T ^[58]	0.8924	0.8404	0.8357	0.8438
HiFuse-T ^[59]	0.9151	0.8708	0.8538	0.8813
TAMax-Net	0.9318	0.8831	0.8067	0.9325

(1) 与深度学习方法进行对比

为全面评估 TAMax-Net 分类性能,本研究利用与湘雅医院联合构建的临床数据集 Xiangya-PPGL 开展系统性验证,该数据集包含经严格筛选标注的 CT 影像数据,实验在 2.3.3 节所述硬件平台与软件环境中进行以确保结果可靠性和可重复性,所有对比模型均采用默认训练策略和超参数设置。在 Xiangya-PPGL 数

数据集分类任务中, TAMax-Net 表现显著优于现有方法, 准确率、AUC 和敏感度等关键指标均达最优水平, 验证了本研究方法的有效性。

相比之下, ViTamin 和 ConvNeXt 在本研究中的表现相对较弱。这主要是因为二者的网络架构设计针对大规模数据集, 其复杂的结构和大量可学习参数需要海量数据才能充分发挥性能。然而, Xiangya-PPGL 数据集包含 425 例患者的 CT 图像, 在深度学习领域属于小规模数据集。因此, 轻量级架构在有限数据下展现出更优的泛化能力。MobileNet、ShuffleNet 和 EfficientNet 作为专为移动端和边缘计算设计的高效模型, 在本研究中展现出更强的适应性。

BiFormer 的 AUC 值达 0.8476, 其独特双向注意力机制可同时建模 CT 图像局部细节与全局上下文信息, 这一特性与 CT 影像中肿瘤区域多尺度特征高度契合, 与 TAMax-Net 局部细节优化模块在本章任务中的优势相似, 本研究通过局部细节优化模块进一步增强细微特征提取能力, 在医学图像分析任务中表现更优。

2017 年提出的 DenseNet 在本研究中表现出色, 其核心优势是采用“密集连接”结构, 使每一层输出直接连接至所有后续层以实现高效信息流动与特征传播, 该设计提升了特征复用能力, 特别适用于医学图像分类任务(医学影像病变区域通常较小, 需充分利用多层次特征信息)。本文提出的跨层级特征融合增强模块在设计理念上与 DenseNet 具有一定相似性, 即通过高效整合多分辨率特征图优化特征传播, 但本研究方法避免了层间直接连接, 降低训练中对中间特征图的存储需求、提高计算效率, 且末端注意力机制的引入使模型能动态调整特征权重, 弥补了 DenseNet 在特征选择灵活性方面的不足, 更精准聚焦于关键诊断特征以提高分类性能。

Ms-Mamba 在本研究中表现优异, 得益于其独特双分支结构和高效特征融合机制, 该模型结合卷积神经网络局部特征提取能力与状态空间模型全局特征建模能力, 其调制交互特征聚合模块能有效融合局部与全局信息以提升医学图像分析准确性。本文提出的 TAMax-Net 进一步优化这一思路: 首先, 跨层级特征融合增强模块高效整合多尺度特征, 既保留空间细节信息又提升特征判别能力; 其次, 末端注意力机制动态调整特征权重使模型更关注关键诊断区域; 最后, 局部细节优化模块增强对肿瘤边缘特征和内部异质性的表征能力, 使模型在医学图像分类任务中取得更优性能。

HiFuse 在嗜铬细胞瘤基因亚型分类任务中展现卓越性能和较高特异性, 其架构设计与医学影像特点高度契合, 针对 CT 影像噪声多、类内差异大、类间相似度高挑战, HiFuse 采用并行全局与局部特征提取模块分别提取肿瘤整体形态特征和局部细节特征, 并通过自适应层次特征融合在不同语义层次融合, 实现对病变区域多尺度特征全面捕捉, 其空间-通道双重注意力机制能自动聚焦关键

诊断区域并通过动态权重调整突出与基因亚型分类相关的特异性特征,自适应层次特征融合模块中的倒残差多层感知机进一步优化特征融合过程、使模型自适应强化不同层次特征的判别性表达。TAMax-Net 在此基础上进一步优化特征提取和融合策略:首先,局部细节优化模块结合多尺度特征提取和细节增强机制,有效提升模型对肿瘤边缘特征和内部异质的表征能力、提高病灶识别精度;其次,跨层级特征融合增强模块在不同分辨率特征图之间实现高效信息交互,保留空间细节信息的同时增强特征判别能力;最后,末端注意力模块通过动态调整特征权重使模型更精准聚焦关键诊断特征,因此 TAMax-Net 在医学图像分类任务中展现更优分类性能。

综上所述,CT 影像分析具有其独特的特点,直接应用在自然图像领域表现优异的模型到嗜铬细胞瘤影像分析任务上,未必能取得理想效果。只有结合医学影像的特性,进行针对性设计和优化,才能有效提升模型的分类性能。

(2) 与人类医生进行对比

表 3-2 Xiangya-PPGL 数据集上人类医生的实验结果

Model	AUC	Accuracy	Specificity	Sensitivity
Urologist(Senior)	-	0.6900	0.5000	0.7500
Urologist(Junior)	-	0.6500	0.6400	0.6600
Radiologist(Senior)	-	0.6900	0.6000	0.7200
Radiologist(Junior)	-	0.6200	0.5600	0.6500
Multidisciplinary Team	-	0.7100	0.6400	0.7300
TAMax-Net	0.9318	0.8831	0.8067	0.9325

为全面评估模型临床应用价值,本研究构建多学科专家评估体系,邀请湘雅医院放射科和泌尿科资深临床医师(均有 10 年以上临床经验)及多学科诊疗专家组对医学影像进行独立分类诊断。研究通过设计“AI-医师”对比实验范式,既评估模型分类性能又挖掘其在临床辅助诊断中的应用潜力,为 AI 技术医疗落地提供实证依据。其中放射科医生包括 1 名有超过 15 年肾上腺 CT 解读经验的高级医师和 1 名有 5 年以上肾上腺 CT 解读经验的初级医师,泌尿科医生包括 1 名有 10 年以上肾上腺外科经验的资深医师和 1 名有 3 年以上肾上腺外科经验的初级医师,多学科专家团队还包括 1 名有 30 年以上肾上腺 CT 解读经验的高级放射科医师、1 名在肾上腺内分泌疾病领域有超过 15 年经验的高级内分泌学家及 1 名在肾上腺手术方面有超过 15 年经验的资深泌尿科医师。实验组对 PPGL 肿瘤分类诊断时所有成员均未接触组织病理学结果及临床资料,以确保评估独立性与客观性,所有临床医生均根据临床经验对肿瘤进行分类。

实验结果表明,TAMax-Net 的分类性能显著优于人类放射科医生、泌尿科医生及多学科专家组的诊断水平。值得注意的是,TAMax-Net 在保持高准确率

的同时,其特异性和敏感度指标也显著优于人类医生。实验结果表明,TAMax-Net在辅助临床诊断、提高诊断效率和减少人为误差方面具有显著优势,有望成为放射科医生的重要决策支持工具。

3.3.3 消融实验

为了分析 TAMax-Net 中各个组件的重要性,本研究在基准模型(EfficientNet-B0)中逐步应用 LDO、CFE 和 TAM 来验证它们的有效性。消融实验在 Xiangya-PPGL 数据集上进行,采用五折交叉验证的方式测试性能。表 3-3 显示了增加或减少每个模块对评估指标的影响,其中√表示使用模块,×表示不使用模块,各指标最优结果均已加粗。

表 3-3 各个模块在 Xiangya-PPGL 数据集上的消融实验

Exp.	LDO	CFE	TAM	AUC	Accuracy	Specificity	Sensitivity
1	×	×	×	0.7978	0.7187	0.7476	0.7000
2	√	×	×	0.8557	0.7425	0.7124	0.7500
3	×	√	×	0.8498	0.7411	0.7562	0.7313
4	×	×	√	0.8093	0.7226	0.7562	0.7104
5	√	√	×	0.9182	0.8473	0.6419	0.9125
6	√	×	√	0.8958	0.8382	0.8095	0.8438
7	×	√	√	0.8300	0.7186	0.7857	0.6938
8	√	√	√	0.9318	0.8831	0.8067	0.9325

(1) LDO

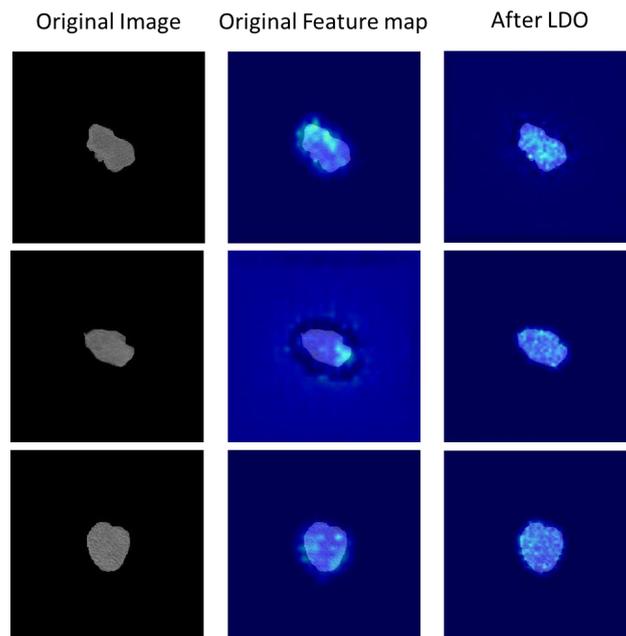


图 3-9 LDO 对特征提取的影响

实验结果（表 3-3）显示，基准模型引入局部细节优化模块（LDO）后检测性能显著提升，AUC 从 0.7978 升至 0.8557、Sensitivity 从 0.7 升至 0.75、Accuracy 从 0.7187 增至 0.7425，这主要得益于 LDO 模块的多尺度特征提取机制使其更有效捕捉局部细节特征、明显增强小尺寸肿瘤识别能力，但 Specificity 从 0.7476 降至 0.7124，可能因模块强化局部特征学习时放大图像噪声和伪影干扰导致部分误判。为进一步验证 LDO 对模型特征学习的影响，本研究在图 3-9 中对模型应用 LDO 前后的注意力分布进行可视化对比，热力图中颜色越亮表明模型对该区域关注度越高，分析显示引入 LDO 后模型注意力机制更聚焦细粒度特征，表明该模块有效提升模型对微小结构和关键细节的感知能力。

（2）CFFE

表 3-3 实验结果显示，引入 CFFE 后所有评估指标显著优化，准确率从 0.7187 提升至 0.7411、敏感度从 0.7 改进至 0.7313，这证明 CFFE 通过整合浅层细节特征与深层语义特征构建互补性表达方式、形成更全面鲁棒的特征表示：浅层细节特征保留图像高分辨率空间信息、精确捕捉肿瘤边缘和纹理等局部特性，对提高敏感度和特异性至关重要；深层语义特征提供高层次语义理解能力、解析肿瘤整体结构，显著提升整体准确率及 AUC。这种互补性融合机制通过自适应加权策略实现，使特征融合编码器可根据输入数据特性动态调整浅层和深层特征的贡献比例，不仅增强模型对不同尺度和类型肿瘤的认识能力，还确保特征表示的多样性和准确性。

（3）TAM

表 3-4 显示添加 TAM 可改善所有评估指标，本文还研究不同常见注意力机制对模型的影响，对比空间注意力（Spatial Attention, SA）^[60]、SENet（Squeeze-and-Excitation Network）^[61]以及 CBAM（Convolutional Block Attention Module）^[62]并简要介绍如下：空间注意力通过学习空间权重矩阵调整不同空间位置特征响应以强调特定空间位置、增强对关键区域的关注；SENet 作为典型通道注意力机制，通过挤压操作将二维特征图转换为单个值，再经激励操作计算特征图重要性权重实现动态重标定；CBAM 结合通道与空间注意力，通过串行方式实现双重注意力机制以帮助模型更有效聚焦重要信息。注意力机制消融实验结果（表 3-4，各指标最优结果已加粗）显示，空间注意力机制因强调特定空间位置但未有效利用全局信息，各方面指标相对较低；SENet 通过压缩和激励操作提升模型性能（Accuracy 0.8327、AUC 0.9063、Sensitivity 0.8750），但因局限于单一尺度特征增强，对复杂图像多尺度信息捕捉能力不足；CBAM 引入空间和通道联合注意力机制，在 Accuracy 和 Sensitivity 上显著提升，其多尺度特征提取优势可对不同尺度特征联合建模、提高肿瘤图像分析中的细粒度特征识别能力；

TAM 由多个大核卷积组合而成, 通过动态调整特征权重扩展模型对多尺度特征的感受野, 且引入全局自注意力机制显著增强对全局信息的捕捉能力, 在 Accuracy、Sensitivity、Specificity 和 AUC 上均达最高水平, 尤其在全局信息提取上超越传统卷积层局限、拓宽感受野, 使模型面对多样性和复杂性特征时仍保持高效分类性能。

表 3-4 注意力机制消融实验结果

Method	AUC	Accuracy	Specificity	Sensitivity
SA	0.8986	0.8162	0.7543	0.8375
SENet	0.9063	0.8327	0.7671	0.8750
CBAM	0.9263	0.8559	0.7971	0.8938
TAM	0.9318	0.8831	0.8067	0.9325

(4) 损失函数

根据 2.2.3 节的数据分析结果, 本研究所用数据集存在类别不平衡问题, 阳性样本与阴性样本比例约为 3:1。为缓解类别分布不均对模型性能的影响, 本研究采用加权交叉熵损失函数 (Weighted Cross-Entropy Loss, WCE Loss) 作为解决方案。该方法的选取基于以下考虑: 首先, 本研究数据集仅包含 425 例患者样本, 若采用欠采样策略会进一步减少本已有限的训练数据; 其次, 由于研究对象为肿瘤 CT 影像, 常规的图像增强方法如几何变换、颜色空间调整等, 可能改变病灶的形态学特征, 影响医学影像的病理相关性; 最后, 合成少数类的过采样技术 (如 SMOTE) 在图像数据上的应用效果有限。相较而言, 通过损失函数加权的方式能在不改变原始数据分布的前提下, 有效提升模型对少数类样本的关注度, 因此被选为本研究的最佳解决方案。

损失函数的消融实验结果如表 3-5 所示, 最优结果已加粗。第一行为使用普通交叉熵损失函数 (Cross-Entropy Loss, CE Loss) 的结果, 第二行为使用加权交叉熵损失函数的结果, 其中每个样本的权重为该样本所在类的数量的倒数。可以看出在使用加权交叉熵损失函数后, 算法的 Specificity, Accuracy, AUC 有所提升, Sensitivity 略微有所下降。说明 WCE Loss 可以帮助模型更好的识别阳性样本, 提高模型的整体表现。

表 3-5 损失函数消融实验结果

Method	AUC	Accuracy	Specificity	Sensitivity
CE Loss	0.9247	0.8604	0.7190	0.9510
WCE Loss	0.9318	0.8831	0.8067	0.9325

3.3.4 可视化分析

临床常规 CT 扫描目前主要依靠目视评估, 难以准确识别与 PPGL 基因突变相关的影像学特征, 但常规 CT 图像可能包含肿瘤异质性及肿瘤组织中蛋白质、基因等微观层面成分的代谢和生物学信息^{[19][63]}, 这些高维成像特征虽超出人眼识别能力, 却可通过深度学习计算分析发现隐藏信息, 且 PPGL 患者潜在的基因突变和基因亚型与临床表现密切相关, 先进影像学分析或在临床实践中发挥重要作用。本研究推测深度学习 CNN 框架识别的影像学特征可能潜在反映 PPGL 的肿瘤异质性和代谢, 已知 Cluster 1 型肿瘤会导致 HIF-2 α 稳定进而引发假性缺氧^[4], 肿瘤倾向于通过改善葡萄糖代谢、改变肿瘤外微环境^[64]适应缺氧, 且有报道称 HIF-2 α 可增加血管生成^[65] (一种已知与肿瘤异质性相关的病理特征), 这些变化会导致组织学水平肿瘤异质性改变并在 CT 图像上反映, 如图 3-10 所示, Cluster 1 型 PPGL 肿瘤边缘及内部高密度区 (平均 HU \approx 45) 与 CAM 强激活区域吻合, 模型可能捕获到血管密集分布的特征。

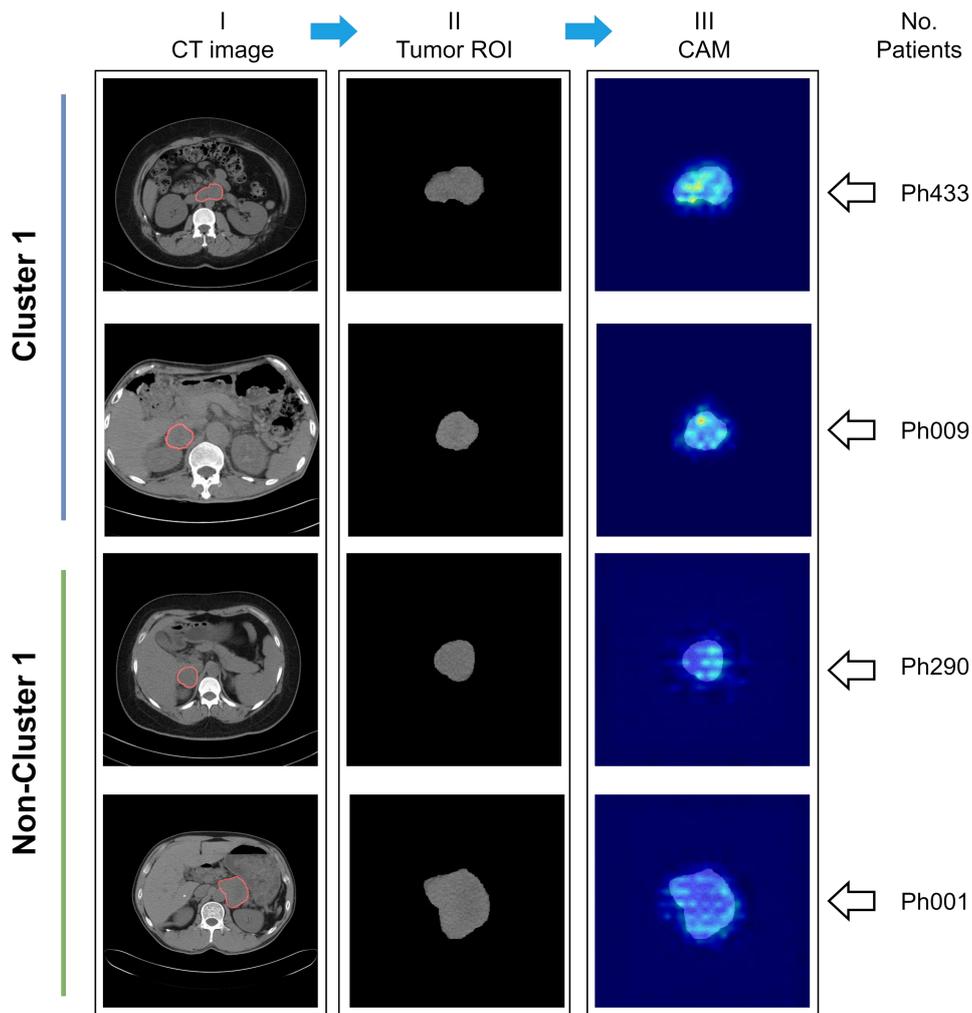


图 3-10 部分分类结果可视化展示

3.4 本章小结

本章节的主要研究目的是利用患者 CT 序列中肿瘤区域面积最大的一帧图像来识别患者 PPGL 的基因亚型。针对肿瘤尺寸差异显著和小面积占比条件下的嗜铬细胞瘤基因亚型分类难题，本章提出了 TAMax-Net 模型来为临床提供更精准的辅助诊断工具。具体而言，首先，通过 LDO 模块来提取多尺度特征，优化模型对多尺度融合能力和局部细节的感知能力，提高模型对不同尺寸肿瘤区域的适应性；其次，通过 CFFE 模块对不同语义信息的特征图进行融合，提高了最终分类用的特征图中的浅层信息，缓解低肿瘤面积占比图像在特征传递过程中的信息损失；此外，设计了 TAM 模块帮助模型在混合了浅层信息和深层信息的特征图中聚焦重要特征，进一步提高分类性能。通过对比实验和消融实验证明，该模型可以有效地解决 PPGL 基因亚型识别任务，有望为临床提供可靠的辅助诊疗。

第 4 章 MFF-Net: 基于肿瘤多帧融合网络的分类模型

第三章针对肿瘤区域面积最大帧的分类问题对 EfficientNet 模型进行创新性改进并提出 TAMax-Net 模型, 尽管模型性能显著提升, 但基于单帧的分析范式存在固有局限性: (1) 无法充分捕捉肿瘤三维空间分布信息; (2) 临床研究表明肿瘤区域面积最大的帧不能全面反映患者整体病情特征^[66], 这种基于单一帧的表征方法在临床适用性方面存在局限。

本章致力于探索基于多帧信息融合的肿瘤基因亚型识别方法。本章在 EfficientNet 架构的基础上进行创新性改进。研究重点解决了多帧信息融合过程中的两个关键科学问题: 其一, 多帧输入引发的特征退化现象; 其二, 肿瘤空间连续性信息的有效提取与表征。基于此, 本章提出了一种新的多帧融合网络 (**Multi-frame Fusion Network, MFF-Net**), 并且在自建数据集 Xiangya-PPGL 上进行实验, 验证本章提出 MFF-Net 模型的可行性和有效性。

4.1 概述

4.1.1 总体思路

深度学习技术在医学影像诊断尤其是疾病识别方面展现出巨大潜力, 传统方法常选特征和病变表现突出的单幅图像训练算法^[67]。但神经内分泌肿瘤等罕见病临床表现复杂、影像学缺乏典型特征, 仅靠单帧 CT 图像分析难以满足精确诊断需求^[68], 临床医生通常需综合分析多个 CT 切片制定诊疗方案, 因此计算机辅助识别系统需综合分析 CT 切片序列构建三维视图提升诊断准确性和可靠性。

尽管深度学习在图像与视频分类成果显著, 医学序列影像分析仍面临挑战。肿瘤 CT 序列图像与普通序列图像相比具有独特性: 医学影像采集特性使相邻切片存在高度空间相关性和信息冗余, 增加计算复杂度且可能降低模型泛化能力; 肿瘤空间异质性导致其不同切片影像表现差异大, 模型需具备捕捉空间连续性特征的能力。

基于上述情况, 本研究构建深度学习框架挖掘切片间空间关联信息, 该框架可自动识别提取关键特征并保留肿瘤形态学空间特征。研究提出改进的 EfficientNet-B0 架构, 针对 CT 序列中肿瘤区域面积小的问题, 在模型前端引入特征聚焦模块 (FFM), 采用特征驱动的自适应图像分析减少背景干扰, 利用 ROI-Align 机制^[69]增强特征提取的尺度不变性和一致性, 提升模型处理多尺度目标能力、改善特征提取质量。为挖掘多帧图像序列帧间关联特征, 设计空间特征提取模块 (SFE), 通过自适应机制生成肿瘤帧对象查询, 用自注意力机制提取

帧间关联特征, 经交叉注意力计算强化帧内特征, 依据原始分类器预测的每帧置信度对对象查询加权聚合降低低质量帧的干扰。为了消除混杂因素、捕捉切片序列时空连续性特征, 提出序列感知多帧融合模块 (SMF), 通过深度聚合对象查询生成判别性肿瘤特征表示。这些创新设计提升了模型诊断性能, 为医学影像序列分析提供新技术路线, 实验表明该方法显著提高诊断准确率, 为临床辅助诊断提供可靠技术支持。

4.1.2 模型总体结构

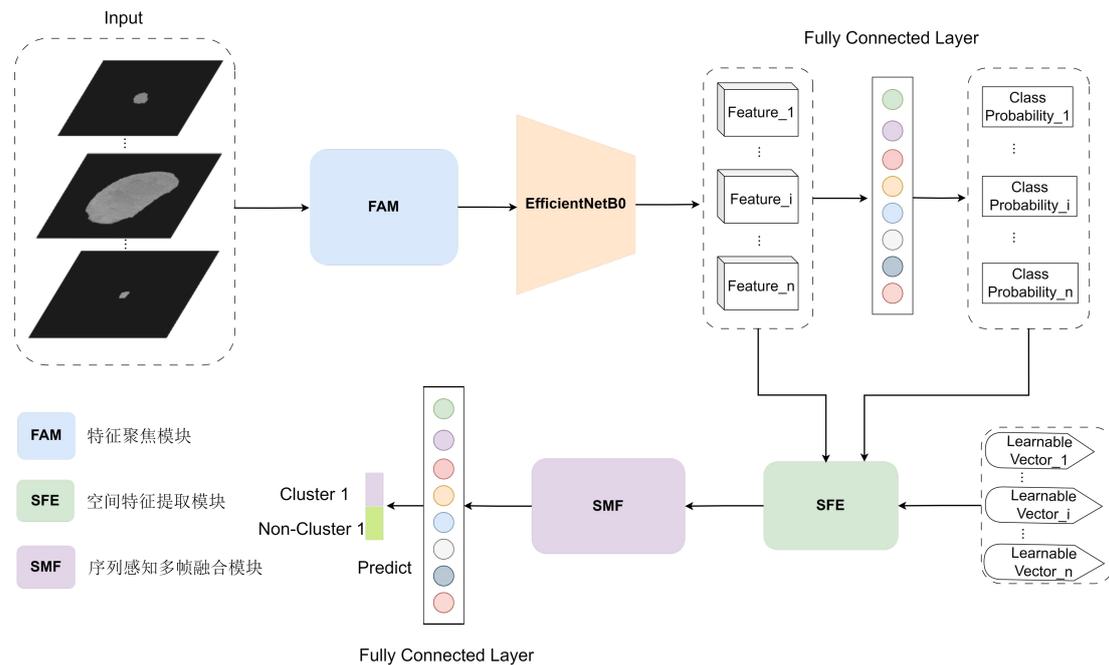


图 4-1 MFF-Net 网络结构图

整体网络流程结构图如图 4-1 所示, 具体设计与实现如下:

(1) 输入数据初始化处理

对输入的图像数据进行预处理。对每位患者的 CT 序列进行等间距抽帧, 每位患者抽取 8 帧, 并将每帧图像的尺寸统一调整为 224×224 像素大小。将预处理后的图像数据转换为张量形式, 作为输入数据馈入主干网络 MFF-Net, 以便进行后续的特征提取与分析。

(2) 帧级特征提取

基于 CT 图像中像素集中区域, 生成一个精确包围肿瘤区域的最小外接矩形框。对该矩形框内区域实施 ROI-Align 操作, 通过将不同大小肿瘤区域进行放大, 以强化模型对肿瘤特征的聚焦, 有效缓解因肿瘤区域面积较小而可能导致的帧退化问题。利用 EfficientNet-B0 网络提取图像特征, 获取其全连接层之前最后一层

的输出特征图及全连接层输出的类别概率，为后续患者级决策的生成奠定基础。

(3) 患者级特征提取

利用可学习的向量作为输入，每张图对应一个一维向量。考虑到帧质量问题，将每一帧的类别置信度与对应的可学习向量进行加权运算，并用自注意力机制处理所有帧的对象查询，有效传播空间上下文信息。利用交叉注意力机制使可学习向量聚焦于其对应帧的特征，学习与对象相关的视觉内容。

(4) 最终分类结果

将患者帧序列对应的可学习嵌入分别进行局部序列依赖提取和全局动态变化提取，其中局部序列依赖分支通过空洞卷积捕捉帧间长距离依赖关系，全局动态变化分支利用高比例随机失活增强特征鲁棒性，有效缓解序列数据的过拟合问题。最终将两个分支特征进行融合，经过全连接层得到患者的类别概率。

4.2 模型主要模块

4.2.1 特征聚焦模块

鉴于本章的核心任务是基于患者肿瘤 CT 序列的特征来实现基因亚型的分类，由第二章数据分析的结果可知，预处理后仍有 48% 的切片肿瘤面积占比在整张图的 5% 以下。本研究面临一个关键挑战：CT 序列中存在大量肿瘤区域占比极小的帧。直接对这些帧进行特征提取会出现特征稀疏，这主要体现在以下两个方面：

- (1) 由于有效特征区域有限，特征提取过程难以获得具有判别性的特征表示；
- (2) 特征分布的稀疏性导致模型难以建立有效的特征关联和空间连续性。尽管第三章提出的局部细节优化模块在一定程度上缓解了这一问题，但其主要针对肿瘤区域面积相对较大的单帧图像，无法有效处理 CT 序列两端肿瘤区域显著缩小导致的特征稀疏性问题。

为解决这一关键挑战，本研究在 EfficientNet-B0 前端创新性设计特征聚集模块，该模块采用特征驱动的自适应图像分析机制精确计算最小外接矩形框坐标，并在此基础上执行基于双线性插值的 ROI-Align 操作以确保特征提取的几何精度和空间一致性，通过 ROI-Align 放大肿瘤区域，这种设计不仅显著提升微小肿瘤区域的特征密度，还有效保留肿瘤形态学信息和空间结构特征，为后续基因亚型分类提供高质量特征表示。

(1) 特征驱动的自适应图像分析

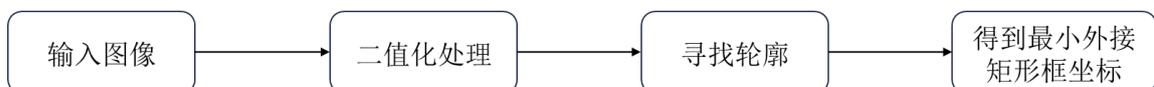


图 4-2 特征驱动的自适应图像分析流程图

特征驱动的自适应图像分析流程如图 4-2 所示。首先是输入图像，给定输入图像 $x \in R^{H \times W \times C}$ ，其中， H 表示输入图像的高度， W 表示输入图像的宽度， C 表示输入图像的通道数。因为图像本身为灰度图，所以不需要灰度化，可以直接进行二值化处理。

二值化是一种将灰度图像转换为黑白图像的技术，目的是为了更清晰地区分目标区域与背景。该过程需要设定一个阈值 T ，根据这个阈值将图像中的每个像素分为两部分：肿瘤区域和背景。公式如 4-1 所示：

$$I_{binary}(x, y) = \begin{cases} 1 & \text{if } I_{gray}(x, y) > T \\ 0 & \text{otherwise} \end{cases} \quad (4-1)$$

这里的 $I_{binary}(x, y)$ 表示二值化后的图像在坐标 (x, y) 处的像素值。选择合适的阈值 T 对于成功分割目标非常重要，这里的 $T = 125$ 。这一步骤大大简化了图像数据，使得轮廓检测等后续操作更加容易执行。

在完成二值化后，接下来是对图像中肿瘤区域进行轮廓检测。寻找轮廓实际上是识别出图像中所有连续的白色像素区域，这些区域的边界被称为轮廓。轮廓可以用来确定物体的形状和位置。每找到一个轮廓，就可以将其视为一系列有序的点集 C_i 。

一旦找到了所有的轮廓，下一步就是确定每个轮廓的最小外接矩形。最小外接矩形是指能完全包含某个轮廓的最小面积矩形。其主要用途在于精确定位和测量目标区域的位置和大小。通过分析轮廓的边界框，可以得到最小外接矩形的坐标。具体来说，对于轮廓 C_i 的点集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 可以通过以下步骤计算最小外接矩形，公式如下所示：

$$x_{\min} = \min(x_1, x_2, \dots, x_n) \quad (4-2)$$

$$x_{\max} = \max(x_1, x_2, \dots, x_n) \quad (4-3)$$

$$y_{\min} = \min(y_1, y_2, \dots, y_n) \quad (4-4)$$

$$y_{\max} = \max(y_1, y_2, \dots, y_n) \quad (4-5)$$

综上所述，通过上述四个步骤，该模块可以根据图像的特征来自适应的生成肿瘤区域最小外接矩形坐标，为后续 ROI-Align 操作提供精确的空间定位信息。

(2) ROI-Align

在计算机视觉领域，ROI-Align 技术作为目标检测和实例分割任务核心特征提取机制已获广泛认可。本研究创新地将 ROI-Align 机制引入医学图像分类任务的特征聚焦模块，用于精确提取肿瘤图像中的肿瘤区域并统一至标准尺寸。传

统应用中 ROI-Align 用于将不同大小特征图统一为指定大小, 鉴于肿瘤区域通常较小, 本研究依据特征驱动的自适应图像分析所得肿瘤区域最小外接矩形对肿瘤区域进行放大。相比直接放大原图, ROI-Align 能保证特征提取的几何精度与空间一致性, 比像素空间插值更稳健, 可使边缘和纹理等信息更连续。在本研究应用场景下, 创新性地将 ROI-Align 机制整合进分类网络的特征聚焦模块, 该模块先定位肿瘤区域, 再利用 ROI-Align 提取固定尺寸特征表示, 最后将特征输入后续分类网络, 此方法既保留肿瘤区域完整空间信息, 又通过特征统一化处理有效解决因肿瘤区域尺寸差异导致的特征退化问题。

与传统的特征提取方法不同, ROI-Align 避免了任何可能导致信息丢失的量化步骤。具体而言, 给定一个输入特征图 F 和一个位于原始图像上的 ROI, 首先需要将该 ROI 映射到特征图上。假设 ROI 在特征图上的坐标为 $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$, 并计划将其划分成 $k \times k$ 个 bins 用于池化操作。对于每一个 bin, ROI-Align 选择 $n \times n$ 个采样点 (通常 $n = 2$ 或 $n = 4$), 并使用双线性插值来计算每个采样点的特征值。假设某个采样点的坐标为 (x_p, y_p) , 那么该点处的特征值 $f(x_p, y_p)$ 可以通过以下公式 4-6 计算:

$$\begin{aligned} f(x_p, y_p) = & (1 - \Delta y) ((1 - \Delta x) f(x_0, y_0) + \Delta x f(x_1, y_0)) \\ & + \Delta y ((1 - \Delta x) f(x_0, y_1) + \Delta x f(x_1, y_1)) \end{aligned} \quad (4-6)$$

其中, $\Delta x = x_p - x_0$, $\Delta y = y_p - y_0$, x_0 , x_1 , y_0 , y_1 分别表示最近邻像素点的坐标。

此外, 为了进一步增强精度, ROI-Align 在每个 bin 内均匀地选取多个采样点, 并对这些点的特征值求平均作为最终输出。最终每个 bin 的输出计算公式如下 4-7 所示:

$$output_{bin} = \frac{1}{n^2} \sum_{i=1}^{n^2} f(x_{pi}, y_{pi}) \quad (4-7)$$

这种方法不仅减少了量化误差的影响, 而且提供了比简单上采样更为丰富的特征信息, 确保特征提取的几何精度和空间一致性, 比像素空间插值更鲁棒, 边缘和纹理等信息更连续。

4.2.2 空间特征提取模块

基于单帧图像的肿瘤基因亚型识别方法存在显著固有局限: 空间维度上, 单帧图像仅提供肿瘤二维截面信息, 无法完整呈现三维空间立体分布特征与生长模式, 易致关键诊断信息丢失; “最大肿瘤面积帧” 选择标准缺乏循证医学依据, 可能遗漏重要生物学特征。计算效率方面, 相邻切片因空间位置相近存在高度空间相关性与信息冗余, 增加计算开销且易引发模型过拟合, 肿瘤空间异质性使连

续切片影像学表现差异显著, 涵盖形态学、纹理和功能等多维度特征, 鲁棒的肿瘤基因亚型识别模型需有效捕捉肿瘤空间连续性特征变化规律。

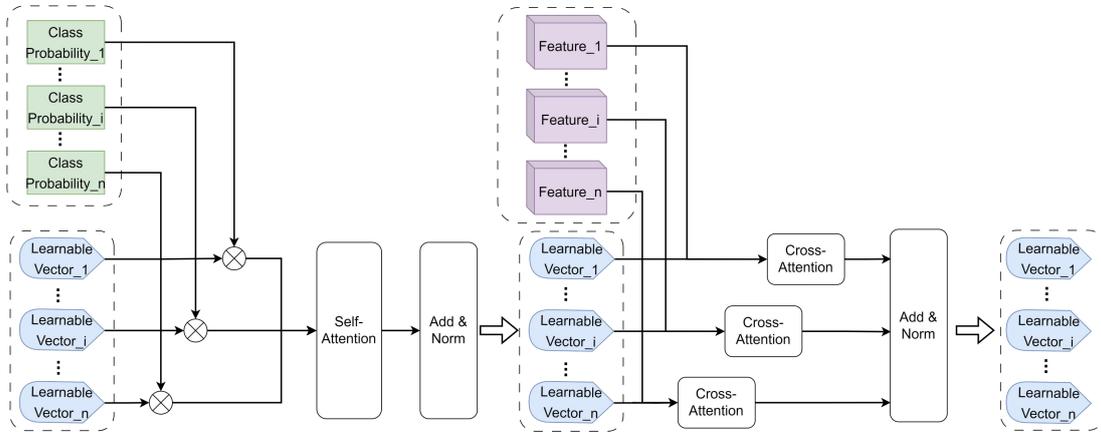


图 4-3 空间特征提取模块结构图

为此本研究提出空间特征提取模块, 以可学习向量为输入 (每张图对应一维向量), 将帧类别置信度与可学习向量加权运算解决帧质量问题, 用自注意力机制处理所有帧对象查询传播空间上下文信息, 引入交叉注意力机制使对象查询聚焦对应帧特征学习相关视觉内容, 该方法提升肿瘤特征捕捉精度, 增强模型适应性与泛化能力, 为精准医学提供有力支持。

空间特征提取模块结构如图 4-3 所示。具体而言, 假设输入为患者的 N 张 CT 图像, 经过主干网络 EfficientNet-B0 的特征提取后, 得到 N 张对应的特征图 $F = \{f_1, f_2, \dots, f_N\}$ 。其中, $f_i \in R^{C \times H \times W}$ 表示第 i 张 CT 图像的特征图, C 为通道数, H 和 W 分别为特征图的高度和宽度。同时, 主干网络还输出每张特征图对应的类别概率 $P = \{p_1, p_2, \dots, p_N\}$, 其中, $p_i \in R^1$ 表示第 i 张 CT 图像属于 Cluster 1 类别的概率分布。

为了进一步增强特征表示, 本研究引入了 N 个可学习的向量 $E = \{e_1, e_2, \dots, e_N\}$ 。其中, $e_i \in R^D$ 表示与第 i 张 CT 图像对应的嵌入向量, D 是嵌入向量的维度, 其具体值为 384。每个可学习的嵌入向量首先与对应特征图的类别概率进行加权融合, 具体公式如 4-8 所示:

$$e'_i = e_i \cdot p_i (i = 1, 2, \dots, N) \quad (4-8)$$

其中, $e'_i \in R^D$ 表示加权后的嵌入向量, \cdot 表示逐元素乘法。接下来, 所有加权后的嵌入向量 $E' = \{e'_1, e'_2, \dots, e'_N\}$ 通过自注意力机制进行计算, 以捕捉嵌入向量之间的全局依赖关系, 有效的传播空间上下文信息。计算过程如公式 4-9 所示:

$$e''_i = LN(e'_i + MHA(e'_i, E', E')) \quad (4-9)$$

其中, LN 表示层归一化, $MHA(e'_i, E', E')$ 表示使用查询 e'_i 、键 E' 和值 E' 进行的多头注意力操作, 在这里键和值的内容一致, 均为 E' 。通过计算, 得到更新后的嵌入向量 $E'' = \{e''_1, e''_2, \dots, e''_N\}$, 在自注意力层之后, 每个对象查询都通过整个 CT 序列的上下文信息进行增强。

随后, 更新后的嵌入向量 E'' 与对应的特征图 F 进行交叉注意力计算, 以融合嵌入向量与特征图之间的信息。计算过程如公式 4-10 所示:

$$e''_i = LN(e''_i + MHA(e''_i, f_i, f_i)) \quad (4-10)$$

通过交叉注意力计算, 建立语义表示与空间信息嵌入之间的相关性, 得到进一步更新的嵌入向量 $E''' = \{e'''_1, e'''_2, \dots, e'''_N\}$ 。

上述自注意力与交叉注意力计算过程重复四次, 最终得到优化后的可学习嵌入向量 $E_{final} = \{e_{final,1}, e_{final,2}, \dots, e_{final,N}\}$ 。这些嵌入向量不仅包含了每张 CT 图像的局部特征信息, 还通过全局依赖关系的建模增强了特征表示的判别能力, 以细粒度的方式融合了帧级和实例级特征, 为后续的基因亚型分类任务提供了高质量的特征输入。

4.2.3 序列感知多帧融合模块

在处理 3D 数据时多帧融合是通过整合多帧信息增强模型对输入数据理解能力的常见策略, 但其实际应用面临诸多挑战: 传统方法常假设帧间独立, 缺乏对时序依赖的建模, 这在医学影像分析中尤其不适用 (前后帧间往往存在显著因果关系, 简单拼接或平均融合会忽略细粒度时序信息、导致关键信息丢失), 且多采用静态特征提取和直接拼接, 难以捕捉帧间长距离依赖关系、无法有效建模全局动态变化, 序列数据中普遍存在的过拟合问题也进一步限制其在复杂动态场景中的表现, 因此有效建模帧间长距离依赖关系、捕捉全局动态变化并缓解过拟合是多帧融合领域亟待解决的关键挑战。

基于此本研究提出序列感知多帧融合模块, 该模块通过双路并行架构分别建模局部序列依赖和全局动态变化, 其中序列卷积分支利用空洞卷积捕捉帧间长距离的依赖关系, 动态感知分支通过高比例随机失活 (Dropout=0.5) 增强特征鲁棒性、有效缓解序列数据中的过拟合问题, 并在此基础上采用特征拼接与时域压缩策略实现了多维度特征的深度融合, 在保留帧间细粒度时序信息的同时有效控制计算复杂度, 该方法突破了传统多帧融合中时序建模不足的局限, 显著提升了模型对复杂动态场景的建模能力, 为视频分析、医学影像处理等时序敏感任务提供了新的解决方案。

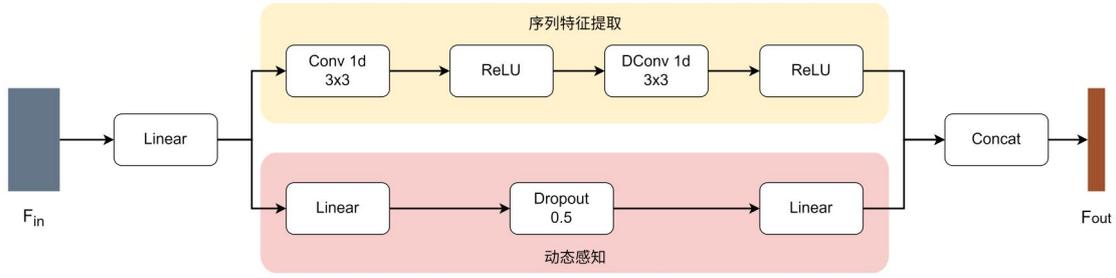


图 4-4 序列感知多帧融合模块结构图

SMF 模块的结构设计如图 4-4 所示。其核心目标是将原始输入特征映射到一个低维表示，并结合局部序列依赖和全局动态变化，生成具有更高表达能力的特征向量。该模块由特征投影层、序列建模分支和动态感知分支组成，其核心是通过双路并行架构实现时序特征的细粒度建模与高效融合。给定输入特征 $X \in R^{B \times N \times D}$, B 为批次大小, N 为帧数, D 为特征维度, 这里 D 设为 384, 输入特征首先经过线性投影层将特征维度压缩至 $d = D/2$, 公式如 4-11 所示:

$$H_1 = XW_1 + b_1 \quad (4-11)$$

其中, $W_1 \in R^{D \times d}$ 和 $b_1 \in R^d$ 为可学习参数。随后, 模块通过双路分支分别建模序列特征和全局动态特征。

序列建模分支采用膨胀卷积捕捉帧间长距离依赖特征, 公式如下所示:

$$H_2 = \text{ReLU}(\text{Conv1D}(H_1, K_1)) \quad (4-12)$$

$$H_3 = \text{ReLU}(\text{Conv1D}(H_2, K_2)) \quad (4-13)$$

其中, K_1 和 K_2 分别为卷积核, 第二层卷积采用膨胀率 $r = 2$ 以扩大感受野。

动态感知分支通过高比例随机失活 (Dropout=0.5) 增强特征鲁棒性, 缓解时序数据中的过拟合问题, 公式如下所示:

$$H_4 = \text{Dropout}(\text{ReLU}(H_1W_2 + b_2), p = 0.5) \quad (4-14)$$

$$H_5 = H_4W_3 + b_3 \quad (4-15)$$

其中, $W_2 \in R^{d \times d}$ 、 $b_2 \in R^d$ 、 $W_3 \in R^{d \times d}$ 、 $b_3 \in R^d$ 为可学习参数。最后, 模块通过特征拼接与时域平均操作实现特征的深度融合, 公式如 4-16 所示:

$$H_{out} = \text{Mean}(\text{Concat}(H_3, H_5), \text{dim} = 1) \quad (4-16)$$

该模块通过层级特征交互与动态信息融合, 有效解决传统多帧融合方法中时序依赖建模不足、动态特征丢失及计算效率低下的问题, 为视频分析、医学影像处理等时序敏感任务提供新解决方案, 且模块各部分可独立优化以增强模型灵活

性和可扩展性。

4.3 实验结果与分析

4.3.1 实验配置

本章研究基于患者的 CT 序列展开,本研究将采用等间距抽帧的方式来选帧。主要是考虑到等间距采样可确保图像帧在整个扫描范围内均匀分布,全面覆盖目标解剖结构,减少区域信息缺失或冗余,并且等间距采样具有确定性和可重复性。相比随机采样或基于病灶的选择性采样,该方法在保证数据代表性的同时,避免了复杂采样策略带来的额外计算成本,为模型优化提供了稳定的数据基础。

具体的抽帧流程如图 4-5 所示。对患者的 CT 序列进行等间隔抽取,共提取了八帧图像。在后续关于帧数的消融实验中,均采用等间隔抽帧的方法。如果遇到序列帧数不足的情况,则以纯黑图像进行填充替代。其余实验配置与 2.3.3 节描述保持一致。

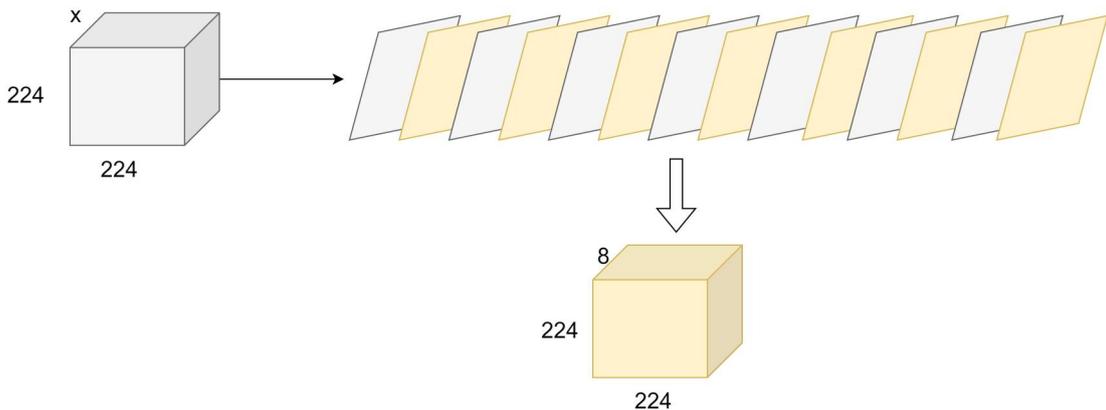


图 4-5 数据生成示意图

4.3.2 对比实验

为了全面评估 MFF-Net 的性能,本研究采用与湘雅医院联合构建的临床数据集 Xiangya-PPGL 进行验证。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。结合国内外研究现状可知,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。为进一步深入分析 MFF-Net 的性能优势,本研究从视频级分类的角度设计了对比实验。选取了近年来深度学习领域的代表性模型进行对比,全面验证 MFF-Net 的有效性和先进性。对比实验结果如表 4-1 所示,各指标最优结果均已加粗。

表 4-1 Xiangya-PPGL 数据集上各算法的实验结果

Model	AUC	Accuracy	Specificity	Sensitivity
MViT ^[70]	0.7849	0.7168	0.7667	0.6979
Video-Swin-T ^[71]	0.5461	0.5528	0.5714	0.5428
Suman 等 ^[40]	0.9143	0.8342	0.7967	0.8438
DUALPAT ViT-B/16 ^[72]	0.6183	0.5925	0.6063	0.5833
ZeroI2V ViT-B/16 ^[73]	0.6620	0.6083	0.7321	0.5625
TAda2D ^[74]	0.8837	0.7905	0.8224	0.7696
TAMax-Net	0.9318	0.8831	0.8067	0.9325
MFF-Net	0.9590	0.9267	0.8250	0.9583

表 4-1 实验结果显示 MFF-Net 在 AUC、Accuracy 和 Sensitivity 等关键指标上性能最佳，本文进一步探讨不同模型性能差异及潜在原因。

MViT 作为多尺度视觉 Transformer 通过构建多层次特征金字塔捕捉时空信息，但其未针对医学影像特点专门优化，在 Xiangya-PPGL 数据集表现一般（AUC 值 0.7849、准确率 0.7168，均显著低于 MFF-Net），可能因建模 CT 序列时未充分关注关键病灶区域且多尺度特征金字塔在医学影像分析中适应性有限，而 MFF-Net 通过特征聚焦模块精准捕捉肿瘤区域提升判别能力。

Video-Swin-T 作为基于 Swin Transformer 的视频分类模型在自然视频任务表现优异，但在本研究中 AUC 仅 0.5461、Accuracy 为 0.5528、Specificity 和 Sensitivity 分别为 0.5714 和 0.5428，明显低于其他方法，可能因其主要依赖大规模自然图像预训练数据优化，而医学影像数据特性与自然视频差异较大导致迁移能力受限，相比之下 MFF-Net 通过序列感知多帧融合模块有效建模医学影像序列时空关系并结合自适应加权策略减少低质量帧影响，整体性能大幅领先 Video-Swin-T。

DUALPAT ViT-B/16 和 ZeroI2V ViT-B/16 作为图像到视频的迁移学习方法，尝试在预训练图像 Transformer 基础上引入时序建模能力，但其 AUC 仅分别为 0.6183 和 0.6620、Accuracy 分别为 0.5925 和 0.6083，整体性能不及 MFF-Net。尽管通过适配器机制增强了一定时序建模能力，但医学影像中病灶空间位置、形态变化及帧间关联信息较自然视频更复杂，仅依赖通用适配器难以充分捕捉关键信息，且两者在 Sensitivity 指标上分别仅达 0.5833 和 0.5625，显示识别 Non-Cluster 1 型患者存在较大挑战，而 MFF-Net 通过空间特征提取模块强化帧间关联信息建模、结合对象查询机制增强特征表达，提升了分类可靠性。

Suman 等人提出的结合注意力机制的 CNN-LSTM 方法在实验中表现相对较好（AUC 0.9143、Accuracy 0.8342、Specificity 0.7967、Sensitivity 0.8438），表明 CNN 空间特征提取结合 LSTM 时序特征融合的策略在医学影像分析中有效，但相较 MFF-Net 仍有局限，可能因 LSTM 对长程依赖建模能力有限导致部分关键帧信息在时序建模中被弱化，且其注意力机制主要作用于帧间特征加权、

未充分考虑帧内特征空间关系，而 MFF-Net 引入自注意力机制并结合 FFM 和 SFE 提升特征表达能力，分类性能更优。

TAda2D 作为基于自适应卷积的时空建模方法在多个视频分类任务中表现优异，本实验中其 AUC 达 0.8837、Accuracy 为 0.7905、Specificity 和 Sensitivity 分别为 0.8224 和 0.7696，整体性能较好但仍低于 MFF-Net，可能因 TAda2D 主要依赖动态卷积核权重进行空间建模，而医学影像序列中病灶区域形态变化较大，仅通过自适应卷积建模易导致特征丢失，相比之下 MFF-Net 通过 SMF 增强帧间特征聚合鲁棒性、结合 FFM 聚焦关键区域，显著提升分类性能。

此外，本章提出的 MFF-Net 与第三章提出的 TAMax-Net 对比显示，MFF-Net 各项指标均更优——TAMax-Net 采用多种策略增强肿瘤区域特征提取，而 MFF-Net 采用标准 EfficientNet 进行单帧特征提取、通过空间特征提取模块和序列感知多帧融合模块实现多帧特征融合，在数据集上表现更优，表明多帧融合方法能利用更丰富信息，有效提升模型对 CT 影像中基因亚型特征的识别能力，相较单帧方法更具优势。

综上所述，现有模型在医学影像序列分类任务中存在小病灶区域识别能力不足、帧间时空建模能力受限等局限，而 MFF-Net 通过特征聚焦模块、空间特征提取模块和序列感知多帧融合模块的结合，实现精准病灶区域识别、高效时空特征建模及自适应帧间信息融合，在 Xiangya-PPGL 数据集上取得最优分类性能，实验结果验证了 MFF-Net 在医学影像序列分析中的有效性，为后续相关研究提供新思路和方法。

4.3.3 消融实验

表 4-2 各个模块在 Xiangya-PPGL 数据集上的消融实验

Exp.	FAM	SFE	SMF	AUC	Accuracy	Specificity	Sensitivity
1	×	×	×	0.7978	0.7187	0.7476	0.7000
2	√	×	×	0.8215	0.7224	0.7371	0.7125
3	×	√	√	0.9249	0.8875	0.7875	0.9125
4	√	√	√	0.9590	0.9267	0.8250	0.9583

为了评估 MFF-Net 中各个组件的贡献，本节在基准模型中逐步应用本章提出的模块来验证它们的有效性。SMF 的核心功能是融合 SFE 生成的多帧特征，因此在消融实验中，这两个模块应同时保留或移除。为了进一步验证 SFE 的有效性，将其与其他常见的多帧聚合模块进行实验。此外，为了探究参考帧数量对模型性能的影响，本研究还设置了参考帧数量的消融实验。消融实验在 Xiangya-PPGL 数据集上进行，采用五折交叉验证的方式测试性能。表 4-2 显示了增加或减少每个模块对模型性能的影响，其中 √ 表示使用模块，× 表示不使用

模块，各指标最优结果均已加粗。

(1) FAM

表 4-2 显示，基准模型增加特征聚焦模块后，AUC 从 0.7978 提升至 0.8215、Sensitivity 从 0.7 提升至 0.7125、Accuracy 从 0.7187 提升至 0.7224，但 Specificity 从 0.7476 下降至 0.7371，这一变化趋势与第三章 LDO 模块消融实验结果一致，二者均通过增强小尺度肿瘤特征表达能力提升模型性能，但同时引入少量噪声导致 Specificity 小幅下降。第四章研究选择 FAM 而非 LDO 作为模块主要基于以下考虑：(1) FAM 无需引入额外参数，更适合多帧场景下的计算效率要求；(2) FAM 通过直接放大肿瘤区域特征，能更有效缓解 CT 序列中肿瘤两端切片的帧退化问题，这些优势使 FAM 在多帧肿瘤任务中表现出更强的适用性和鲁棒性。

(2) SFE 和 SMF

表 4-2 实验结果显示，引入多帧数据进行肿瘤基因亚型分类可显著提升所有评估指标，敏感度从 0.7 提升至 0.9125、准确率从 0.7187 增至 0.8875，这证明通过患者多帧数据能有效捕捉肿瘤动态特征和空间异质性、为基因亚型分类提供更具判别性的特征。

(3) SMF

SMF 用于融合 SFE 输出的多帧特征，为验证该模块有效性，本文研究多种常见多帧特征融合模块对模型的影响，对比全局平均池化(Global Average Pooling, GAP)、全局最大池化(Global Max Pooling, GMP)、混合池化(Mixed Pooling)、自注意力池化(Self-Attention Pooling)并简要介绍如下：全局平均池化通过计算每一帧特征同一位置平均值融合多帧特征；全局最大池化通过计算每一帧特征同一位置最大值融合多帧特征；混合池化结合最大池化和平均池化、通过可学习权重参数决定两者比例；自注意力池化通过注意力机制动态计算每个位置重要性并加权求和。

表 4-3 多帧融合模块消融实验结果

Method	AUC	Accuracy	Specificity	Sensitivity
Global Average Pooling	0.8839	0.8531	0.7125	0.9000
Global Max Pooling	0.9260	0.8719	0.7500	0.9125
Mixed Pooling	0.8942	0.8594	0.8875	0.8500
Self-Attention Pooling	0.9589	0.9260	0.8625	0.9458
SMF	0.9590	0.9267	0.8250	0.9583

多帧融合模块消融实验结果见表 4-3 (各指标最优结果已加粗)，实验显示全局平均池化和全局最大池化在 Accuracy 和 Specificity 上表现欠佳，因其仅通过简单池化操作处理特征，难以捕捉序列数据长距离依赖关系；本文方法借助序列卷积分支的空洞卷积，可显式建模帧间长距离依赖。混合池化在 Specificity 表

现最佳但 Sensitivity 下降, 说明其平衡正负类样本存在局限; 自注意力池化在 Sensitivity 和 AUC 表现良好, 但 Accuracy 略低, 或因注意力机制过度关注全局特征致分类性能不稳定。本文方法通过动态感知分支高比例 Dropout 增强特征鲁棒性、缓解过拟合, 结合特征拼接与时域压缩策略实现多维度特征深度融合, 既保留细粒度时序信息又控制计算复杂度, 最终在 Accuracy 和 AUC 取得最佳性能, AUC 达 0.959 显著优于其他方法。

(4) 帧数

本研究测试不同帧数对模型性能的影响, 所有抽帧方式与 4.3.1 实验配置描述方法一致, 实验结果如表 4-4 所示 (各指标最优结果已加粗)。帧数为 8 时模型 AUC 最高, Accuracy、Specificity、Sensitivity 均为最优, 表明 8 帧输入可为模型提供足够时序信息并避免冗余或噪声干扰; 帧数较少 (如 2 或 4) 时模型性能显著下降, 准确率和特异性较低, 显示时序信息不足限制模型对肿瘤基因亚型关键特征的捕捉能力; 帧数过多 (如 16 或 32) 时性能略有下降, 推测因帧数增多引入更多噪声。综合计算资源与模型性能, 最终选择 8 帧作为每个患者的使用帧数。

表 4-4 帧数对模型性能影响

Frame Number	AUC	Accuracy	Specificity	Sensitivity
2	0.8635	0.7500	0.7000	0.7667
4	0.8844	0.8094	0.7750	0.8208
8	0.9590	0.9267	0.8250	0.9583
16	0.9454	0.8944	0.7375	0.9393
32	0.9258	0.8639	0.8125	0.8786

4.4 本章小结

本章节主要研究目的是利用患者 CT 序列数据识别 PPGL 基因亚型, 针对多帧输入引发的特征退化及肿瘤空间连续性信息有效提取难的问题, 提出 MFF-Net 模型: 首先通过 FFM 模块聚焦肿瘤区域以解决序列帧特征退化问题, 其次通过 SFE 模块提取肿瘤空间连续性信息以提高特征捕捉精度并增强模型在不同场景下的适应性和泛化能力, 此外设计 SMF 模块融合多帧特征以提升模型对复杂动态场景的建模能力。对比实验和消融实验证明, 该模型可有效解决 PPGL 基因亚型识别任务, 有望为临床提供可靠辅助诊疗。

第 5 章 总结与展望

5.1 论文总结

嗜铬细胞瘤及其副神经节瘤（PPGL）是发病率仅为 0.46/10 万的罕见神经内分泌肿瘤，其基因亚型与肿瘤恶性风险评估、治疗方案选择及患者预后密切相关。传统基因检测方法（如 RNA 测序或穿刺活检）因创伤性高、成本昂贵、普及性差等问题制约了常规临床应用。随着医学影像技术与人工智能发展，深度学习在肿瘤基因表型预测中应用渐广，本研究构建深度学习模型实现影像无创预测 PPGL 基因亚型，旨在为临床提供低成本、高效且可推广的辅助工具以优化个体化诊疗策略并改善患者预后，主要研究工作及临床应用价值如下：

（1）构建湘雅嗜铬细胞瘤影像 - 基因数据集（Xiangya-PPGL）。PPGL 领域当前缺少公开可用的标准数据集，本研究以湘雅医院为主体联合长沙及周边医院，回溯 2010 年至 2022 年临床病例，整理得到 425 例含基因测序结果、CT 序列影像及病灶勾画的完整标签患者数据。研究对影像数据实施格式转换、CT 值归一化、图像裁剪与填充等预处理，发现 PPGL 肿瘤普遍存在尺寸差异大、区域占比小的特征。该数据集既为本研究提供数据支撑，也为同行开发 AI 辅助诊断模型提供标准化资源，未来可用于训练影像基因预测工具、支持医院构建智能化辅助诊疗平台。

（2）提出基于肿瘤区域面积最大帧的单帧分类模型以优化表征能力提升分类准确率。针对 PPGL 肿瘤在 CT 影像中区域占比小、尺寸差异大的问题，研究从三方面优化模型结构：设计局部细节优化模块通过多层次特征提取与细粒度调整提升对局部病灶的感知能力；引入特征融合编码器缓解低肿瘤占比图像的信息丢失问题；结合多尺度特征设计末端注意力机制增强对关键特征的聚焦能力，构建高效鲁棒的肿瘤分类框架。该模型可应用于临床术前辅助评估，为影像医生提供初步基因亚型提示，在无创条件下支持治疗分型决策，尤其适用于基层医院或资源有限环境中 PPGL 的初筛与转诊判断。

（3）针对单帧模型空间信息不足局限提出多帧融合分类模型以全面提取肿瘤整体形态与空间特征。传统依赖面积最大帧的分析方法难以全面反映肿瘤形态特征，本研究设计多帧融合框架：构建特征聚焦模块提升 ROI 细节保留能力，引入空间特征提取模块通过可学习向量与自注意力机制捕捉帧间上下文，提出序列感知模块以双路结构实现多帧信息深度融合，显著提升模型对基因亚型的识别能力。该多帧模型适用于临床术前精准分型与高危风险判断，为手术策略、靶向治疗或长期随访提供依据，也可嵌入医院 PACS 系统作为辅助诊断模块以提升

复杂病例诊断效率与准确性。

本研究并行设计单帧与多帧深度学习模型并验证有效性,结果显示多帧融合模型较单帧模型特征提取能力更强、预测准确率更高,在临床影像数据中识别 PPGL 基因亚型具巨大潜力。未来该方法有望作为无创、快速、低成本的基因检测替代方案,在无法穿刺或基因检测资源不足场景中为临床医生提供可靠辅助,推动罕见肿瘤精准医疗落地应用。

5.2 工作展望

5.2.1 研究局限性

(1) 数据泛化性不足:本研究模型训练与验证完全基于自建的 Xiangya-PPGL 数据集,未在更广泛临床环境中验证,由于不同地区和医院的 CT 设备参数、扫描协议及患者群体特征存在差异,仅依赖单一的数据集训练的模型在实际临床应用中可能因数据分布改变而导致性能下降,难以确保在多样化临床场景中稳定运行。

(2) 低样本数据适应性差:嗜铬细胞瘤等罕见病本身发病率低,尽管本研究尽力收集大量影像数据,但数据规模仍受极大限制,实际临床中新出现的罕见基因亚型病例可能因缺乏足够数据支持而使模型无法准确预测。

(3) 人工标注依赖性强:模型输入准确性依赖医生手动标注肿瘤区域,该过程耗时耗力、严重影响工作效率,且标注结果受医生个人经验和主观判断影响较大,不同医生标注差异会导致训练数据不一致,进而影响模型稳定性和可靠性,难以实现标准化临床应用。

(4) 单模态信息不全面:研究仅基于 CT 影像进行 PPGL 基因亚型分类,CT 影像虽能清晰显示肿瘤形态、大小等解剖结构信息,但无法全面反映病灶代谢、血流灌注等重要生理信息,单一模态的局限性使模型难以获取完整病灶特征,限制诊断准确性进一步提升,无法满足临床对复杂病情精准诊断的需求。

5.2.2 未来研究方向

(1) 拓展数据验证范围:依托湘雅医院持续收集胰腺神经内分泌肿瘤、骨恶性纤维组织细胞瘤、腺泡状软组织肉瘤等多种罕见病患者 CT 数据,构建多中心大规模数据集,将本研究算法应用于不同类型罕见病基因亚型分类任务,通过多样化临床数据测试全面评估模型泛化能力、验证稳健性,为算法临床广泛应用奠定基础。

(2) 优化低样本数据性能:通过系统性逐步减少训练集样本量,深入测试算法在不同数据规模下的性能表现,分析模型预测准确性和稳定性随样本量变化

的规律，探索满足临床识别精度要求且最大限度减少数据依赖的数据量阈值，结合迁移学习、数据增强等技术提高模型在低样本条件下的学习能力和预测准确性，增强对罕见病例的适应性。

(3) 实现自动化分割技术：探索 U-Net、DeepLabV3+ 等先进深度学习分割网络，利用其特征提取和图像分割能力实现 PPGL 肿瘤自动分割，尝试采用弱监督或多实例学习方法在部分标注数据情况下训练分割模型以降低对大量手动标注数据的依赖，最终构建涵盖影像预处理、肿瘤自动分割到基因分型预测的端到端自动化诊断系统，提高诊断效率和准确性、减少人为因素干扰。

(4) 推进多模态融合研究：开展 CT 与 MRI、PET 等多模态影像融合技术研究，挖掘不同模态影像互补信息，如结合 PET-CT 的肿瘤代谢活性信息和 MRI 的高软组织对比度与精细解剖细节，通过设计高效多模态特征提取网络实现不同模态特征有机融合以提升模型对病灶特征的全面感知能力，同时将临床数据（如患者病史、症状、实验室检查结果等）与多模态影像数据结合，优化个性化诊断和预测策略，为患者提供更精准全面的诊疗方案。

参考文献

- [1] 唐小林, 刘云飞. 机器人辅助小儿二次腹腔镜嗜铬细胞瘤切除术麻醉管理一例[J]. 中国临床新医学, 2024, 17(08): 930-932.
- [2] Nölting S, Bechmann N, Taieb D, et al. Personalized management of pheochromocytoma and paraganglioma[J]. *Endocrine Reviews*, 2022, 43(2): 199-239.
- [3] Antonio K, Valdez M M N, Mercado-Asis L, et al. Pheochromocytoma/paraganglioma: recent updates in genetics, biochemistry, immunohistochemistry, metabolomics, imaging and therapeutic options[J]. *Gland Surgery*, 2020, 9(1): 105.
- [4] Crona J, Taïeb D, Pacak K. New perspectives on pheochromocytoma and paraganglioma: toward a molecular classification[J]. *Endocrine Reviews*, 2017, 38(6): 489-515.
- [5] Bechmann N, Moskopp M L, Ullrich M, et al. HIF2 α supports pro-metastatic behavior in pheochromocytomas/paragangliomas[J]. *Endocrine-related Cancer*, 2020, 27(11): 625-640.
- [6] Rose N R, McDonough M A, King O N, et al. Inhibition of 2-oxoglutarate dependent oxygenases[J]. *Chemical Society Reviews*, 2011, 40(8): 4364-4397.
- [7] Berends A M, Buitenwerf E, de Krijger R R, et al. Incidence of pheochromocytoma and sympathetic paraganglioma in the Netherlands: a nationwide study and systematic review[J]. *European Journal of Internal Medicine*, 2018, 51: 68-73.
- [8] Graeff J D, Horak B. The incidence of phaeochromocytoma in the Netherlands[J]. *Acta Medica Scandinavica*, 1964, 176(5): 583-593.
- [9] Lenders J W, Duh Q Y, Eisenhofer G, et al. Pheochromocytoma and paraganglioma: an endocrine society clinical practice guideline[J]. *The Journal of Clinical Endocrinology & Metabolism*, 2014, 99(6): 1915-1942.
- [10] Liu Z, Zhu L, Roberts R, et al. Toward clinical implementation of next-generation sequencing-based genetic testing in rare diseases: Where are we?[J]. *Trends in Genetics*, 2019, 35(11): 852-867.
- [11] Aim L B, Maher E R, Cascon A, et al. International initiative for a curated SDHB variant database improving the diagnosis of hereditary paraganglioma

- and pheochromocytoma[J]. *Journal of Medical Genetics*, 2022, 59(8): 785-792.
- [12] Pang Y, Li M, Jiang J, et al. Impact of body composition and genotype on haemodynamics during surgery for pheochromocytoma and paraganglioma[J]. *Journal of Cachexia, Sarcopenia and Muscle*, 2022, 13(6): 2843-2853.
- [13] Liu H, Guan X, Xu B, et al. Computed tomography-based machine learning differentiates adrenal pheochromocytoma from lipid-poor adenoma[J]. *Frontiers in Endocrinology*, 2022, 13: 833413.
- [14] Yi X, Guan X, Zhang Y, et al. Radiomics improves efficiency for differentiating subclinical pheochromocytoma from lipid-poor adenoma: a predictive, preventive and personalized medical approach in adrenal incidentalomas[J]. *The EPMA Journal*, 2018, 9(4): 421-429.
- [15] Yi X, Guan X, Chen C, et al. Adrenal incidentaloma: machine learning-based quantitative texture analysis of unenhanced CT can effectively differentiate sPHEO from lipid-poor adrenal adenoma[J]. *Journal of Cancer*, 2018, 9(19): 3577-3582.
- [16] Lu C F, Liao C Y, Chao H S, et al. A radiomics-based deep learning approach to predict progression free-survival after tyrosine kinase inhibitor therapy in non-small cell lung cancer[J]. *Cancer imaging : the Official Publication of the International Cancer Imaging Society*, 2023, 23(1): 9.
- [17] Yin W, Wang W, Zou C, et al. Predicting tumor mutation burden and EGFR mutation using clinical and radiomic features in patients with malignant pulmonary nodules[J]. *Journal of Personalized Medicine*, 2022, 13(1).
- [18] Zhu H, Song Y, Huang Z, et al. Accurate prediction of epidermal growth factor receptor mutation status in early-stage lung adenocarcinoma, using radiomics and clinical features[J]. *Asia-Pacific Journal of Clinical Oncology*, 2022, 18(6): 586-594.
- [19] Lambin P, Leijenaar R T H, Deist T M, et al. Radiomics: the bridge between medical imaging and personalized medicine[J]. *Nature reviews. Clinical Oncology*, 2017, 14(12): 749-762.
- [20] Gardin I, Grégoire V, Gibon D, et al. Radiomics: Principles and radiotherapy applications[J]. *Critical Reviews in Oncology/Hematology*, 2019, 138: 44-50.
- [21] Shiri I, Amini M, Nazari M, et al. Impact of feature harmonization on radiogenomics analysis: Prediction of EGFR and KRAS mutations from non-small cell lung cancer PET/CT images[J]. *Computers in Biology and*

- Medicine, 2022, 142: 105230.
- [22] Zhu X, Dong D, Chen Z, et al. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer[J]. European Radiology, 2018, 28(7): 2772-2778.
- [23] Zhou Y, Zhan Y, Zhao J, et al. CT-Based radiomics analysis of different machine learning models for discriminating the risk stratification of pheochromocytoma and paraganglioma: A multicenter study[J]. Academic Radiology, 2024, 31(7): 2859-2871.
- [24] Yang F, Chen W, Wei H, et al. Machine learning for histologic subtype classification of non-small cell lung cancer: A retrospective multicenter radiomics study[J]. Frontiers in Oncology, 2020, 10: 608598.
- [25] Chen P T, Chang D, Liu K L, et al. Radiomics utilization to differentiate nonfunctional adenoma in essential hypertension and functional adenoma in primary aldosteronism[J]. Scientific Reports, 2022, 12(1): 8892.
- [26] Maggio R, Messina F, D'Arrigo B, et al. Machine learning-based texture analysis in the characterization of cortisol secreting vs. non-secreting adrenocortical incidentalomas in CT scan[J]. Frontiers in Endocrinology, 2022, 13: 873189.
- [27] Krishan A, Mittal D. Multi-Class liver cancer diseases classification using CT images[J]. The Computer Journal, 2021, 66(3): 525-539.
- [28] Lin W, Hasenstab K, Moura Cunha G, et al. Comparison of handcrafted features and convolutional neural networks for liver MR image adequacy assessment[J]. Scientific Reports, 2020, 10(1): 20336.
- [29] Yang M, He X, Xu L, et al. CT-based transformer model for non-invasively predicting the Fuhrman nuclear grade of clear cell renal cell carcinoma[J]. Frontiers in Oncology, 2022, 12: 961779.
- [30] Kusunoki M, Nakayama T, Nishie A, et al. A deep learning-based approach for the diagnosis of adrenal adenoma: a new trial using CT[J]. The British Journal of Radiology, 2022, 95(1135): 11066.
- [31] Yasaka K, Akai H, Abe O, et al. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: A preliminary study[J]. Radiology, 2018, 286(3): 887-896.
- [32] Santra B, Jha A, Mukherjee P, et al. Anatomical location-guided deep learning-based genetic cluster identification of pheochromocytomas and

- paragangliomas from CT images[C]//Wu S, Shabestari B, Xing L. Applications of Medical Artificial Intelligence. Cham: Springer Nature Switzerland, 2024: 62-71.
- [33] Sori W J, Feng J, Godana A W, et al. DFD-Net: lung cancer detection from denoised CT scan image using deep learning[J]. *Frontiers of Computer Science*, 2021, 15: 1-13.
- [34] Li D, Yuan S, Yao G. Classification of lung nodules based on the DCA-Xception network[J]. *Journal of X-ray Science and Technology*, 2022, 30(5): 993-1008.
- [35] Lei H, Tong G, Su H, et al. Dual branch CNN and transformer for cardiac atherosclerotic plaque classification[C]//2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Istanbul, Turkiye: IEEE, 2023: 2007-2010.
- [36] Zhu J, Zou L, Xie X, et al. 2.5D deep learning based on multi-parameter MRI to differentiate primary lung cancer pathological subtypes in patients with brain metastases[J]. *European Journal of Radiology*, 2024, 180: 111712.
- [37] Li Y, Yu R, Chang H, et al. Identifying pathological subtypes of brain metastasis from lung cancer using MRI-Based deep learning approach: A multicenter study[J]. *Journal of Imaging Informatics in Medicine*, 2024, 37(3): 976-987.
- [38] Saihood A, Karshenas H, Naghsh-Nilchi A R. Multi-Orientation local texture features for guided attention-based fusion in lung nodule classification[J]. *IEEE Access*, 2023, 11: 17555-17568.
- [39] Tian J, Zhang Y, Lei J, et al. Lightweight classification network for pulmonary tuberculosis based on CT images[J]. *Journal of Artificial Intelligence and Technology*, 2023, 3(1): 25-31.
- [40] Suman S, Singh G, Sakla N, et al. Attention based CNN-LSTM network for pulmonary embolism prediction on chest computed tomography pulmonary angiograms[M]//De Bruijne M, Cattin P C, Cotin S, et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021: Vol. 12907. Cham: Springer International Publishing, 2021: 356-366.
- [41] Shen W, Zhou M, Yang F, et al. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification[J]. *Pattern Recognition*, 2017, 61: 663-673.

- [42] Shafiq-ul-Hassan M, Zhang G G, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels[J]. *Medical Physics*, 2017, 44(3): 1050-1062.
- [43] Li X, Morgan P S, Ashburner J, et al. The first step for neuroimaging data analysis: DICOM to NIfTI conversion[J]. *Journal of Neuroscience Methods*, 2016, 264: 47-56.
- [44] Wiesler S, Ney H. A convergence analysis of log-linear training[J]. *Advances in Neural Information Processing Systems*, 2011, 24.
- [45] 李旭辉. 基于放射基因组学的 Ki-67 指数非侵入式预测[D]. 中南大学, 2023.
- [46] 冯茜茜, 韩福刚, 肖燕, 等. 基于磁共振常规序列的纹理分析在鉴别四肢软组织肿瘤良恶性中的应用价值[J]. *中国医学计算机成像杂志*, 2021, 27(04): 313-317.
- [47] 高玉妮. 基于深度学习的影像组学模型在脑胶质瘤鉴别诊断及基因分型中的应用[D]. 重庆医科大学, 2023.
- [48] 林佳璐, 陈佳雪, 娄鉴娟, 等. 术前多区域多参数乳腺 MRI 影像组学对 N2-3 期腋窝淋巴结的预测价值[J]. *南京医科大学学报(自然科学版)*, 2025, 45(02): 185-195.
- [49] Zhang Q, Jiang Z, Lu Q, et al. Split to be slim: An overlooked redundancy in vanilla convolution[J]. *arXiv preprint arXiv:2006.12085*, 2020.
- [50] Finder S E, Amoyal R, Treister E, et al. Wavelet convolutions for large receptive fields[C]//*European Conference on Computer Vision*. Springer, 2024: 363-380.
- [51] Woo S, Debnath S, Hu R, et al. Convnext v2: Co-designing and scaling convnets with masked autoencoders[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: 16133-16142.
- [52] Chen J, Yu Q, Shen X, et al. Vitamin: Designing scalable vision models in the vision-language era[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024: 12954-12966.
- [53] Zhu L, Wang X, Ke Z, et al. Biformer: Vision transformer with bi-level routing attention[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: 10323-10333.
- [54] 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.
- [55] Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 116-131.
- [56] Koonce B. EfficientNet[M]//Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization. Springer, 2021: 109-123.
- [57] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 2261-2269.
- [58] Zou S, Zhang Z, Zou Y, et al. Microscopic-Mamba: Revealing the secrets of microscopic images with just 4M parameters[J]. arXiv preprint arXiv:2409.07896, 2024.
- [59] Huo X, Sun G, Tian S, et al. HiFuse: Hierarchical multi-scale feature fusion network for medical image classification[J]. Biomedical Signal Processing and Control, 2024, 87: 105534.
- [60] Zhu X, Cheng D, Zhang Z, et al. An empirical study of spatial attention mechanisms in deep networks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 6688-6697.
- [61] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.
- [62] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.
- [63] Rogers W, Thulasi Seetha S, Refaee T A G, et al. Radiomics: from qualitative to quantitative imaging[J]. The British Journal of Radiology, 2020, 93(1108): 20190948.
- [64] Ajith T A. Current insights and future perspectives of hypoxia-inducible factor-targeted therapy in cancer[J]. Journal of Basic and Clinical Physiology and Pharmacology, 2018, 30(1): 11-18.
- [65] Anan M, Abe T, Matsuda T, et al. Induced angiogenesis under cerebral ischemia by cyclooxygenase 2 and hypoxia-inducible factor naked DNA in a rat indirect-bypass model[J]. Neuroscience Letters, 2006, 409(2): 118-123.

- [66] Sun P, Bush S J, Wang S, et al. STMiner: Gene-centric spatial transcriptomics for deciphering tumor tissues[J]. *Cell Genomics*, 2025, 5(2): 100771.
- [67] Sun A, Zhang Z, Lei M, et al. Boosting breast ultrasound video classification by the guidance of keyframe feature centers[M]//Greenspan H, Madabhushi A, Mousavi P, et al. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*: 14224. Cham: Springer Nature Switzerland, 2023: 441-451.
- [68] Yang X, Chin B, Silosky M, et al. Learning with synthesized data for generalizable lesion detection in real PET images[M]//Greenspan H, Madabhushi A, Mousavi P, et al. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*: 14224. Cham: Springer Nature Switzerland, 2023: 116-126.
- [69] Yang G Y, Nakayama G K, Xiao Z K, et al. Semantic-aware transformation-invariant RoI align[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*: 38. 2024: 6486-6493.
- [70] Fan H, Xiong B, Mangalam K, et al. Multiscale vision transformers[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 6824-6835.
- [71] 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.
- [72] 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.
- [73] Li X, Zhu Y, Wang L. Zeroi2v: Zero-cost adaptation of pre-trained transformers from image to video[C]//*European Conference on Computer Vision*. Springer, 2024: 425-443.
- [74] Huang Z, Zhang S, Pan L, et al. Temporally-adaptive models for efficient video understanding[J]. *arXiv preprint arXiv:2308.05787*, 2023.

攻读学位期间主要的研究成果

一、发表的学术论文

[1] Deep learning for predicting molecular subtype of pheochromocytomas and paragangliomas: A multicenter retrospective cohort study (一作, 投稿至 *Physica Medica*, IF=3.3, JCR Q1, 对应本文中第三章部分内容)

[2] A Three-Dimensional Multiframe Integration Network for Genotypic Subtyping of Pheochromocytomas[C]//Proceedings of 2024 5th International Symposium on Artificial Intelligence for Medicine Science, ISAIMS 2024. Amsterdam, Netherlands, 2025: 143-147. (一作, 对应本文中第四章部分内容)

[3] Image Analysis and Diagnosis of Skin Diseases - A Review [J]. *Current Medical Imaging*, 2023, 19(3): 199-242. (三作, IF:1.315, JCR Q3)

二、主持和参与的科研项目

(1) 国家重点研发子课题: 老年失能分类评估标准体系及预警模型研究, 项目编号: 2020YFC2008501, 2021.9-2024.2, 参与.

(2) 湖南省自然科学基金: 基于深度学习影像-基因组学评估 PPGL 血流动力学不稳定和转移风险的多中心研究, 项目编号: 2022JJ30979, 2022.01-2024.12, 参与.

(3) 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, 项目编号: 718010185, 2023.8-2024.8, 参与.

(4) 企、事业单位委托项目: 基于图像的灵芝病虫害 AI 检测技术研究, 项目编号: 718010299, 2024.11-2025.11, 参与.

三、个人获奖情况

(1) 国家级, 中国高校大数据挑战赛, 国家级一等奖, 2022.

(2) 国家级, Mathorcup 高校数学建模挑战赛, 国家级二等奖, 2023.

(3) 校级, 中南大学 2022-2023 学年优秀学生, 2024.

(4) 校级, 中南大学 2022-2023 学年校级一等奖学金, 2023.

致 谢

“沧浪之水清兮，可以濯吾缨；沧浪之水浊兮，可以濯吾足。”听着故乡的童谣，回首三年研究生生涯，从沧浪洲畔到橘子洲头，见到了很多，经历了很多，也成长了很多。

学贵得师，亦贵得友。感谢中南大学提供的优秀师资。读研三年最应该感谢的就是谢老师，他是一位有格局，讲奉献，善作为，敢担当，时时刻刻以大局为重的优秀导师。他的一言一行都对我产生了深刻的影响，让我没有虚度我的研究生生涯。生病时的关心、择业迷茫时的指引、每次工作结束后的那一句“ok 辛苦了”都是我最珍贵的回忆。感谢刘乐书记，刘雯老师，张巧丽老师，胡艳老师，谭小钰老师等院办老师在学业和生活中给予的帮助和指导。

同道为朋，共研究生辉。感谢实验室同门的帮助。振南带着我去发现长沙的美食，让读研时光多一份烟火气；清锋在我低落的时候帮我理清事实，让读研之路没有迷失方向；子清在我最辛苦的时候帮我分担任务，让我少熬了很多个夜晚；思航勇于从我手中接过“管家”的重任，让我研三可以专心于工作和论文；硕清、德坤、周鑫、炼爽和我一起攻克 PPT 难题，在我自顾不暇时，及时出手；嘉明、邵康、锦伦、欣玉和我一起应对一个个读研路上的难题。

幸遇同檐，暖岁寒年。感谢室友的陪伴。“柳书记”经常在宿舍普及国家大事，引导着我们向党组织靠近，经常带我去参加运动，岳麓山顶的牡丹、百里毅行的江景、明月山巅的松柏等都是友谊的见证；“刘班长”主动承担起宿舍卫生工作，经过他打扫的厕所总是最干净的，过节的时候记给大家送小礼物，让宿舍显得更加温暖；“李总”的心灵辅导，让我的很多疑问都得到解答，更是教会了我要敢于斗争善于斗争，要去争取自己的合理合法合规的权益。

春晖寸草，山海难量。感谢父母这么多年的养育，您们给我提供了默默地支持让我能够走到这里，我见过你们所没见过的风景，你们错过的风景未来我也会陪你们补上。

2025 年 5 月 20 日

于湖南长沙