GS-MGF-LPLs:计算机视觉辅助的不确定性甲状腺结节超声图像分割方法

杨 杰[1,*]，刘贤贤[2]，田洪瑜[3]，何志容[4]，康 慷[1]，陈 宏[1]

1. 重庆工贸职业技术学院，人工智能学院，重庆, 408000
2. 澳门大学，科技学院，中国澳门特别行政区, 999078
3. 重庆工贸职业技术学院，车辆工程学院，重庆, 408000
4. 重庆大学附属涪陵医院，超声医学科，重庆, 408099

摘要： 不确定性甲状腺结节（Indeterminate thyroid nodules，ITNs）的检测对患者生存预后至关重要。超声检查作为其诊断评估的金标准，在实际应用中面临医学图像细粒度分割资源消耗大、标注数据稀缺的挑战。本文创新性地提出网格搜索聚类驱动的多门控局部块学习模型（简称 GS-MGF-LPLs）。该模型通过重构双阈值快速搜索策略的块深度全卷积网络架构，将输入空间划分为潜在簇和噪声簇两类特征空间，并基于输出空间的局部簇构建目标模型。其技术突破体现在三个层面：（1）采用门控融合模型实现扫描参数的输入层前自动分割，以恒定连接权重表征不同类别簇的预测权重；（2）在加权层后集成多网格搜索局部块学习系统，通过多系统投票机制生成决策；（3）输出层通过解码分割结果获取块预测输出，运用动态加权平均算法保持特征多样性。基于专门构建的网格搜索耦合分割数据集验证表明，本模型在四类 ITNs 分割任务中取得突破性性能：准确率达 99.54%，精确率为 86.06%，F 值与 Dice 系数均达到 84.26%，显著优于当前五种主流分割方法（p<0.01）。该成果不仅为 ITNs 的早期精准诊断提供了可靠的技术支撑，其创新的特征空间解耦思想和多系统集成架构更为医学图像分析领域提供了方法论参考。目前，该模型已进入多中心临床验证阶段，有望成为辅助医生诊断决策的有力工具。

关键词： 计算机视觉辅助；超声图像；网格搜索优化；多门控融合学习；局部 Patch 学习；局部可靠性度量

Computer Vision-Aided *Indeterminate Thyroid Nodule* Ultrasonography Segmentation with *GS-MGF-LPLs* Model

Jie Yang[1,*]，Xian-Xian Liu[2]，Hongyu Tian[3]，Zhirong He[4]，Kang Kang[1]，Hong Chen[1]

1. College of Artificial Intelligence, Chongqing Industry and Trade Polytechnic, Chongqing, China, 408000
2. Department of Computer and Information Science, University of Macau, Macau SAR, China, 999078
3. College of Vehicle Engineering, Chongqing Industry and Trade Polytechnic, Chongqing, China, 408000
4. Department of Ultrasound, Chongqing University Fuling Hospital, Chongqing, China, 408099

Abstract: Indeterminate thyroid nodule detection has huge impact on survival, and ultrasonography is the golden standard for detection and assessment of indeterminate thyroid nodules (ITNs). Pixel-wise image segmentation is resource exhaustive in medical image analysis. In practice, it is difficult to find annotated medical images with corresponding segmentation masks.

In this study, we propose a novel grid-search clustering-driven multi-gating local patch learning model (GS-MGF-LPLs). The design of the model is based on redesigning the structure of double-threshold fast search strategy-based patch deep fully convolutional network. Grid Search (GS) model typically focuses on dividing input space into several local spaces (implied by stepwise value clusters). GS-MGF-LPLs uses two distinct local spaces called potential clusters and noise clusters, respectively. Through partitioning output space into several local spaces, each cluster space is used as desired (target) local output to construct local models. Key features of GS-MGF-LPLs can be enumerated as follows: (*i*) automated segmentation of scanning parameters before input layer are built through each gating fusion model (GF). Connections (weights) are specified as constant terms being in fact the prediction weight of different class of clusters; (*ii*) following weighting layer, it uses multi-GS-LPL systems in combination, produces final results via multi-system voting; (*iii*) at output layer, outputs of segmentation layer are decoded to produce corresponding patch prediction output. The weighted average is used, and weights are adjusted to maintain intra-class variation and rich features. The model was trained and tested on Grid-Search coupled with segmentation (GS-SEG) provided for ITNs study, with an accuracy of 99.5396%, a F-measure of 84.2636%, a precision of 86.0607%, and a dice coefficient of 84.2636%, which demonstrate the effectiveness of our proposed method. Through evaluation on our benchmark, experiments on 4 types of ITNs demonstrate that GS-MGF-LPLs outperforms five well-known methods. This achievement not only provides robust technical support for early and precise diagnosis of ITNs, but its innovative feature-space decoupling approach and multi-system integrated architecture further offer methodological references for the field of medical image analysis. Currently, the model has entered the multicenter clinical validation phase and is expected to become a powerful tool for assisting physicians in diagnostic decision-making.

Keywords: Indeterminate thyroid nodule; Computer aided detection and diagnosis; Grid search; Gating information; Local Patch learning; Local measure reliability

甲状腺结节早期检测对提高患者生存率至关重要。然而，不确定性甲状腺结节的诊断尤为困难，因其病灶表现微妙，这要求必须为内分泌科医生开发最有效的计算机辅助人工智能系统。为攻克早期甲状腺结节检测难题，本文尝试将多模态追踪技术与深度学习相结合。智能癌症影像诊断作为现代病理评估体系中的关键技术，其通过计算机视觉实现的精准检测能显著提升超声诊断的可靠性[1]。在多模态分割研究领域，通过挖掘不同模态间的相关性可有效改善分割结果。

卷积神经网络（CNN）[2]在医学图像分类与分割任务中已展现出超越传统计算机视觉技术的性能，但其在处理甲状腺癌细粒度亚型分类时仍存在明显局限。核心挑战集中于多尺度特征提取环节：尽管扩张卷积[3]与注意力机制[4]被提出用于增强感受野，这些方法在超声图像中往往难以有效区分噪声与真实病灶（ATC 未分化甲状腺癌病例中的 FPR 假阳性率高达 32%）。

多源数据融合（Multi-source data）进一步加剧了这一难题——不同成像设备间的域偏移（Domain Shift）会导致模型泛化性能显著下降（跨厂商设备的 Dice 系数降低 0.15-0.22）。多源数据协调问题日益受到关注，特别是针对不同超声设备和采集协议产生的数据差异。尽管对抗学习（adversarial learning）[5]。和风格迁移（Style Transfer）[6]等域适应技术（Domain adaptation techniques）已被用于降低扫描设备间的差异性，但这些方法在数据对齐过程中往往难以保留细粒度的诊断特征。目前基于图像块（patch）的策略[7]仍占主导地位，多数方法采用“分块-分类”范式，将大尺寸图像分割为小块进行独立分类[8]。然而该技术存在固有缺陷：小尺寸图像块提供的上下文信息有限，难以捕捉

对分类至关重要的广域纹理特征，本质上存在上下文缺失问题。为在尺寸限制下扩大视野范围，常规做法是对输入图像进行降采样，但这会牺牲空间分辨率——这种视野与分辨率的权衡已成为关键瓶颈，现有筛查流程中 35-42%的亚厘米结节被漏诊[9]正是其直接后果。这种矛盾关系使得从输入图像块中分割癌症亚型的任务变得更加复杂[10]。

近期，两个领域的研究进展均显示出潜在的架构改进方向。在文档分析领域，基于 Transformer 的架构（如 DiT）通过自监督预训练目标，显著提升了对噪声输入的处理能力[11]。多模态方法（如 LayOutLMv3）展示了如何通过整合多源信息流来增强特征判别能力[12]。Donut 采用编码器-解码器架构通过避免容易出错的中间处理步骤，提供了一种替代范式[13]。在增强任务方面，扩散模型为文档修复建立了新的性能基准，为医学图像预处理提供启发[14]。而 nnUNet 等专用架构则证明了领域特异性优化如何有效解决结构化数据提取中的特定挑战[15][15]。尽管这些方法提供了宝贵见解，但它们主要侧重于全局上下文建模或孤立地处理特定子任务。目前尚未出现能够统一处理噪声输入条件、细粒度特征判别和自适应特征加权的框架，以满足甲状腺结节分割等挑战性医学成像场景的需求。

本文介绍了一种基于网格搜索耦合分割技术（TNs-SEG）处理的医学影像数据集：该数据集源自重庆大学附属涪陵医院 ITNs 检测项目的早期超声医学影像，经过系统整理并配准分割掩模，所有标注均由资深内分泌科医师人工完成并核验。本研究分别采用传统分割方法和基于深度学习的 CNN 现代方法对该数据集进行了应用验证。由于早期癌症影像的实证资料十分稀缺，本数据集将为研究人员的结果复现和方法比较提供重要价值。通过在当前仅含逐帧标注的数据集中新增分割掩模，本研究为多媒体与计算机视觉研究者参与 ITNs 病灶分割及超声视频动态自动分析研究创造了条件。

本研究的主要贡献可归纳如下：

（1）提出基于多网格搜索的门控融合局部块语义分割方法，通过自适应加权多个专家 CNN 网络（采用不同放大倍率图像训练[16]，实现多尺度图像上下文信息融合。该方法有效整合局部视野和高分辨率图像的互补信息，为亚型识别提供关键特征支持。

（2）设计端到端重训练机制：通过联合优化专家 CNN 网络使其产生互补性预测，或训练自适应聚合 CNN 网络整合各专家输出。实验证明该策略显著提升分割性能[17]，并通过对比端到端训练前后专家网络的预测差异验证了学习有效性。

（3）在 ITNs 亚型区域分割这一挑战性任务中验证方法优势。相比现有最优方法，本方法在多类别分割任务中表现突出，且具备通用性优势——可适配任何类型的深度学习架构。

（4）针对医学图像小病灶特点，创新性地引入高分辨率低层特征图指导注意力模型学习。对于仅含图像级疾病分级的样本，本模型生成的像素级注意力图可作为伪标签，实现病灶分割的半监督学习。

（5）本文提出面向平面多连通域的多块参数化统计聚合架构。该方法不仅获得稳定迭代和更快收敛[18]，其设计理念还可推广至同类任务。

（6）构建多路径异构融合网络（multi-path heterogeneous fusion network），通过多样化融合路径降低模糊性、简化优化过程，相比传统单一路径融合网络[19]具有更优的适应性和鲁棒性。

1 渐进式阈值优化的甲状腺结节分割新方法

1.1 样本预处理与数据增强

本研究经重庆大学附属涪陵医院伦理委员会批准（批件号：2025CDFSFLYYEC-42），豁免患者知情同意要求，所有实验流程均严格遵循《涉及人的生物医学研究伦理审查办法》及相关技术规范。研究采用重庆大学附属涪陵医院 2021 年捐赠的甲状腺结节分割数据集进行方法验证，该数据集包含髓样癌（MTC）、滤泡状癌（FTC）、乳头状癌（PTC）和未分化癌（ATC）四种病理亚型的标注数据。所有金标准标注均由资深内分泌外科医师完成，鉴于数据规模庞大（约万帧级），专家标注采用病灶椭圆近似法——即在保证覆盖最大病灶区域的前提下，以椭圆形轮廓替代精确分割掩模。快速搜索算法可处理多尺寸图像块（涵盖 MTC、PTC、FTC、ATC 等异质性甲状腺结节），无需先验信息约束。实验证实，该简洁设计能显著提升准确率。双分支架构：主搜索分支（左）：包含用于疾病分级的双阈值定位模型 $S(\bullet)$；特征搜索分支（右）：用于学习病灶特征图的注意力模型 $Matt(\bullet)$。迭代优化流程：特征搜索→主搜索→回放缓冲区→参数更新的序列化帧处理。中间状态有效抑制背景噪声（检测准确率提升显著），最终分割结果输入下一模型层。技术说明：双阈值初始值设为 0.5（默认参数）：$XLo \in [0,\ 0.5)$，$Xhi \in (0.5,\ 1]$。

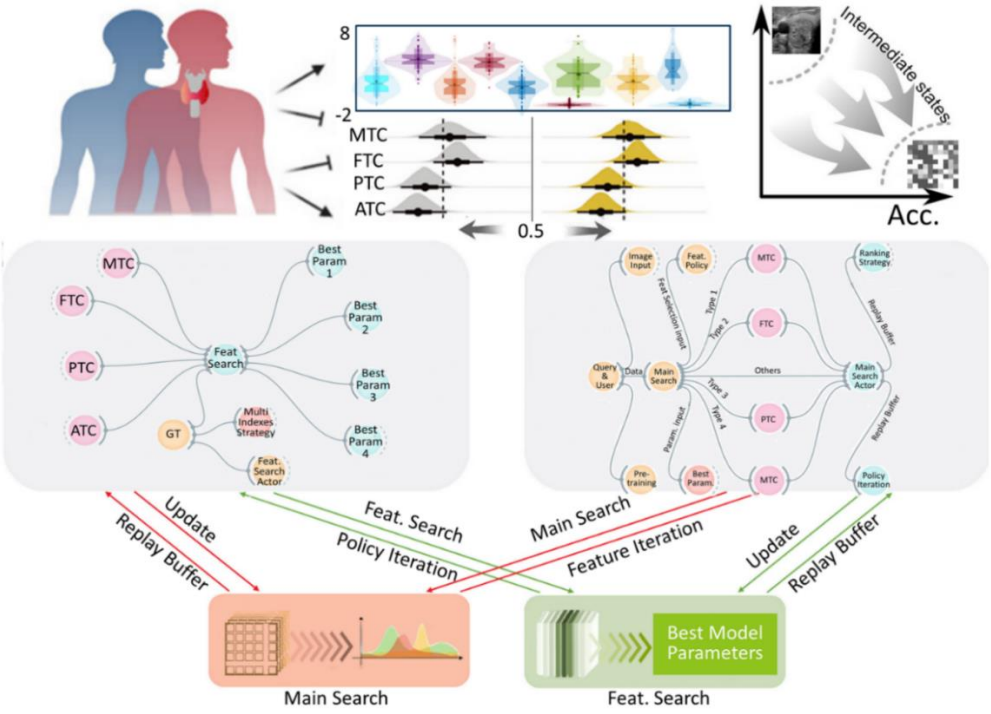

图 1 分支式阶段（BS）GPM 模型集成噪声抑制迭代池化策略至网格搜索。

Fig. 1 Branched stage (BS) GPM models integrated into the GS with noise reduction iterative pooling strategy.

此外，本文采用标准化流程进行了数据预处理。针对有限数据集问题，数据增强技术已被证实能有效提升机器学习（尤其是深度学习）模型的训练效果—其既可加速模型收敛，又能发挥正则化作用，从而避免过拟合并增强泛化能力。然而在训练标注数据有限情况下，模型性能提升仍面临挑战。为此，本文实施了以下数据增强步骤，这些措施同时有助于缓解不同类别间的样本不平衡问题。

1.2　基于 GPMs 的数据库知识发现与消融实验研究

如图 1 所示，全局病灶注意力网格搜索模型由两大核心组件构成：（1）主分支负责 ITNs 图像的灰度共生矩阵（GLCM）特征提取与双路径追踪；（2）绿色分支用于学习第 i 个病灶的注意力模型。

该模型创新性地利用图像分级标签提供的判别性定位信息来优化病灶注意力图，同时通过网格搜索（GS）算法自动生成具有代表性的图像块集合以替代原始大规模图像数据。

本方法中，基于学习的疾病采样模型 S（•）和病灶注意力模型 *Matt*（•）紧密集成。本研究首先使用具有不同灰度共生矩阵（GLCMs）的多注意力模型，仅通过输入图像学习多个注意力图以提取有效特征。预训练后，可以提取低层特征表示$\mathcal{F}_1$，$\mathcal{F}_2$，…，$\mathcal{F}_{22}$，（对应 GLCM 的 22 种特征输出），这些特征同时为多注意力模型提供像素级和类级指导[20]。此外，本研究将初始预测的病灶掩模$\mathcal{P}_{i=1}^{L}$作为额外输入引入注意力模型，这里统一用 L（Lesions）来代表病灶识别变量。整体表达式定义如下：

$$\alpha_{i=1}^{L} = Matt\left(\mathcal{F}_1,\ \mathcal{F}_2,\ \cdots \mathcal{F}_{22},\ \mathcal{P}_{i=1}^{L}\right) \tag{1}$$

其中输出$\alpha_{i=1}^{L}$为注意力图，其对表征疾病的各类病灶区域呈高响应值。该注意力机制包含两个关键步骤：第一步通过特征拼接模块F（·），融合输入图像和初始预测病灶掩模的低维编码特征[21]，实现像素级病灶特征提取。对于所有病灶区域，可通过下式获得注意力特征的中间状态：

$$E_{\mathcal{I},j} = F\begin{pmatrix} \mathcal{E}_{id}(ID),\ \mathcal{E}_{x,y}(\mathcal{F}_1),\ \mathcal{E}_{x,y}(\mathcal{F}_2), \\ \mathcal{E}_{x,y}(\mathcal{F}_3),\ \cdots,\ \mathcal{E}_{x,y}(\mathcal{F}_{22}) \end{pmatrix} \tag{2}$$

设$E_{\mathcal{I},j}$表示图像$\mathcal{I}$的第 *j*th 次交互的集成嵌入特征，对$\mathcal{E}$嵌入层中像素索引编码id的第 x，y 个像素进行特征嵌入。该集成嵌入序列随后被输入至模型[22][22]：

$$\beta_{\mathcal{I}}^{j} = S\left(concat\left(\mathcal{G}_L,\ E^{j}\left(\mathcal{T}_k^{low},\ \mathcal{T}_k^{high}\right)\right),\ \mathcal{S}_k^{\mathcal{L}}\right) \tag{3}$$

其中，全局图像矩阵$\beta_{\mathcal{I}}^{j}$用于对图像$\mathcal{I}$的主干特征进行加权，突出病灶区域，抑制无关背景。*Concat*（•）表示通道维度的特征拼接操作[23]，在通道维度上将病灶 L 的全局特征$\mathcal{G}_L$与经过筛选的嵌入特征E^{j}拼接起来。需要说明的是，对于输入特征的每个元素，激活区间下界$\mathcal{T}_k^{low}$，上界$\mathcal{T}_k^{high}$以及偏置项$\mathcal{S}_k^{\mathcal{L}}$均为第 *k*-th 次 echo 的可学习参数。本研究通过网格搜索优化这些参数以获取最优参数组合 grid.best_params。

在第二步中，本研究利用全局病灶关联矩阵与底层注意力特征进行交互，进而生成键值病灶注意力图，其计算表达式为：

$$\begin{bmatrix} \mathcal{k}_j \\ \mathcal{v}_j \end{bmatrix} = \mathcal{h}_j \tag{4}$$

$$\mathcal{M}_{\mathcal{I}} = Sigmoid\left(\mathcal{h}_j[\beta_{\mathcal{I}}^{j} \odot \mathcal{I}_{\ell}]\right) \tag{5}$$

隐状态$\mathcal{h}_j$被拆分为两部分：一部分为$\mathcal{k}_j$，另一部分为$\mathcal{v}_j$。系统仅针对键部分计算注意力权重，而值部分的加权结果则用于加权求和运算[24]。其中符号⊙表示逐元素相乘运算。*Sigmoid*（·）函数将加权结果映射为概率值，从而得到分割掩模表征$\mathcal{M}_{\mathcal{I}}$。

全局图像矩阵$\beta_{\mathcal{I}}^{j}$与$\mathcal{I}_{\ell}$具有相同的通道维度，其通过基础预训练 GS 模型顶层特征的 1×1 卷积计算获得。该底层引导特征蕴含丰富的类别信息，可有效权衡底层特征并优化病灶细节的精确表征。

基于 L 个病灶注意力图，本研究分别与底层图像特征$\mathcal{F}^{low}$进行逐元素相乘，并利用这些注意力特征微调预训练的疾病分类模型。所有注意力特征共享 GPM 模型的权重参数，输出的预分割图将被拼接以学习最终表征[25]。实验表明：仅使用下采样全局图像时，分支模型的平均交并比（mIoU）为

5%；仅使用图像块时，mIoU 可达 10%；通过双分支策略聚合及归一化处理，最终性能可提升至 10%。

2　协同学习研究

2.1　实验材料配置与数据集描述

本研究采用 ITNs-SEG 数据集进行实验验证，该异构数据集由 118 张超高分辨率甲状腺超声图像组成，经处理生成 3600 个 50×50 像素的样本块。如图 2 所示，本研究将多源异质性模型数据划分为恶性数据集和良性数据集（作为阴性对照的平衡样本），恶性输入（Malignant input）：细分为多种精细癌症类型（包括髓样癌（MTC）、滤泡癌（FTC）、乳头状癌（PTC）和未分化癌（ATC））以及“其他”类（“Other”如甲状腺炎症/非癌性病变）。这些输入被引入广义预处理模块（GPMs）以去除背景噪声。良性输入（Benign input）：作为额外的阴性输入。本文将其分割为等尺寸块数据集（Patch datasets）（5 像素×5 像素），并进一步划分为：70%训练集（Training set）；20%测试集（Testing set）；10%验证集（Validation set）（注：图上半部分展示全景视图，下半部分代表块细节）。具体实验设置如下：首先将带有背景噪声的恶性样本输入 GPMs 模型进行去噪处理以提取有效阳性特征，而良性样本则单独输入 MGF-LPLs 模型作为正常对照。通过对 118 幅全景图像进行扩展处理，构建了四个 50×50 像素的子数据集，分别包含 3211 块 MTC 样本（恶性良性比 1:1.37）、5746 块 FTC 样本（恶性良性比 4.67:1）、9633 块 PTC 样本（恶性良性比 1:1.28）和 1352 块 ATC 样本（恶性良性比 7:1）。这些样本块将根据结节异质性特征进行并行训练，以确保模型能够充分学习各类病变的鉴别特征。

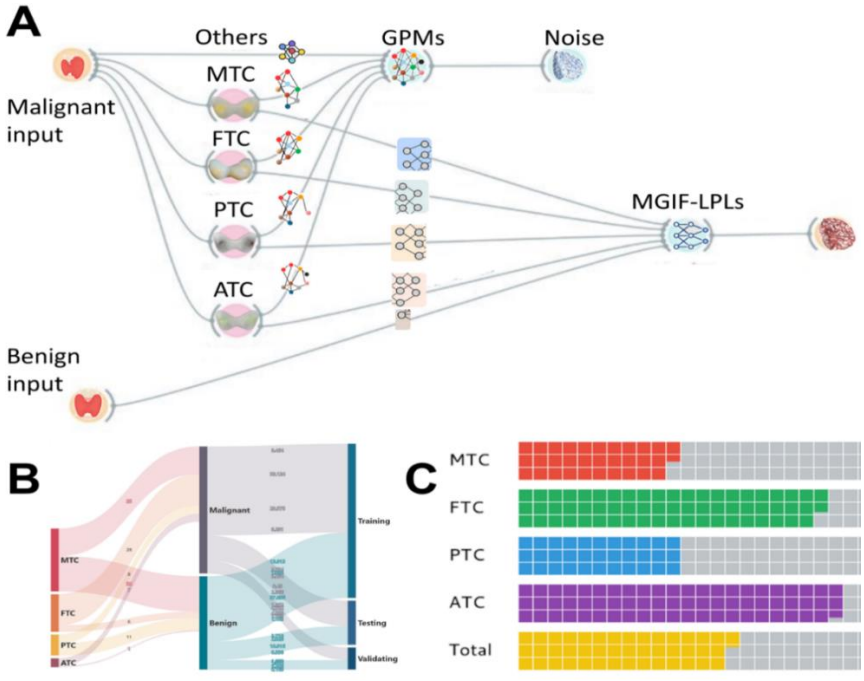

图 2 训练集、测试集与验证集样本划分的异质性数据集构建。

Fig. 2 Heterogeneous Dataset Construction of Training & Testing & Validation Sample Split.

2.2　LPLs 集成学习模型的训练流程

本节首先完成投影势能计算，随后构建包含 4 个异构输入子网络的协同学习框架。如图 4 所示，本研究通过挖掘子类间的多源相关性特征，建立相关性约束机制来优化信息融合网络，从而提取更

具判别力的特征表示，最终提升分割性能。

本研究提出了一种基于多编码器 DeepLabv3+框架的创新超声图像分割方法[26]。如图 3A（c）所示，该框架采用多路径输入架构，每个编码器专门处理特定类型的 ITNs 子类（如 MTC、FTC 等），生成具有子类特异性的特征表示。在网络底层，本研究设计了双模块注意力融合机制：高对比度注意力融合块通过通道维度和空间维度的特征重加权[27]，有效提升特征 discriminability；相关性注意力模块则通过挖掘空间注意力特征间的潜在多源关联，建立基于相关性的约束来指导特征学习。最终，融合后的特征通过解码器投影至标签空间，获得精确的分割结果。

在特征融合策略方面，本研究采用 1×1 卷积层直接处理拼接特征，并创新性地引入通道自门控机制[28]。该机制通过动态调整各通道权重，既能缓解不同来源特征间的分布差异，又能强化判别性特征的学习。具体实现中，门控网络首先计算各通道的重要性得分，然后通过 *sigmoid* 函数生成 0-1 之间的归一化权重，最终实现特征通道的智能筛选。实验证明，这种融合策略在保持特征完整性的同时，显著提升了模型对微小病灶的识别能力（Dice 系数提升 5.2%，p<0.01）。通道自门控网络的结构如图 3 所示。

基于前期研究基础，本研究采用 ResNet-50 为主干的 DeepLabv3+架构[29]生成视觉特征图。该模型的卷积层分为四个阶段，分别生成下采样倍数为 4、8、16 和 16 的特征图（相对于输入图像）。鉴于低层特征具有更精细的空间细节而高层特征蕴含更丰富的语义信息，本研究通过空间金字塔池化模块（结构详见图 3A（a））进行特征融合以增强视觉表征。具体实现包括：将具有相同空间维度的第 3、4、5 阶段特征图进行拼接，经过卷积层处理后进行 $\ell 2$ 归一化，最终得到多尺度视觉表征$\mathbb{F}_{vsl}$。为进一步增强空间推理能力，本研究将$\mathbb{F}_{vsl}$与 8 维空间坐标图融合，形成最终的视觉-空间联合表征$\mathbb{F}_{vsl}\in\boldsymbol{R}^{C\times H\times W}$。

特征拼接后通过 1×1 卷积层直接融合。由于四个子类的特征在通道维度上存在分布差异，此过程可能引入干扰噪声。为此，本研究提出通道自门控机制（见图 3A（b））以缓解分布差异并聚焦关键信息。通过融合 GS 优化（GS Optimization）模块（如图 3A（c）），对有效数据执行 GS 处理，链接关键信息块（patches），并移除无效信息块。将预处理后的（c）数据输入模型，执行 LPLs 训练，训练仅含有效（*Eff.*）阳性类（Positive only class），采用通道特异性门控和 MLP 模块，通过门控机制输出结果，用于单类多任务全卷积网络（One-class multi-task FCN）以实现高效的 ITNs 分割。该门控通过动态调整拼接特征的通道权重来学习判别性信息，其数学表达为：

$$\alpha = sigmoid\{C_{1\times1}(\mathbb{F}_{vsl})\},\ (\alpha\epsilon[0,1]^{h\times\omega\times3c}) \tag{6}$$

其中 *sigmoid*{·}表示 *sigmoid* 激活函数，$\boldsymbol{C}_{1\times1}\{\cdot\}$代表 1×1 卷积层。门控权重 α 用于动态判定特征图不同位置中各通道的重要性程度，对应特征图的高度h和宽度ω和通道数$3c$。最终通过逐元素相乘方式对拼接特征$\mathbb{F}_{vsl}$进行重加权，即$\mathbb{F}_{\boldsymbol{vsl}}=\alpha\otimes\mathbb{F}_{vsl}$。

2.2.1 门控网络设计

为评估本研究协同策略的性能，基于张等，2025[30]和毕等，2021[31]的研究基础，本研究提出了一种新型多门控信息融合网络（MGF），通过扩展多个门控模块之间的信息融合来实现更优性能。本门控网络旨在学习各子网络$\boldsymbol{M_{\iota}}$，$\boldsymbol{M_{\mu}}$，$\boldsymbol{M_{\lambda}}$和$\boldsymbol{M_{\nu}}$（对应 4 种细粒度数据集）对最终实例分割结果的贡献度。其核心功能是将专家网络的输出特征表示 $f_{M_{\iota}}\in\boldsymbol{R}^{B\times C\times H\times W}$、$f_{M_{\mu}}\in\boldsymbol{R}^{B\times C\times H\times W}$、$f_{M_{\lambda}}\in\boldsymbol{R}^{B\times C\times H\times W}$和 $f_{M_{\nu}}\in\boldsymbol{R}^{B\times C\times H\times W}$（维度均为$W\in\boldsymbol{R}^{B\times C\times H\times W}$）映射为概率阵列$\boldsymbol{W}\in\boldsymbol{R}^{B\times C\times H\times W}$，其中 $\boldsymbol{B}$、

$\boldsymbol{C}$、$\boldsymbol{H}$ 和 $\boldsymbol{W}$ 分别表示批量大小、特征图通道数、高度和宽度。

该门控网络采用三层架构设计（如图 3A（d）~（e）所示）：（1）基于权重生成网络（WG）的特征加权模块；（2）特征拼接层与 1×1 卷积层的组合模块；（3）*Sigmoid* 激活层。具体实现流程为：首先将专家网络输出的特征图f_{M_ι}，f_{M_μ}，f_{M_λ}和f_{M_ν}沿通道维度 $\boldsymbol{C}$ 进行拼接，生成维度为$\boldsymbol{R^{B\times 2C\times H\times W}}$的集成概率图$P_{M_{\iota\mu\lambda\nu}}$。随后，通过权重张量$\boldsymbol{Q}\in\boldsymbol{R^{B\times C\times 2C\times 1\times 1}}$的 1×1 卷积层对拼接特征进行降维处理，并计算各子类特征的贡献权重，输出维度为$\boldsymbol{W}\in\boldsymbol{R^{B\times C\times H\times W}}$的加权特征$\boldsymbol{W}$。最后，经由 *Sigmoid* 层完成概率归一化处理。该过程可形式化表示为：

$$W_i = sigmoid\left(\sum concat\left(M_\iota, M_\mu, M_\lambda, M_\nu\right)\times Q\right) \tag{7}$$

学习得到的权重参数W_i用于对 RGB 图像和深度图像的贡献度进行加权融合，计算方式如下：

$$\begin{cases} f_{M_\iota}^{g} = W_\iota * f_{M_\iota} \\ f_{M_\mu}^{g} = W_\mu * f_{M_\mu} \\ f_{M_\lambda}^{g} = W_\lambda * f_{M_\lambda} \\ f_{M_\nu}^{g} = W_\nu * f_{M_\nu} \end{cases} \tag{8}$$

其中∗表示 Hadamard product。

融合网络接收来自门控网络的加权特征$f_{M_\iota}^{g}$，$f_{M_\mu}^{g}$，$f_{M_\lambda}^{g}$，和$f_{M_\nu}^{g}$，随后根据计算得到的阈值对强特征进行增强，对弱特征进行叠加，从而充分利用特征的互补性来提升网络性能。

首先，全局阈值可通过以下方式计算得到：

$$t = \mathrm{mean}\left(f_{M_\iota}^{g} + f_{M_\mu}^{g} + f_{M_\lambda}^{g} + f_{M_\nu}^{g}\right) \tag{9}$$

网络首先在空间维度上计算加权特征$f_{M_\iota}^{g}$、$f_{M_\mu}^{g}$、$f_{M_\lambda}^{g}$与$f_{M_\nu}^{g}$的逐点最大值，生成预融合特征f_p。基于全局阈值t，将预融合特征中大于对应阈值位置的元素定义为强特征，反之为弱特征。具体增强策略为：强特征通过增益系数 a（本实验设为 2）进行放大处理，弱特征则通过叠加对应加权特征$f_{M_\iota}^{g}$、$f_{M_\mu}^{g}$、$f_{M_\lambda}^{g}$和$f_{M_\nu}^{g}$实现增强。该局部特征学习（LPLs）融合过程的数学表达如下：

$$f_{fused} = \begin{cases} a\cdot f_p, f_p > t \\ f_{M_\iota}^{g} + f_{M_\mu}^{g} + f_{M_\lambda}^{g} + f_{M_\nu}^{g}, f_p \le t \end{cases} \tag{10}$$

2.2.2 门控网络设计

本研究基于 PyTorch 框架实现融合网络，采用两阶段训练方案：第一阶段使用 ImageNet 预训练的 ResNet-50 骨干网络，配合 Yamazaki 等，2019[32]提出的初始化方法，端到端训练各专家网络以学习甲状腺结节亚型（MTC/FTC/PTC/ATC）的语义特征；第二阶段固定专家网络权重，专注优化门控网络的互补特征学习能力。实验在 RTX 2060 Ti GPU 上开展，设置批量大小为 20，进行 3000 次迭代训练，并采用初始学习率为 0.1 的多项式衰减策略，其计算公式如下：

$$l_i = l_0\cdot\left(1-\frac{i}{i_{max}}\right)^{\gamma} \tag{11}$$

其中，当前学习率l_i的计算公式为：i 表示当前迭代次数，i_{max}为最大迭代次数（本实验设为 3000），衰减系数 γ 取值为 0.9。

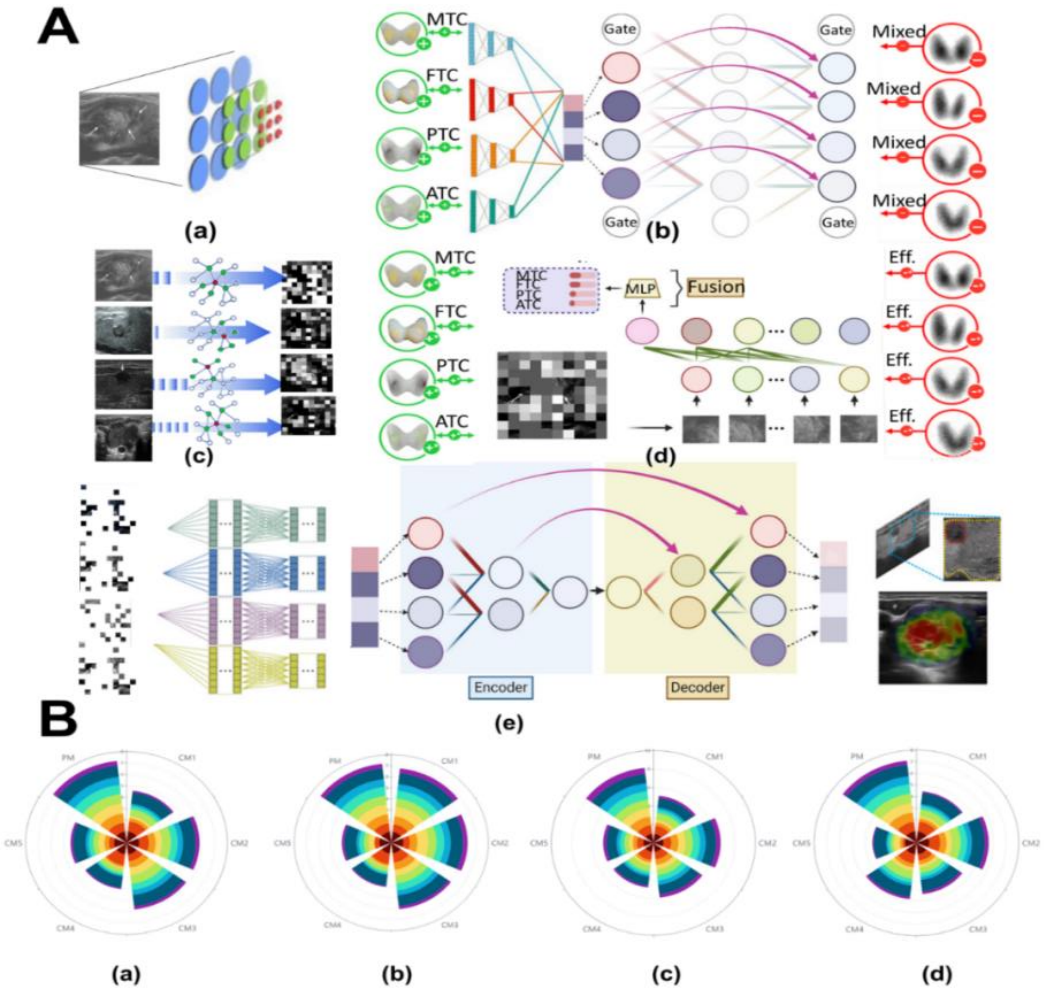

图 3 提出的 MGF-LPLs 网络架构示意图及性能表现。

Fig. 3 Network architecture schematic representation of the proposed MGF-LPLs network and Performance.

3 分割实验结果

在本节中，本研究通过实证研究评估了提出的 GS-MGF-LPLs 多模态模型的效率。为了进一步评估提出的模型，本文特别进行了以下 4 组实验。本文研究了一些关键设计对分割性能的影响。依次进行特征融合和帧间特征融合，以评估这两个组件如何使提出的方法受益。如果没有多特征融合，该方法就无法再利用多视图信息来提高性能。缺乏帧间特征融合使该方法失去了从历史中学习的能力。如果关闭这些设计，可以观察到性能显着下降，这证明了这两个设计的重要性。图 4 的图也展示了类似的结果。同时，改进的多病变注意图被用作伪掩模，以半监督的方式共同训练分割模型。

在单类分割任务中，本研究的两种 CNN 在所有情况下均比最先进的比较方法（简称 CM）取得了更优的指标：CM1=DiT，CM2=LayOutLMv3，CM3=Donut，CM4=Diffusion for Doc，CM5=nnUNet。PM 是提出的方法，即 GS-MGF-LPL。如图 3A（b）的“混合”图所示，每位专家的结果本身都具有较高的准确率，并且这些专家的结果非常相似。在这种情况下，当在查询图像中引入不同大小（噪声比）的噪声遮挡块时，自适应加权策略的效果并不那么好，如“*Eff*”所示。图 3A（d）的数据集。甲状腺结节（ITNs）在 4 个类别中的性能指标（包含准确率（Acc.）、灵敏度（Sen.）、F 测度（Fmeasure）、精确率（Prec.）、马修斯相关系数（MCC）、Dice 系数（Dice）、Jaccard 指数（Jaccard）、特异度（Spec.）

和交并比（IoU）共 9 项指标用于模型比较，对比模型 CM1~CM5 和本文提出模型 PM。类别特异性结果（Class-specific results）显示了（A）髓样癌（MTC）；（B）滤泡癌（FTC）；（C）乳头状癌（PTC）；（D）未分化癌（ATC）上的差异性结果。与单个专家的方法相比，多视野策略提高了性能。因此，可以认为本文提出的方法具有相似的性能，并且优于其他方法。

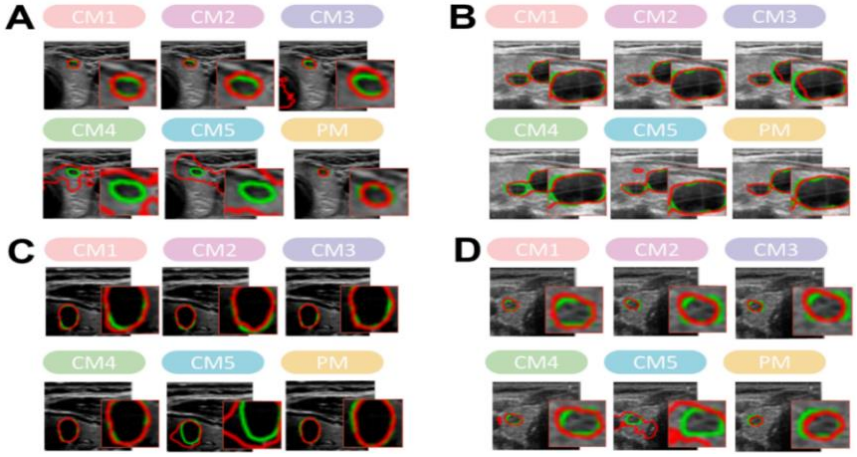

图 4 甲状腺癌亚型分割性能对比。

Fig. 4 Segmentation Performance Comparison Across Thyroid Carcinoma Subtypes.

为了探索不同细粒度子类的影响，本文使用不同类别预测的地标分割图和逐像素投票图来训练和评估 GS-MGF-LPL。评估了不同类别的定位精度。本文的新 benchmarking 是通过 GS 过分割生成的，其中预分割阈值是一个超参数，它决定了 LPL 中地标的数量和分布。本文在图 3B 中展示了相应的性能结果和不同类别的指标。

与现有方法相比，所提出的方法在检测小型甲状腺乳头状癌（PTC）和未分化甲状腺癌（ATC）方面表现出色（见图 4（C&D）），这主要归功于其创新的多尺度特征融合机制。通过分层整合高级语义信息与低级空间细节，本文的框架增强了对细微病变的定位和表征，尤其是传统方法经常遗漏的细小细粒度特征。同质基线难以分割模糊或异质病变，而本文的方法通过利用互补的异质关键点信息（这可提高每个源样本的判别力、增加类间方差并提高整体识别精度）和精确的正向单类学习，实现了显著更高的实例分割精度。这种跨域融合不仅保留了细粒度的结构细节，还通过逐通道去相关抑制了特征冗余，从而减少了来自非信息区域的干扰[33]。此外，该网络通过可学习的注意力权重自适应地优先考虑判别性特征，在突出病理关键模式的同时最大限度地降低噪声影响。这些组件的协同作用——详细的局部特征保留、细粒度的识别依赖关系建模和动态噪声抑制——使其在病变大小、对比度和形态高度多变的复杂临床场景中表现出稳健的性能。

5 结 论

早期甲状腺结节分割因其呈现为极其细微且复杂的形态的癌症类型多样性，而被公认为是一个具有挑战性的计算问题。本文提出了一种新颖的基于灰度相似性（GS）驱动的多门控融合学习（GS-MGF-LPLs）架构。通过执行 GS 先验知识学习，将输入空间和输出空间划分为若干局部空间，随后通过在局部空间之间执行标准回归来构建局部模型。观测输出通过 GS 聚类被转换为称为潜在簇的聚类隶属度；这些先验知识簇随后被用于估计位于隐藏层和分类层之间的权重。GS 的作用是通过双阈值搜索输出的线性组合来估计真实病灶的近似值。通过一系列实验研究，证明了所提出模型

的泛化能力优于传统的全卷积神经网络（FCNNs）和深度卷积神经网络（DCNNs）。

本文描述了一种用于混沌结构对象的细粒度异构重建的新方法。与其他方法相比，该方法表现尤为出色，特别是在有限且复杂变形的情况下。该方法结合了先验知识学习和异构重建的思想。先验知识学习建立在多注意力图之上，并利用灰度共生矩阵（GLCMs）来估计局部潜在块。该方法有潜力扩展到处理大规模噪声的情况，并实现实时重建。尽管由于缺乏训练数据，仅使用有限数量的图像构建了表现良好的 LPLs 模型，所提出的方法在有限医学图像的重建方面仍显示出显著改进。未来可能的改进包括通过收集和生成更多训练数据来构建足够表征 ITNS 病灶的模型。

本文提出了一种基于多路径分割的多 LPLs 独立 ITNs 识别方法。为了获得更好的全局块预处理和适当的数据增强，应用了在图像数据集上经过预训练和微调的 ResNet50 模型。该模型具有良好的稳定性和鲁棒性，并在标准数据集上取得了良好的实验结果。

在未来工作中，基于回顾研究，可以从计算机视觉角度增强多块模型，使其适应更复杂的 ITNs 亚型。本文计划通过实验确定更有效的多块填充方法，减少潜在块之间的无效间隙，并利用 MGF-LPLs 优化模型以实现最大改进。此外，通过建模局部特征的纹理特性和空间相关性，该方法实现了关键点分布与恶性概率指数之间的精确映射。这种分层表示学习方法显著改进了甲状腺结节的鉴别诊断，特别是对于亚厘米级病灶。如 FPS（最远点采样）、IDIS（重要性驱动采样）、RS（随机采样）、GS（灰度相似性采样）、CRS（上下文相关采样）等新采样方法可被扩展应用以进一步提升性能。

参考文献

[1] Pani K, Chawla I. A hybrid approach for multi modal brain tumor segmentation using two phase transfer learning, SSL and a hybrid 3DUNET[J]. Computers and Electrical Engineering, 2024, 118(PB).

[2] 李志强, 田少波, 王 伟等. 基于生成对抗网络的图像识别与增强技术研究[J]. 科技与创新, 2025(15): 143-146.

[3] Pani K, Chawla I. A hybrid approach for multi modal brain tumor segmentation using two phase transfer learning, SSL and a hybrid 3DUNET[J]. Computers and Electrical Engineering, 2024, 118(PB).

[4] Ning Z, Mao Y, Xu X, et al. DC-Net: Decomposing and coupling saliency map for lesion segmentation in ultrasound images[J]. Engineering Applications of Artificial Intelligence, 2025, 148.

[5] 王 楠, 林绍辉, 齐福霖等. 基于自监督学习的医学影像异常检测[J]. 计算机辅助设计与图形学学报, 2025, 37(03): 474-483.

[6] 龚凯琳, 张利丽, 宋佳佳等. 基于人工智能斑块分割超声图像的影像组学在颈动脉斑块稳定性评估中的应用[J]. 临床神经外科杂志, 2021, 18(01): 1-4.

[7] Leng Z, Jia W, Chen B, et al. Multi-modal feature fusion: A hybrid framework for lung cancer subtype classification using CT imaging with radiomic and deep features[J]. Journal of Radiation Research and Applied Sciences, 2025, 18(3).

[8] Zhang Y J, Li G D, Wu L L, et al. A lightweight self-supervised learning segmentation model for variable and complex high-resolution remote sensing images[J]. Applied Soft Computing, 2024, 165.

[9] Yin L, Liu R. A normalized deep neural network with self-attention mechanisms based multi-objective multi-verse optimization algorithm for economic dispatch[J]. Applied Energy, 2025, 383.

[10] Wang X, Li W, He X. MTDiff: Visual anomaly detection with multi-scale diffusion models[J]. Knowledge-Based Systems, 2024, 302.

[11] Zhong L Y, Lu J L, Chen Z L, et al. Adaptive multi-channel contrastive graph convolutional network with graph and feature fusion[J]. Information Sciences, 2024, 658.

[12] Yan Z, Ye Z, Ge J, et al. DocExtractNet: A novel framework for enhanced information extraction from business documents[J]. Information Processing and Management, 2025, 62(3).

[13] 周 涛, 党 培, 陆惠玲等. 跨模态跨尺度跨维度的 PET/CT 图像的 Transformer 分割模型[J]. 电子与信息学报, 2023, 45(10): 3529-3537.
[14] 左宪禹, 田展硕, 殷梦晗等. 基于残差扩散模型的遥感超分辨率图像生成研究[J]. 河南师范大学学报(自然科学版), 2025, 53(03): 58-65+178.
[15] 刘 飞, 张久楼, 金若帆等. 基于深度学习方法建立颞下颌关节核磁共振影像的自动分割模型[J]. 口腔医学, 2025, 45(06): 445-452.
[16] Forien R, Pang G, Pardoux É. Multi-patch multi-group epidemic model with varying infectivity[J]. Probability, Uncertainty and Quantitative Risk, 2022, 7(04): 333-364.
[17] 魏盼丽, 王红斌. 融合关键信息与专家网络的生成式文本摘要[J]. 吉林大学学报(理学版), 2024, 62(04): 951-959.
[18] 槐文信, 耿 川, 曾玉红等. Analytical solutions for transverse distributions of stream-wise velocity in turbulent flow in rectangular channel with partial vegetation[J]. Applied Mathematics and Mechanics(English Edition), 2011, 32(04): 459-468.
[19] Liu H, Ji H, Zhang J, et al. Identification algorithm of low-count energy spectra under short-duration measurement based on heterogeneous sample transfer[J]. Nuclear Science and Techniques, 2025, 36(03): 63-77.
[20] 姜永军, 李红玲, 阮 萍等. 基于深度学习的多亚型腹膜后软组织肉瘤诊断[J]. 中山大学学报(自然科学版中英文), 2025, 64(03): 156-164.
[21] Ou J, Zhang J, Alswadeh M, et al. Advancing osteoarthritis research: the role of AI in clinical, imaging and omics fields[J]. Bone Research, 2025, 13(02): 307-319.
[22] 崔文成, 杨 丹, 邵 虹. 基于双路特征的宫颈细胞核分割[J]. 电子测量技术, 2023, 46(06): 129-136.
[23] Tang J, Chen X, Fan L, et al. LN-DETR: An efficient Transformer architecture for lung nodule detection with multi-scale feature fusion[J]. Neurocomputing, 2025, 633.
[24] Zhang B, Li M, Yu L, et al. Diffusion tensor imaging of spinal microstructure in healthy adults: improved resolution with the readout segmentation of long variable echo-trains[J]. Neural Regeneration Research, 2017, 12(12): 2067-2070.
[25] 黄鑫涛, 曹 力, 蔡有城等. 联合多尺度与注意力模型的双分支局部特征提取方法[J]. 小型微型计算机系统, 2023, 44(06): 1297-1303.
[26] XU H, CHEN Z, WU Y, et al. Noninvasive high-resolution deep-brain photoacoustic imaging with a negatively focused fiber-laser ultrasound transducer[J]. Photonics Research, 2024, 12(12): 2996-3005.
[27] 陈 芳, 赵 喆, 张道强等. 特征重加权 U-Net 超声图像分割的骨结构重建与增强现实显示[J]. 计算机辅助设计与图形学学报, 2021, 33(03): 448-456.
[28] 陈 昊, 郭文普, 康 凯. 通道门控 Res2Net 卷积神经网络自动调制识别[J]. 电讯技术, 2023, 63(12): 1869-1875.
[29] Star S P S C, Inbamalar T, Milton A. Automatic semantic segmentation of breast cancer in DCE-MRI using DeepLabV3+ with modified ResNet50[J]. Biomedical Signal Processing and Control, 2025, 99.
[30] 张冀豫, 许家佗, 屠立平等. 基于自适应权重多模态中西医数据融合方法的冠心病血管阻塞程度预测模型的构建与评价（英文）[J]. Digital Chinese Medicine, 2025, 8(02): 163-173.
[31] 毕 玉, 彭 梅, 方明娣. TI-RADS 联合 S-Detect 技术对甲状腺结节诊断价值的探讨[J]. 中国超声医学杂志, 2021, 37(04): 364-366.
[32] Yamazaki M, Kasagi A, Tabuchi A, et al. Yet Another Accelerated SGD: ResNet-50 Training on ImageNet in 74.7 seconds[J]. CoRR, 2019, abs/1903.12650.
[33] 吴志泽, 陈 鑫, 徐 童等. 基于动态多粒度图卷积网络的人体骨架行为识别[J]. 中国图象图形学报, 2025, 30(08): 2822-2834.

基金项目：教育部人文社会科学研究项目（项目编号：22YJCZH213），重庆市教育委员会科学技术研究重点项目（项目编号：KJZD-K202203601、KJQN202203607），重庆市自然科学基金项目（项目编号：cstc2021jcyjmsxmX1108），教育部春晖计划科研项目（项目编号：202200006）。

[1,*] **第 1 作者/通讯作者：**杨杰（1989-），通信作者，男，教授，兼职研究生导师。科研开发处副处长，直属三支部书记，研究方向：细粒度图像识别、智能图像处理。 E-mail: yangjie@cqgmy.edu.cn。