

基于跨模态生成与交叉注意力的乳腺超声诊断

傅力嘉¹, 林艳萍¹, 李娜², 贾超^{3,4}, 李刚^{3,4}, 杜联芳^{3,4}, 李凡^{4,5}

1. 上海交通大学机械与动力工程学院, 上海 200240; 2. 上海理工大学健康科学与工程学院, 上海 200093; 3. 上海市第一人民医院超声医学科, 上海 201620; 4. 上海交通大学医学院, 上海 200025; 5. 上海市胸科医院超声科, 上海 200030

【摘要】目的:针对常规B超在乳腺癌筛查中因读图者间一致性不佳、微小病灶恶性征象不典型导致的准确性受限问题,以及超声造影因成本高、操作复杂而难以普及的临床矛盾,提出一种“分割-生成-分类”三阶段级联诊断框架,旨在仅利用常规B超图像实现高精度的乳腺肿瘤良恶性诊断。**方法:**首先,采用CNN-Transformer架构的分割网络从B超图像中提取肿瘤区域。接着,构建基于条件生成对抗网络的跨模态生成模型,利用预训练的ResNet-34编码器从分割后的B超中生成对应的超声造影图像。最后,设计以Inception-v3为骨干的双分支交叉注意力网络,通过动态特征融合实现对真实B超与生成造影信息的深度交互,完成对肿瘤良恶性的精确诊断。**结果:**在包含677例患者的乳腺超声数据集上进行验证,该框架的受试者工作特征曲线下面积(AUC)达到0.930,显著优于多个单模态B超基线模型(最佳AUC为0.868)。消融实验证实病灶分割、跨模态生成以及交叉注意力特征融合模块对提升诊断性能的必要贡献。**结论:**本研究提出的框架验证了通过深度学习方法生成并利用多模态超声信息进行高精度诊断的可行性与有效性,能够以低成本、高效率的方式提供多模态诊断增益,有望为乳腺癌临床筛查提供有力的辅助决策支持,降低乳腺癌临床筛查的漏诊与误诊风险。

【关键词】乳腺癌; 乳腺超声; 超声造影; 辅助诊断; 生成对抗网络; 跨模态生成; 交叉注意力; 特征融合

【中图分类号】R318; TP391.41

【文献标志码】A

【文章编号】1005-202X(2026)07-0889-09

Breast ultrasound diagnosis based on cross-modal generation and cross-attention

FU Lijia¹, LIN Yanping¹, LI Na², JIA Chao^{3,4}, LI Gang^{3,4}, DU Lianfang^{3,4}, LI Fan^{4,5}

1. School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; 2. School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 3. Department of Ultrasound, Shanghai General Hospital, Shanghai 201620, China; 4. School of Medicine, Shanghai Jiao Tong University, Shanghai 200025, China; 5. Department of Ultrasound, Shanghai Chest Hospital, Shanghai 200030, China

Abstract: Objective To address the limited accuracy of conventional B-mode ultrasound (US) in breast cancer screening, resulting from poor inter-reader agreement and atypical malignant signs of tiny lesions, and resolve the clinical dilemma where contrast-enhanced ultrasound (CEUS) is hard to popularize due to its high cost and operational complexity, this study proposes a three-stage cascaded diagnostic framework of "segmentation-generation-classification" to achieve high-precision diagnosis of benign and malignant breast tumors relying solely on conventional US images. **Methods** A CNN-Transformer framework-based segmentation network was first employed to extract tumor regions from US images. Subsequently, a cross-modal generation model based on a conditional generative adversarial network was constructed, utilizing a pre-trained ResNet-34 encoder to synthesize corresponding CEUS images from the segmented US images. Finally, a dual-branch cross-attention network with Inception-v3 as the backbone was designed to achieve deep interaction between the real US and the generated CEUS information through dynamic feature fusion, thereby accomplishing accurate benign and malignant tumor classification. **Results** Validated on a breast ultrasound dataset including 677 patients, the proposed framework achieved an area under the receiver operating characteristic curve (AUC) of 0.930, significantly outperforming several single-modal US baseline models which reached a maximum AUC of 0.868. Ablation experiments verified the essential contributions of lesion segmentation, cross-modal generation, and cross-attention feature fusion modules in enhancing diagnostic performance. **Conclusion** This study validates the feasibility and effectiveness of generating and utilizing multimodal ultrasound

【收稿日期】2025-11-25

【基金项目】国家自然科学基金(52175020);上海市自然科学基金(21ZR1451400)

【作者简介】傅力嘉, 硕士研究生, 研究方向: 医学图像处理, E-mail: fulili@sjtu.edu.cn

【通信作者】李凡, 博士, 主任医师, 副研究员, 博士生导师, 研究方向: 超声生物治疗、超声造影及多模态超声诊疗, E-mail: medicineli@sjtu.edu.cn

information with deep learning for high-precision diagnosis. The proposed framework provides multimodal diagnostic benefits in a cost-effective and highly efficient way, holding great promise to provide robust auxiliary decision support for clinical breast cancer screening and reduce the risks of missed diagnoses and misdiagnoses.

Keywords: breast cancer; breast ultrasound; contrast-enhanced ultrasound; auxiliary diagnosis; generative adversarial network; cross-modal generation; cross-attention; feature fusion

前言

乳腺癌的早期精准筛查与诊断是降低患者死亡率、改善预后的关键环节^[1-2]。临床实践中,B模式超声(B-mode Ultrasound, US)因其无创、低成本及实时成像等优势,已在乳腺疾病筛查与诊断流程中获得广泛的临床应用^[3-4]。然而,常规US主要反映组织的声阻抗差异等结构信息,受限成像原理,图像信噪比低、对比度差,易受斑点噪声干扰,特别是在面对早期微小病灶或不典型恶性肿瘤时,单纯依赖形态学特征往往导致诊断特异性不足,观察者间一致性较差^[5-6]。

相比之下,超声造影(Contrast-Enhanced Ultrasound, CEUS)通过静脉注射微泡造影剂,动态显示肿瘤内部的微血管灌注模式,提供关键的功能性血流信息,显著提升良恶性鉴别的准确率^[7-9]。即使是CEUS视频序列中的关键单帧图像,也能提供关于血管形态的丰富信息,具有重要的诊断价值^[10]。尽管如此,CEUS检查存在成本高昂、需有创操作以及对操作者技术要求高等弊端,这些弊端限制其在基层医疗机构和大规模机会性筛查中的可及性^[11]。同时临床实践表明,在血流灌注模式上,一些用于鉴别良恶性的关键影像征象较为微妙,依赖人眼观察时辨识度有限^[12]。这就形成了一个核心的临床矛盾:B超易于获取但信息有限且存在判读差异,而CEUS虽信息丰富却因可及性以及内在诊断模糊性问题无法完全解决临床需求。

近年来,深度学习技术在医学影像分析领域取得突破性进展,为解决上述问题提供新思路^[13]。在病灶分割方面,以U-Net及其变体为代表的卷积神经网络(CNN)已成为基准方法^[14-15]。为了突破CNN的局部感受野限制,该领域最前沿、性能最强大的模型普遍采用CNN与Transformer相结合的混合架构,如SwinUnet和TransUNet^[16-17],结合CNN提取局部特征的优势与Transformer的全局上下文建模的优势,显著提升了复杂背景下的病灶定位精度。在获取功能信息方面,生成对抗网络(GAN)展现出强大的跨模态图像合成能力^[18]。在跨模态医学图像生成领域中,GAN和以pix2pix为例的图像翻译模型,已被证明在MRI合成CT等任务中的有效性^[19-21]。最近,文

献[22]在肿瘤消融评估的背景下,也证明了从B模式超声合成伪CEUS图像的可行性。

然而,现有的乳腺超声智能诊断研究多局限于单模态分析,或直接利用获取困难的真实多模态数据进行融合^[23-25],鲜有研究尝试通过计算生成的方式构建多模态诊断框架。当前,多模态数据的特征级融合已成为主流的研究范式^[26],基于注意力机制的动态融合策略已成为最前沿的研究方向之一。其中,交叉注意力机制成为实现模态间显式、定向信息交互的先进方法^[27]。

本研究针对前述的技术痛点与临床需求,提出一种基于跨模态图像生成与交叉注意力融合的乳腺超声智能诊断框架。本文的主要贡献总结如下:提出一种新颖的“分割-生成-分类”三阶段级联诊断框架,通过模拟并优化临床专家诊断流程,实现仅从常规B超图像进行高精度乳腺肿瘤良恶性分类;验证利用B超图像计算生成超声造影图像的可行性,为解决先进影像技术在常规筛查中应用受限的临床困境提供一种“以计算代替检查”的创新解决方案;构建一个深度优化的跨模态生成模型,通过引入预训练编码器和定制化的损失函数,优先保证生成图像的结构准确性,契合下游诊断任务的真实需求;设计一种基于交叉注意力的双流融合分类网络,融合真实B超图像与生成造影图像两种模态的图像特征,实现真实结构信息与生成式功能信息的深度动态交互,提高乳腺超声肿瘤的分类精度。

1 材料与方法

1.1 数据集与预处理

1.1.1 数据集构建 本研究所有数据均回顾性地收集于上海市第一人民医院超声科,收集2021年3月~2023年6月接受乳腺超声成像的乳腺病变患者的图像数据与临床信息。研究方案已获得该机构伦理审查委员会的批准(IRB编号:No.20230320034817764)。在纳入研究前,所有患者均签署知情同意书。为保护个人隐私,所有患者数据在用于研究前均经过严格的匿名化处理。

本研究构建了一个包含677例患者的双模态乳腺超声数据集。该数据集涵盖368例经病理证实的恶性病例与309例良性病例。所有影像均采用高频

线阵探头的超声诊断仪采集, CEUS 检查以 Sonovue 为造影剂。所有病灶区域均由具备丰富乳腺超声诊断经验的资深放射科医师进行手动勾画。

1.1.2 图像预处理 所有图像在送入网络前均经过一系列标准化的预处理步骤。首先通过非局部均值 (Non-Local Means, NLM) 算法抑制原始 US 图像中的散斑噪声, 随后利用限制对比度的自适应直方图均衡化 (Contrast Limited Adaptive Histogram Equalization, CLAHE) 技术增强肿瘤与周围组织的对比度。接着, 利用本框架第一阶段分割网络获得的肿瘤掩码, 通过掩膜操作精确提取出肿瘤的感兴趣区域 (Region of Interest, ROI)。最后, 对提取的 ROI 进行灰度化处理、裁剪并统一缩放至 256×256 像素, 作为后续生成与分类网络的标准输入。

1.2 总体框架

本研究设计了一种新颖的三阶段级联智能诊断

框架, 将复杂的端到端诊断任务分解为 3 个功能高度专门化子任务, 分别是病灶分割、跨模态图像生成与多模态分类, 如图 1 所示。首先, 在病灶分割环节, 经过预处理的 US 图像被送入一个高精度分割网络, 输出像素级掩码。接着, 图像进入跨模态生成模块, 在分割掩码的引导下, 模型从 US 图像中提取出肿瘤 ROI。该图像随后被送入一个基于生成对抗网络 (conditional Generative Adversarial Network, cGAN) 的图像生成模型, 该模型以肿瘤 ROI 的结构与纹理特征为条件, 生成与之对应的、富含微血管灌注模式等功能性信息的生成式超声造影 (generative Contrast-Enhanced Ultrasound, gCEUS)。最后, ROI-US 图像与对应的 gCEUS 图像组成双模态图像对, 进入一个双分支的分类模型。该模型利用交叉注意力机制实现两种模态的特征融合, 最终输出肿瘤良恶性的诊断结果。

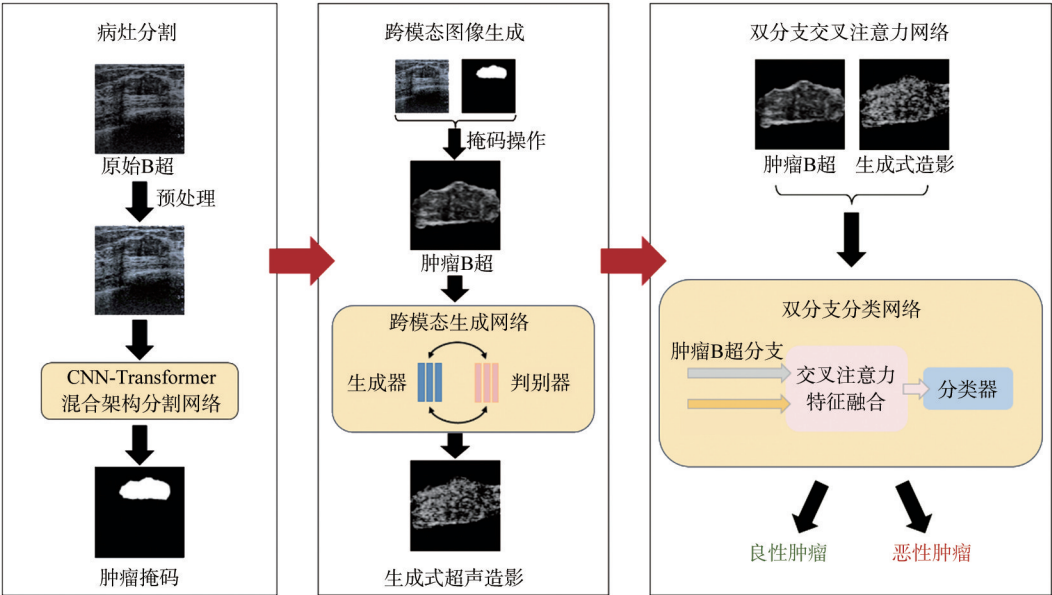


图1 “分段-生成-分类”三级级联智能诊断框架

Figure 1 Three-stage cascaded intelligent diagnostic framework of "segmentation-generation-classification"

1.3 病灶分割

在分割前端, 本研究采用已被广泛验证的 TransUNet 网络作为病灶分割模型。首先, 编码器利用 CNN 提取高分辨率的局部特征图, 随后在网络的瓶颈部分将 CNN 提取的深层特征图展平为序列化的图像块, 送入标准的 Transformer 编码器。最后, 解码器通过上采样和跳跃连接, 将 Transformer 编码的全局上下文信息与 CNN 编码器提供的多尺度高分辨率局部特征进行融合。这种 CNN-Transformer 的混合架构, 既能利用全局信息理解肿瘤的整体形态, 又能利用局部细节信息精准划定边界。

同时, 本研究在构建分割任务的真实标签时, 创新性地采用一种“掩码并集”策略。鉴于 US 图像主要反映肿瘤的结构边界 (MASK_{US}), 而 CEUS 图像则揭示其功能性的血供范围 (MASK_{CEUS}), 两者在边界上可能存在不完全一致。本研究将最终的真实标签 (MASK_{GT}) 定义为两种模态掩码的并集, 即:

MASK_{GT} = MASK_{US} ∪ MASK_{CEUS} (1)

该策略确保分割区域能完整地包含病灶的结构与功能信息, 为后续的生成和分类任务提供最全面的病灶信息。

1.4 跨模态图像生成

该模块是本框架的核心步骤,利用结构性的US图像生成功能性的gCEUS图像,将单模态输入转变为多模态分析。为此,本研究构建了一个基于cGAN的跨模态图像生成模型。

1.4.1 生成网络架构 该生成模型由一个生成器(Generator)和一个判别器(Discriminator)构成,其架

构如图2所示。生成器采用带有预训练编码器的U-Net架构。具体而言,其编码器部分采用在ImageNet数据集上预训练的ResNet-34网络,利用迁移学习的优势,帮助模型理解和重建肿瘤的复杂形态与边界。解码器部分则采用标准的对称结构,通过一系列转置卷积进行上采样,采用跳跃连接融合来自编码器对应层级的特征图。

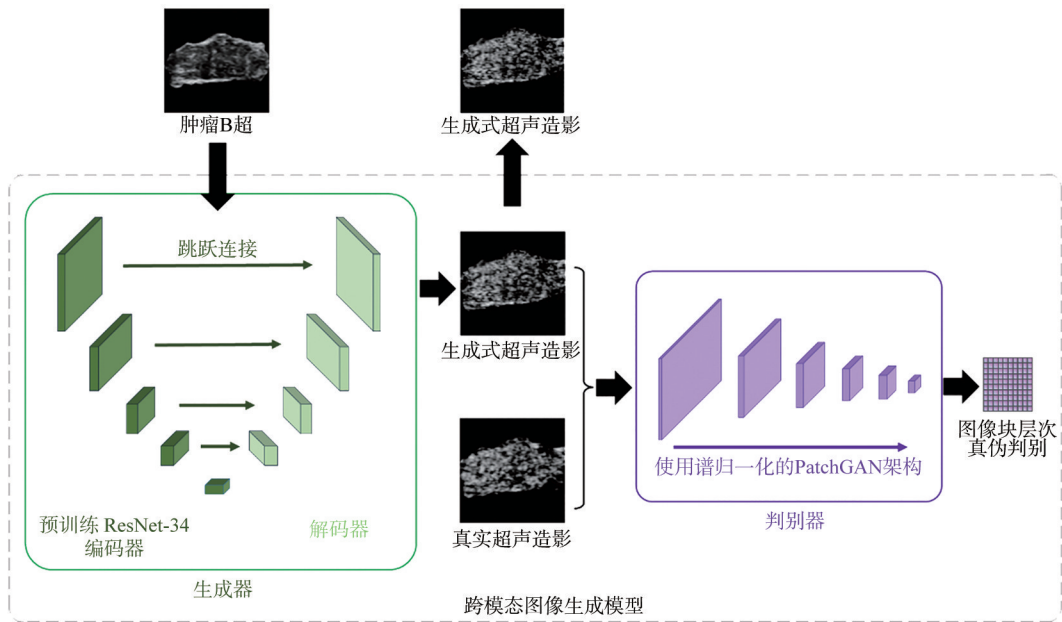


图2 跨模态生成网络架构图
Figure 2 Cross-modal generation network architecture

判别器采用带有谱归一化的 PatchGAN 架构。本研究使用的判别器将输入图像划分为N个输入的图像块,并独立对每个图像块的真实性和进行判别,最终输出一个概率图,图中的每个值代表了其对应图像块为真的概率。这迫使生成器不仅要关注全局的真实性,更要保证图像在局部纹理和细节上的高保真度。同时,为稳定GAN的训练过程,判别器的所有卷积层均应用了谱归一化(Spectral Normalization, SN),对每一层权重矩阵的谱范数进行约束。

1.4.2 目标函数 为了生成结构准确、细节真实且噪声受控的高质量gCEUS图像,生成器的训练采用了一个精心设计的复合目标函数,该函数由4个加权损失项构成:

$$L_G=L_{cGAN}(G,D)+\lambda_{L1}L_{L1}(G)+\lambda_{perc}L_{perc}(G)+\lambda_nL_n(G) \quad (2)$$

其中,G和D分别代表生成器和判别器,λ为各项损失的权重系数。各项损失的具体定义和作用如下:条件对抗损失(L_{cGAN})是GAN框架的核心,采用最小二乘GAN(LSGAN)的目标函数驱动生成器学习真实CEUS图像的数据分布,提升图像的真实感。L1损失

(L_{L1})计算生成图像与真实目标图像之间的平均绝对误差(Mean Absolute Error, MAE)。该损失直接惩罚像素级的差异,保证生成图像在肿瘤的整体轮廓、尺寸和位置等低频信息上与目标图像保持高度一致,其权重λ_{L1}设为50.0。感知损失(L_{perc})采用基于学习的感知图像块相似度(Learned Perceptual Image Patch Similarity, LPIPS)作为优化指标。该损失通过1个在ImageNet上预训练的VGG网络提取生成图像和真实图像的深层特征图,并计算这些特征图之间的距离,在高级语义特征的层面λ_{perc}上进行限制,以此引导模型学习生成具有高诊断价值的微血管细节等功能性信息与纹理。其权重设为1.0。总变差损失(L_n)是正则化项,惩罚图像中相邻像素间的剧烈变化,抑制生成图像中的高频噪声和伪影。其权重λ_n设为0.0001。

1.5 双分支交叉注意力分类网络

为融合US图像的结构信息与gCEUS图像的功能信息,本研究设计一个双分支交叉注意力网络,其整体架构如图3所示。

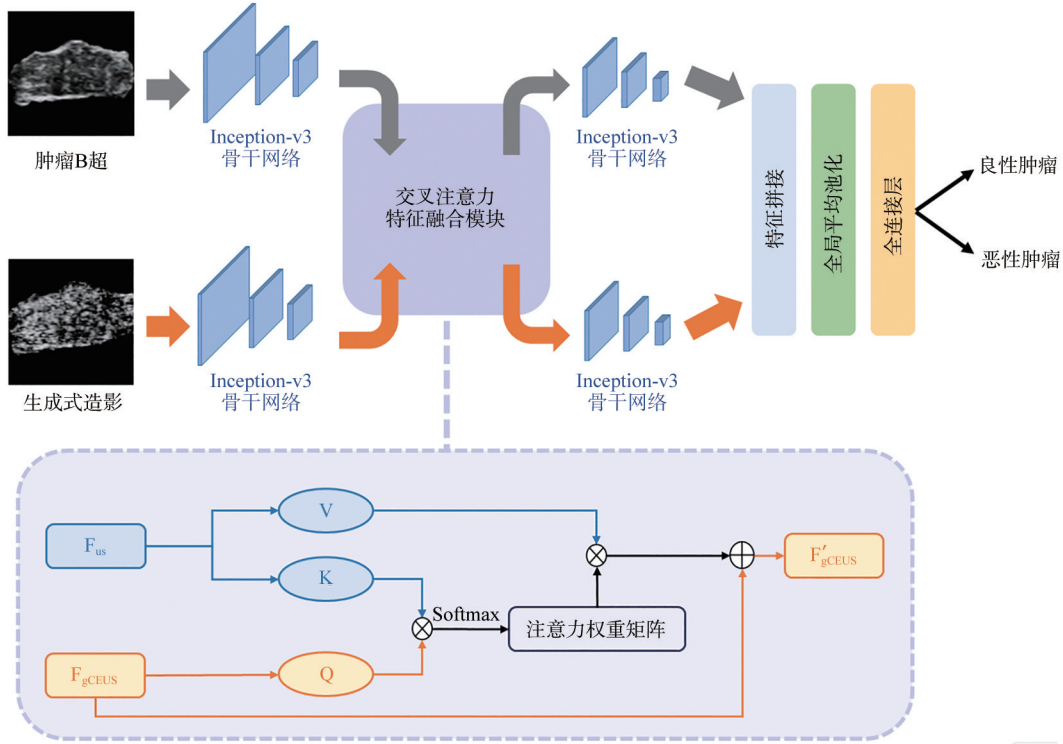


图3 双分支交叉注意力网络架构图

Figure 3 Overall architecture of the dual-branch cross-attention classification network

该网络采用并行的双分支设计,分别处理ROI-US图像与gCEUS图像。每个分支均采用一个在ImageNet数据集上预训练的Inception-v3网络作为骨干特征提取器。两种图像首先并行地通过各自骨干网络的前几个阶段,提取出多尺度的深层特征图。在骨干网络的中间层级,两个分支提取的特征图被送入交叉注意力模块进行深度交互融合。融合后得到的增强特征图,继续通过骨干网络的剩余层,再进行全局平均池化,最后通过一个全连接层和Softmax激活函数,输出最终的良恶性二分类概率。

基于交叉注意力机制的特征融合模块的具体实现如图3所示。在网络中间层级,从两个分支分别提取的特征图 F_{US} 和 F_{gCEUS} 进入交叉注意力融合模块。两种模态的特征融合过程是双向的。在一个方向上,使用gCEUS增强US特征。首先,通过3个独立的可学习线性投影,将输入特征图映射为注意力机制所需的3个核心向量,包括查询(Query, Q)、键(Key, K)和值(Value, V)。在此方向上, F_{US} 被用作查询的来源, F_{gCEUS} 被用作键和值的来源:

$$Q_{US} = F_{US} W_Q, K_{gCEUS} = F_{gCEUS} W_K, V_{gCEUS} = F_{gCEUS} W_V \quad (3)$$

然后,通过计算 Q_{US} 与所有 K_{gCEUS} 的点积相似度,经过Softmax归一化,得到一个注意力权重矩阵,表示 F_{US} 上的某个空间位置对 F_{gCEUS} 上的所有空间位置赋予多大的关注度:

$$\text{Attention}(Q_{US}, K_{gCEUS}) = \text{soft max} \left(\frac{Q_{US} K_{gCEUS}^T}{\sqrt{d_k}} \right) \quad (4)$$

最后,用该权重矩阵去加权 V_{gCEUS} ,得到一个被gCEUS信息调制的特征表示。该结果通过残差连接与原始的 F_{US} 相加,得到由gCEUS增强的US特征 F'_{US} :

$$F'_{US} = F_{US} + \text{Attention}(Q_{US}, K_{gCEUS}) V_{gCEUS} \quad (5)$$

在另一个方向上,使用US增强gCEUS特征。上述过程对称地进行,以 F_{gCEUS} 为查询来源,以 F_{US} 为键和值的来源,得到由US增强的gCEUS特征 F'_{gCEUS} 。

最终, F'_{US} 和 F'_{gCEUS} 继续通过骨干网络的剩余层进行处理,在通道维度上进行拼接,形成一个深度融合的特征表示 $F_{US-gCEUS}$,用于后续的分类任务。

2 实验结果

2.1 实验设置与评估指标

为了充分利用有限的医学影像数据并客观评估模型的泛化能力,本实验采用五折交叉验证策略。数据划分遵循分层采样原则,保持良恶性样本比例不变,将总计677例图像样本随机划分为5个互斥的子集。在每轮独立实验中,依次选取1个子集(20%,共136例,其中恶性74例、良性62例)作为测试集,其余4个子集(80%,共541例,其中恶性294例、良性247例)作为训练集。为了在训练过程中监控模型收

敛状态并进行超参数调优,本研究从每轮的训练集中随机按12.5%的比例划分出验证集。整体上,训练集、验证集、测试集的比例为7:1:2。最终各项分类性能指标均取5轮测试结果的平均值。

本研究采用在临床诊断研究中广泛应用的分类性能评估指标,包括准确率(Accuracy, Acc)、灵敏度(Sensitivity, Sens)、特异度(Specificity, Spec)、F1分数(F1-Score)、受试者工作特征曲线下面积(Area Under the Curve, AUC)。

所有模型均基于PyTorch框架实现,并在NVIDIA TU104 GeForce RTX 2080 GPU上进行训练与评估。分割模块采用Adam优化器进行训练,学习率为 1×10^{-4} ,损失函数为Dice损失与交叉熵损失。生成模块中,跨模态生成网络训练400个周期,批处理大小为16,优化器采用Adam,学习率设置为 2×10^{-4} 。在分类模块中,所有分类模型均训练100个周期,批处理大小为16。优化器同样采用Adam,学习率设置为 1×10^{-4} 。为防止过拟合,在全连接层后加入Dropout,其丢弃率设置为0.5。此外,为评估模型性能的稳定性和对比实验的统计学意义,本研究通过自举法计算AUC的95%置信区间(95% CI),采用

DeLong检验比较不同模型间AUC的差异, $P<0.05$ 为差异具有统计学意义。

2.2 对比实验

本文将提出的智能诊断框架分类性能与一系列仅使用US的单模态基线模型及一个理论性能模型进行全面对比。单模态基线模型涵盖了ResNet-50、Inception-v3等经典架构,以及EfficientNet-B7、Swin Transformer等先进架构。理论性能模型采用与本文方法相同的智能诊断框架,但其输入为真实的US与真实CEUS图像对,代表了在拥有真实多模态数据时的理想性能上限。

对比实验的定量结果如表1所示。本文模型在所有评估指标上均显著优于全部单模态基线模型。本文模型的AUC达到0.930(95% CI: 0.913~0.947),相较于表现最强的单模态基线EfficientNet-B7(AUC=0.868, 95% CI: 0.844~0.892),实现了6.2%的显著提升,且差异具有显著的统计学意义($P<0.001$)。使用真实双模态数据的理论性能模型的AUC为0.946,这与最强单模态基线(AUC=0.868)之间的性能差距代表引入真实CEUS信息所能带来的全部潜在诊断增益,而本框架的性能增益与理论性能模型十分接近。

表1 模型分类性能对比
Table 1 Comparison of model classification performance

模型	模态	AUC(95% CI)	准确率	敏感度	特异度	F1分数	P值
本文模型	US+gCEUS	0.930(0.913~0.947)	0.854	0.831	0.887	0.864	-
理论性能模型	US+CEUS	0.946(0.930~0.962)	0.869	0.896	0.836	0.880	<0.010
EfficientNet-B7	US	0.868(0.844~0.892)	0.804	0.744	0.865	0.802	<0.001
Inception-v3	US	0.861(0.836~0.886)	0.809	0.810	0.813	0.812	<0.001
DenseNet-121	US	0.853(0.825~0.881)	0.789	0.805	0.771	0.802	<0.001
ResNext-50	US	0.844(0.817~0.871)	0.757	0.710	0.876	0.784	<0.001
Swin Transformer	US	0.836(0.806~0.866)	0.772	0.722	0.846	0.778	<0.001
ResNet-50	US	0.833(0.807~0.859)	0.771	0.761	0.783	0.772	<0.001
ConvNeXt	US	0.821(0.792~0.850)	0.783	0.715	0.820	0.762	<0.001
Vision Transformer	US	0.812(0.780~0.844)	0.779	0.702	0.823	0.758	<0.001

2.3 消融实验

本研究设计了一系列的消融实验,系统性地评估分割模块、跨模态生成模块以及交叉注意力融合模块的必要性与有效性。所有消融实验均以Inception-v3为骨干网络,实验结果如表2所示。

分割模块的贡献通过两组对比得以验证。首先,在单模态场景下,对比无分割的基线模型(g组)(AUC=0.843),经过分割的基线模型(f组)(AUC=0.861),

性能提升1.8%。在多模态场景中,对比无分割的完整框架(c组)(AUC=0.893),最终的完整框架(a组)(AUC=0.930)性能提升3.7%。

生成模块的效用通过f、a、b组性能递进关系得到有力证明。首先,在无跨模态生成模型的情况下,对比单模态基线模型(f组)(AUC=0.861),使用真实双模态融合模块的(b组)分类性能取得明显提升(AUC=0.946),证明引入真实的CEUS模态能够提供

表 2 消融实验结果
Table 2 Ablation study results

实验组	分割	US	gCEUS	真实 CEUS	融合机制	AUC(95% CI)	准确率	敏感度	特异度	F1 分数	P 值
a	√	√	√		交叉注意力	0.930(0.913~0.947)	0.854	0.831	0.887	0.864	-
b	√	√		√	交叉注意力	0.946(0.930~0.962)	0.869	0.896	0.836	0.880	<0.010
c		√	√		交叉注意力	0.893(0.871~0.915)	0.801	0.787	0.806	0.796	<0.001
d	√	√		√	拼接融合	0.901(0.880~0.922)	0.811	0.815	0.805	0.823	<0.001
e	√	√	√		拼接融合	0.882(0.857~0.907)	0.807	0.774	0.833	0.807	<0.001
f	√	√			单分支	0.861(0.836~0.886)	0.809	0.810	0.813	0.812	<0.001
g		√			单分支	0.843(0.814~0.872)	0.775	0.771	0.789	0.782	<0.001

高价值的诊断信息。在此基础上,通过引入跨模态生成模块,本研究的完整框架(a组)(AUC=0.930)的分类精度同样明显高于单模态基线(f组),并逼近理论性能模型(b组)。

同时,本研究设计了两组对照实验,验证交叉注意力特征融合相较于传统静态融合的优越性。第一组,在处理生成数据时,相较于采用拼接融合的配置(e组)(AUC=0.882, 95% CI: 0.857~0.907),采用交叉注意力的完整框架(a组)(AUC=0.930, 95% CI:

0.913~0.947)性能提升明显,且差异具有统计学意义($P<0.001$)。第二组,在处理真实数据时,采用交叉注意力的理论性能模型(b组)(AUC=0.946, 95% CI: 0.930~0.962)同样明显优于采用拼接融合的配置(d组)(AUC=0.901, 95% CI: 0.880~0.922)。

2.4 生成图像质量的定性评估

为了直观评估跨模态生成网络的性能,图4展示了良性与恶性肿瘤在B超、生成式造影及真实造影3种模态下的可视化对比结果。

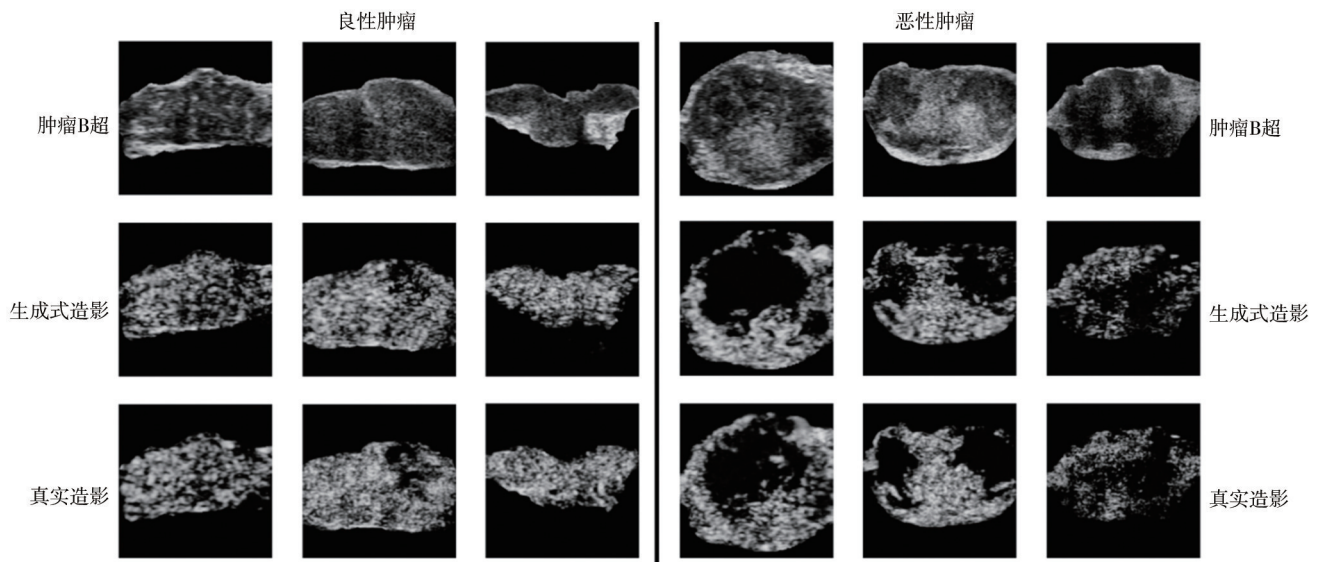


图4 良性和恶性肿瘤的模型生成表现
Figure 4 Generation performance for benign and malignant tumors

从形态学角度观察,生成造影图像在肿瘤的几何形状、尺寸大小及边缘轮廓等结构方面,与输入的原始US图像保持了高度的一致性。这表明生成网络有效地利用编码器提取的结构特征,未出现明显的几何畸变或伪影,能够精确定位病灶区域。从纹理和增强模式角度观察,生成造影图像在视觉上成功复现了真实造

影图像的关键诊断特征。对于恶性肿瘤病例(图4右侧),模型生成典型的不均匀高增强模式,清晰地模拟出肿瘤内部因坏死或血供不足导致的无灌注区,这与真实造影的表现高度吻合。对于良性病例(图4左侧),模型生成了较为均匀的增强纹理,反映了良性肿瘤血供分布相对规则的特点。

3 讨论

本研究提出利用计算的方法生成并融合多模态超声信息以实现高精度诊断的设想,成功构建一个“分割-生成-分类”三阶段级联诊断框架。实验结果验证该框架的可行性和有效性,更证明跨模态生成模块的核心效用以及多模态信息融合的诊断价值。

首先,本研究最重要的发现是,联合使用常规US图像与本框架生成的超声造影(gCEUS)图像进行诊断,其性能(AUC=0.930)显著优于所有仅依赖单模态US的先进基线模型(最佳AUC=0.868)。这一显著的性能提升证实了跨模态生成模型并非简单地对输入图像进行纹理重映射,而是成功学习到了与肿瘤血流灌注相关的关键功能性特征。

在对图4的定性分析中,对于恶性肿瘤,模型生成的gCEUS图像准确呈现了典型恶性征象,对于良性肿瘤则生成了较为均匀的增强模式。这种差异表明模型已具备捕捉良恶性肿瘤在微血管灌注动力学上本质区别的能力。尽管生成图像在像素级上与真实造影图像仍存在细微差异,但其提供的功能性特征弥补B超仅有结构信息的短板,在特征空间中构成对真实CEUS的有效替代。从信息论角度看,本方法成功缩小单模态基线与使用真实双模态数据的理论性能上限(AUC=0.946)之间的差距,实现了“以计算代检查”的临床价值。

针对GAN在跨模态医学图像生成中普遍存在的临床幻觉风险,本研究在方法设计与实验验证上形成了一定的闭环论证。如果图像生成模型过度追求视觉的逼真度,容易产生凭空合成的病理特征,破坏医学影像的物理一致性。为规避这一风险,本框架在跨模态生成模块引入严格的肿瘤结构掩码与多尺度损失函数约束,强制模型在B超解剖结构的先验引导下进行特征映射。在生成结果的物理一致性方面,定性分析表明,对于恶性肿瘤,模型生成的gCEUS图像不仅维持了边缘的空间一致性,更准确再现肿瘤中心因细胞坏死或血管畸形导致的充盈缺损与不均匀高增强模式;对于良性肿瘤,模型则对应生成符合其规则血管特征的均匀增强模式。这种基于B超回声异质性生成符合血流动力学规律的灌注特征能力,证实生成图像的合理性。更关键的是,本框架在下游分类任务中取得0.930的AUC值,说明生成模型不是凭空生成幻觉特征,而是成功捕获并重建具有真实鉴别诊断价值的物理映射关系,保障诊断结果的可靠性。

其次,交叉注意力机制在本框架的分类模块中

起到了实现信息增益的关键作用。消融实验数据表明,无论是使用真实的还是生成的超声数据,采用交叉注意力的双模态分类模型性能都明显优于采用简单静态融合的模式。这种实验结果背后的机理是,交叉注意力通过其动态的、内容自适应的查询与关注过程,实现两种模态在特征层面的深度、智能融合。一方面,US模态提供了稳健的结构与边界信息,其特征图作为“查询”,能够自适应地关注gCEUS特征图中位于肿瘤结构内部的血流信号,从而利用形态学信息为功能性信息提供空间锚定,抑制无关信息的干扰;另一方面,生成的gCEUS模态提供功能性的血供强度信息,其特征图中的高亮区域作为“查询”,可以反过来引导模型更加关注US特征图中那些形态学上可能模糊、但在功能上高度可疑的区域。这种双向的、动态的特征交互过程,使得最终得出的融合特征能够兼顾结构与功能的协同信息,获得比静态拼接更丰富、更具判别力的复合特征。

本研究框架聚焦于良恶性二分类任务,但其通过生成式方法增强图像信息并进行多模态深度融合的核心思想,为解决更具临床挑战性的乳腺癌分子亚型无创预测提供了重要技术路径。乳腺癌的分子亚型是决定患者预后和治疗方案的关键依据,其金标准是侵入性的组织活检和免疫组化分析^[28]。现有研究表明,不同的分子亚型在影像学上常表现出结构形态与血管模式复杂相互作用的显著成像表型^[29-31]。本框架通过独立的US分支提取形态特征、gCEUS分支提取血流特征,并通过交叉注意力进行深度耦合,在理论上非常契合分子亚型预测对多维特征的需求。这为未来通过非侵入性影像手段辅助制定个性化治疗方案提供了重要的技术路径。

尽管本研究已取得积极成果,但依然存在一定的局限性。首先,数据来源的单一性限制了模型的泛化性验证。本研究的数据集来源于单一医疗中心,模型的泛化能力仍有待在更多中心的数据上进行外部验证。其次,生成图像与真实数据间仍存在部分差距。虽然生成的造影图像在宏观诊断特征上与真实图像大致一致,但在高频纹理细节上仍存在一定差距,这反映当前生成模型在完美复现高频纹理和复杂血流灌注细节方面的挑战。最后,级联框架存在潜在的误差传导风险。本研究采用的“分割-生成-分类”三阶段级联框架,虽然逻辑清晰且模块化,但其结构本身带来误差在框架中累积和传递的风险。未来工作将探索端到端的联合优化策略,以降低误差传播风险。

4 结论

针对常规US因判读一致性低、微小病灶征象不明显导致的诊断瓶颈,以及超声造影虽能提供功能信息、却因可及性受限与诊断征象存在模糊性而难以普及的临床矛盾,本研究提出一种基于跨模态生成与交叉注意力融合的智能诊断新方法,该方法通过“分割-生成-分类”三阶段级联设计,成功地从常规B超图像中挖掘并重建具有关键诊断价值的生成造影信息,实现结构与功能信息的深度融合,最终完成对肿瘤的高精度良恶性诊断。

在一个包含677例患者的临床数据集上严格验证表明,该方法在分类性能上(AUC=0.930)显著优于各类仅依赖单模态信息的先进基线模型,逼近真实双模态检查的水平。本研究有力地证实了“以计算代替检查”范式的可行性,即通过深度学习技术挖掘常规影像中的隐式特征,可以低成本地获取高级影像的诊断效能。这一成果不仅为乳腺癌筛查提供了一种高效、经济、无创的技术手段,也为医学影像分析领域解决多模态数据缺失问题、实现乳腺癌分子亚型的无创预测,提供了新的理论视角与技术路径,具有广阔的临床应用前景。

【参考文献】

- [1] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2024, 74(3): 229-263.
- [2] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249.
- [3] Hookey RJ, Scoutt LM, Philpotts LE. Breast ultrasonography: state of the art[J]. Radiology, 2013, 268(3): 642-659.
- [4] Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer[J]. JAMA, 2008, 299(18): 2151-2163.
- [5] Abdullah N, Mesurolle B, El-Khoury M, et al. Breast imaging reporting and data system lexicon for US: interobserver agreement for assessment of breast masses[J]. Radiology, 2009, 252(3): 665-672.
- [6] Romeo V, Cuocolo R, Apolito R, et al. Clinical value of radiomics and machine learning in breast ultrasound: a multicenter study for differential diagnosis of benign and malignant lesions[J]. Eur Radiol, 2021, 31(12): 9511-9519.
- [7] Chen C, Turco S, Kapetas P, et al. Spatiotemporal analysis of contrast-enhanced ultrasound for differentiating between malignant and benign breast lesions[J]. Eur Radiol, 2024, 34(7): 4764-4773.
- [8] Li Q, Hu M, Chen ZK, et al. Meta-analysis: contrast-enhanced ultrasound versus conventional ultrasound for differentiation of benign and malignant breast lesions[J]. Ultrasound Med Biol, 2018, 44(5): 919-929.
- [9] Cheng M, Tong W, Luo J, et al. Value of contrast-enhanced ultrasound in the diagnosis of breast US-BI-RADS 3 and 4 lesions with calcifications[J]. Clin Radiol, 2020, 75(12): 934-941.
- [10] Wan CF, Du J, Fang H, et al. Evaluation of breast lesions by contrast enhanced ultrasound: qualitative and quantitative analysis[J]. Eur J Radiol, 2012, 81(4): e444-e450.
- [11] Sidhu PS, Cantisani V, Dietrich CF, et al. The EFSUMB guidelines and recommendations for the clinical practice of contrast-enhanced ultrasound (CEUS) in non-hepatic applications: update 2017 (long version)[J]. Ultraschall Med, 2018, 39(2): e2-e44.
- [12] Wang J, Zhao R, Cheng J. Diagnostic accuracy of contrast-enhanced ultrasound to differentiate benign and malignant breast lesions: a systematic review and meta-analysis[J]. Eur J Radiol, 2022, 149: 110219.
- [13] 孙雨, 杨琛. 基于自动乳腺超声诊断系统的乳腺肿瘤人工智能诊断研究进展[J]. 中国医学影像学杂志, 2024, 32(11): 1176-1181.
- [14] Sun Y, Yang C. Research progress on artificial intelligence diagnosis for breast tumors based on automated breast ultrasound system[J]. Chinese Journal of Medical Imaging, 2024, 32(11): 1176-1181.
- [15] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
- [16] 王一凡, 刘静, 马金刚, 等. 深度学习在乳腺癌影像学检查中的应用进展[J]. 计算机科学与探索, 2024, 18(2): 301-319.
- [17] Wang YF, Liu J, Ma JG, et al. Application progress of deep learning in imaging examination of breast cancer[J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(2): 301-319.
- [18] Cao H, Wang YY, Chen J, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation[C]//Computer Vision-ECCV 2022 Workshops. Cham: Springer Nature Switzerland, 2023: 205-218.
- [19] Chen JN, Lu YY, Yu QH, et al. TransUNet: transformers make strong encoders for medical image segmentation[EB/OL]. (2021-02-08). <https://arxiv.org/abs/2102.04306>.
- [20] 柴梦婷, 朱远平. 生成式对抗网络研究与应用进展[J]. 计算机工程, 2019, 45(9): 222-234.
- [21] Chai MT, Zhu YP. Research and application progress of generative adversarial networks[J]. Computer Engineering, 2019, 45(9): 222-234.
- [22] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Commun ACM, 2020, 63(11): 139-144.
- [23] Isola P, Zhu JY, Zhou TH, et al. Image-to-image translation with conditional adversarial networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 5967-5976.
- [24] Nie D, Trullo R, Lian J, et al. Medical image synthesis with context-aware generative adversarial networks[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2017. Cham: Springer International Publishing, 2017: 417-425.
- [25] Chen C, Yu JB, Xu ZK, et al. Pseudo-contrast-enhanced US via enhanced generative adversarial networks for evaluating tumor ablation efficacy[J]. Radiol Artif Intell, 2025, 7(3): e240370.
- [26] Wang QC, Chen H, Luo GN, et al. Performance of novel deep learning network with the incorporation of the automatic segmentation network for diagnosis of breast cancer in automated breast ultrasound[J]. Eur Radiol, 2022, 32(10): 7163-7172.
- [27] Zhou HY, Yu YZ, Wang CD, et al. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics[J]. Nat Biomed Eng, 2023, 7(6): 743-755.
- [28] Chen C, Wang Y, Niu JW, et al. Domain knowledge powered deep learning for breast cancer diagnosis based on contrast-enhanced ultrasound videos[J]. IEEE Trans Med Imaging, 2021, 40(9): 2439-2451.
- [29] 何俊, 张彩庆, 李小珍, 等. 面向深度学习的多模态融合技术研究综述[J]. 计算机工程, 2020, 46(5): 1-11.
- [30] He J, Zhang CQ, Li XZ, et al. Survey of research on multimodal fusion technology for deep learning[J]. Computer Engineering, 2020, 46(5): 1-11.
- [31] Chen CFR, Fan QF, Panda R. CrossViT: cross-attention multi-scale vision transformer for image classification[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 347-356.
- [32] Ab Mumin N, Ramli Hamid MT, Wong JHD, et al. Magnetic resonance imaging phenotypes of breast cancer molecular subtypes: a systematic review[J]. Acad Radiol, 2022, 29 Suppl 1: S89-S106.
- [33] Zhang K, Li JW, Zhu Q, et al. Prediction of pathologic complete response by ultrasonography and magnetic resonance imaging after neoadjuvant chemotherapy in patients with breast cancer[J]. Cancer Manag Res, 2020, 12: 2603-2612.
- [34] Gong XT, Li QF, Gu LS, et al. Conventional ultrasound and contrast-enhanced ultrasound radiomics in breast cancer and molecular subtype diagnosis[J]. Front Oncol, 2023, 13: 1158736.
- [35] Ko ES, Lee BH, Kim HA, et al. Triple-negative breast cancer: correlation between imaging and pathological findings[J]. Eur Radiol, 2010, 20(5): 1111-1117.

(编辑:薛泽玲)