

基于多模态超声对比学习的肝癌诊断方法

杨印凯¹, 万鹏¹, 石航¹, 薛海燕², 邵伟¹

(1. 南京航空航天大学人工智能学院脑机智能技术教育部重点实验室, 南京 211106; 2. 南京大学医学院附属鼓楼医院, 南京 210008)

摘要: 近年来, 肝癌已经成为严重威胁人类健康的疾病, 多模态超声成像是其诊断的重要手段之一。与临床医生结合多模态超声诊断肝癌类似, 利用多模态融合的方法集成各模态超声的影像特征有望提高肝癌诊断的准确性。然而, 现有的多模态融合方法在融合过程中往往将各模态的特征信息孤立处理, 未能充分考虑模态内的样本相似性和模态间的语义一致性, 同时忽视了模态的不确定性。因此, 提出了一种基于多模态超声对比学习的肝癌诊断方法, 旨在充分利用各超声模态的特征信息, 提高诊断准确率。该方法利用监督对比学习深入挖掘模态特征, 捕获模态内同类样本之间的相似性信息和不同模态之间样本的语义一致性信息。此外, 该方法基于主观逻辑引入了模态不确定度的度量, 实现了模态信息的动态融合, 具有较好的鲁棒性。多模态超声影像评估结果显示, 本文提出的方法实现了 85.21% 诊断准确率, 相较于主流的多模态融合方法性能得到了提升。

关键词: 多模态融合; 超声; 对比学习; 不确定度; 肝癌诊断

中图分类号: TP18; R445

文献标志码: A

Liver Cancer Diagnosis Method Based on Multi-modal Ultrasound Contrast Learning

YANG Yinkai¹, WAN Peng¹, SHI Hang¹, XUE Haiyan², SHAO Wei¹

(1. Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; 2. Department of Ultrasound, Affiliated Drum Tower Hospital, Medical School of Nanjing University, Nanjing 210008, China)

Abstract: In recent years, liver cancer has become a disease that seriously threatens human health, and multi-modal ultrasound imaging is one of the important diagnostic tools for it. Similar to how clinicians use multi-modal ultrasound to diagnose liver cancer, using multi-modal fusion methods to integrate the image features of each ultrasound modality is expected to improve the accuracy of liver cancer diagnosis. However, the existing multi-modal fusion methods often isolate the feature information of each modality during the fusion process, failing to fully consider the intra-modal sample similarity and inter-modal semantic consistency, while ignoring modality uncertainty. Therefore, this paper proposes a liver cancer diagnosis method based on multi-modal ultrasound contrast learning, aiming to make full use of the feature information of each ultrasound modality to improve the diagnostic accuracy. Specifically, this method employs supervised contrastive learning to deeply explore modality features, capturing both the similarity information among samples within the modality and the semantic consistency information across different modalities. In addition, this method introduces a measure of modality uncertainty based on Subjective

Logic, enabling dynamic fusion of modality information and exhibiting good robustness. Evaluation of multimodal ultrasound imaging shows that the proposed method achieves an 85.21% diagnostic accuracy, demonstrating performance improvement compared to other mainstream multimodal fusion methods.

Key words: multi-modal fusion; ultrasound; contrast learning; uncertainty; liver cancer diagnosis

引 言

原发性肝癌是世界上最常见的癌症之一,是癌症相关死亡的第三大常见原因,中国已成为原发性肝癌发病率最高的国家^[1-2]。然而,只有不超过20%的患者在初次确诊为肝癌时,拥有根治性治疗的机会。原发性肝癌的主要类型包括肝细胞癌(Hepatocellular carcinoma, HCC)和肝内胆管癌(Intrahepatic cholangiocarcinoma, ICC)^[3-4],这两种类型因其独特的发病机制、病理特征、预后和对辅助治疗的反应而被视为不同的肝癌类型^[5]。因此,HCC和ICC的早期鉴别诊断对于肝癌的临床诊疗至关重要。

超声具有安全无创、操作简单、经济便捷的特点,已成为原发性肝癌诊断常用的重要手段之一。其中,B模式超声(BUS)提供了肝脏的形态学信息,超声造影(CEUS)提供了肝脏的血流灌注信息,在此基础上,超声参量成像(PARA)提供不同组织血流初始到达时间,反映了肝组织异质信息,不同模态超声从不同层面反映了肝病灶相关信息。因此,集成不同模态超声的影像特征有望提高肝癌诊断的准确性。

多模态融合的早期尝试^[6]有特征拼接,是在提取特征后立即对其进行集成,这种做法比较简单但未能充分发掘模态信息之间的相关性。基于典型相关分析(Canonical correlation analysis, CCA)^[7]的多模态融合方法^[8-9]为每个模态单独建立网络,可以很好地捕获模态之间的相关性^[10],但仍然受到模态数量的限制。不同于特征层面的融合,决策融合在每个模态做出预测后进行集成,常见的方法有决策平均^[11]、决策投票^[12]以及基于注意力的决策融合^[13]。然而,这些多模态融合方法大都默认各个模态的权重相等或固定,因此难以适应复杂的现实世界。

随着深度学习的发展,多模态融合的研究更加深入。Xu等^[14]提出了一种自适应加权损失融合策略用于多个神经网络进行联合决策,通过灵活调整权重的幂指数来自适应地集成多个神经网络,避免了权重的平凡求解。Han等^[15]提出了动态多模态融合策略对特征和模态信息进行建模,利用稀疏门策略动态获取不同样本的特征信息,从而可靠地集成多种模态。虽然这些方法在多模态融合中取得了一定的进展,但仍存在两个关键问题:(1)无法量化各模态的不确定度,不能给出决策的置信度;(2)孤立处理各模态信息,忽略了模态内的样本相似性和模态间的语义一致性。

Khosla等^[16]将对比学习引入到监督领域中,同类样本“拉近”,异类样本“拉远”,使模型能够有效地利用同类样本的相似性。Jin等^[17]基于对比学习引入了样本对齐的概念,学习到跨视图的一致性信息。Han等^[18-19]在多视图融合时为每一个视图引入了不确定度的度量,并且在分类任务中取得很好的效果。鉴于此,本文提出一种新的多模态融合方法,一方面采用监督对比学习捕获模态内同类样本之间的相似性信息和不同模态之间样本的语义一致性信息;另一方面在主观逻辑下建立超声模态的不确定度量并实现动态融合。在多模态超声影像数据上的实验证明所提方法的有效性和可行性。

本文的主要贡献如下:

(1)提出一种基于多模态超声对比学习的肝癌诊断方法,捕获模态内同类样本之间的相似性信息和不同模态之间样本的语义一致性信息。

(2)建立超声模态不确定度的度量,利用证据理论实现多模态超声的动态融合,提供可靠诊断结果。

1 相关工作

1.1 多视图学习

近年来,各种研究工作^[14-15,18-23]专注于设计可以有效融合多模态数据的新方法。Xu等^[14]提出了一种自适应加权损失融合策略用于多个神经网络进行联合决策,通过灵活调整权重的幂指数来自适应地集成多个神经网络,避免了权重的平凡求解。Han等^[15]提出了动态多模态融合策略,利用稀疏门策略动态获取不同样本的信息特征,引入模态置信度来动态评估不同样本的不同模态的信息量,从而可靠地集成多种模态。Han等^[18]提出了可信多视图分类(Trusted multi-view classification, TMC)框架,巧妙地利用主观逻辑为每一个视图引入不确定度的度量,还利用 Dempster-Shafer 理论^[24]融合不同来源的证据(反映了模态中用于决策的分类概率),实现了动态多视图融合,而且还为融合结果给出不确定度。为了解决 TMC 框架缺少视图交互的问题,Han等^[19]在原有的基础上将所有视图的原始特征拼接起来,从而引入伪视图,一定程度上缓解了 TMC 缺少视图交互的问题。Yang等^[20]提出了面向发散的多模态融合网络(Divergence-orientated multi-modal fusion network, DOMFN),可以基于跨模态发散自适应地融合单模态预测和多模态预测,当跨模态发散过大时,DOMFN 会拒绝多模态预测,避免整体性能下降。Liu等^[21]从意见聚合的角度出发,采用证据理论将意见的不确定性表示为来自不同视图的学习结果,并通过证据积累将意见聚合的不确定性作为多视图融合的结果。Xue等^[22]提出了一种新颖的多模态融合方法,该方法基于一种门函数可以自适应融合多模态数据并在推理过程自适应融合输入,可以即时提供模态级或融合级决策。Zhang等^[23]提出了一种针对低质量多模态数据的动态融合方法,引入了不确定性感知权重在决策层进行不确定加权融合。

1.2 对比学习

近年来,对比学习相关的研究工作^[16-17,25-31]也取得很大的进展。Khosla等^[16]将自监督领域的对比方法创新地扩展到监督领域,将同一类的样本在嵌入空间中拉在一起,同时将来自不同类的样本分开,使模型能够有效地利用类别信息。Jin等^[17]提出新的多视图对比学习,将同一样本在不同视图下的表示作为正样本,将其他样本的视图作为负样本,学习到跨视图的一致性信息。Zhong等^[25]提出了一种新颖的基于图的对比学习策略,结合潜在的类别信息来学习更紧凑的聚类分配,将传统的实例级一致性提升为集群级一致性。You等^[26]提出了一种用于 GNN 预训练的新型图对比学习框架(Graph contrastive learning, GraphCL),通过潜在空间中的对比损失最大化同一样本的两个增强视图之间的一致性,以便针对不同的图结构数据学习对扰动不变的表示;Xiao等^[27]提出了一种图对比学习方法,该方法可以在简单的非对称学习框架中捕获一跳邻域上下文和两跳单一相似性。Li等^[28]提出了一种称为对比聚类(Contrastive clustering, CC)的在线聚类方法,分别在行和列空间中进行实例级和簇级对比学习,可以很好地获得聚类最喜欢的表示。Hager等^[29]将影像和表格两个模态的数据融入到对比学习框架,为无监督多模态预训练和单模态预测提供了一个新的思路。Shan等^[30]将对比学习应用到预训练阶段提出了对比预训练框架,该框架使预训练模型能够从未标记的数据中学习特征表示。Sung等^[31]提出了一种基于对比学习的语义分割方法,利用多尺度上下文对比学习使网络能够捕获局部及全局上下文信息,以获得更好的区分能力。

2 本文方法

如图 1 所示,本文网络由 3 部分组成,分别是实例感知模块、不确定度获取模块和多模态融合模块。实例感知模块由两重对比损失组成,负责约束网络结构捕获模态内的样本相似性信息和模态间的语义一致性信息;不确定度获取模块包含证据网络,主要负责利用证据网络的输出并基于主观逻辑^[32]获取各个模态的不确定度;多模态融合模块由证据组合规则^[33]组成,负责融合实现多模态融合。

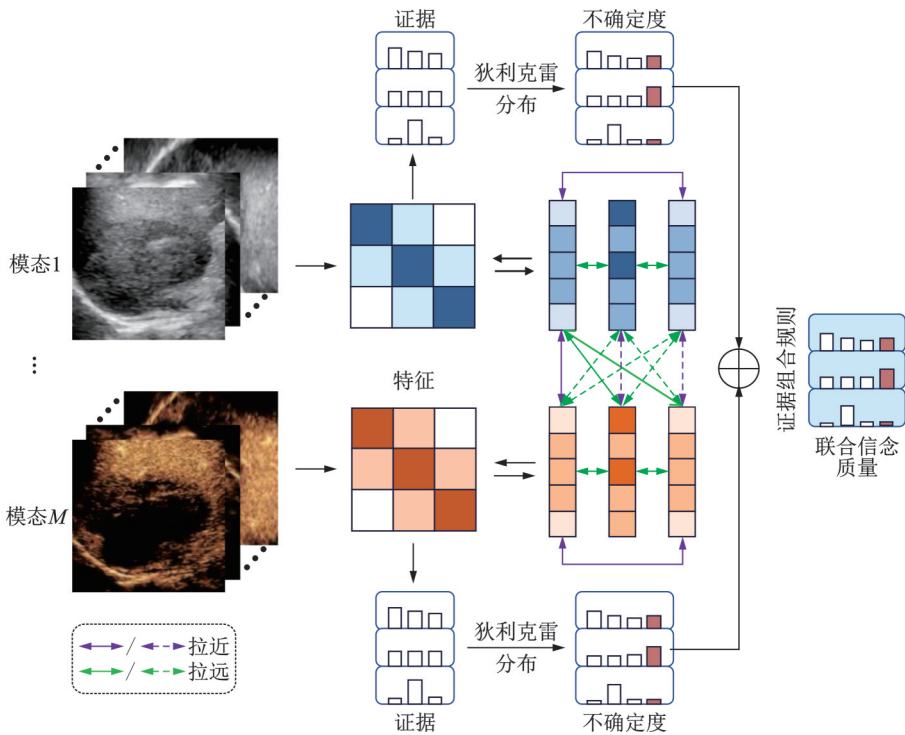


图1 本文方法框架

Fig.1 Framework of the proposed method

2.1 基本定义

定义一个包含 M 个模态和 N 个样本的 K 分类任务。在模态 m 中,样本 i 经过神经网络提取的特征信息记为 f_i^m 。该特征信息在投影空间的投影向量记为 p_i^m ,通过证据网络的输出记作 $e_i^m = \{e_{i,1}^m, e_{i,2}^m, \dots, e_{i,K}^m\}$ 。

2.2 实例感知模块

对比学习的思想是在嵌入空间中 将锚样本和“正”样本拉在一起,并将锚点与许多“负”样本分开^[16]。基于对比学习的思想,本文考虑将同一个超声模态内(如BUS)的实例按照类别区分,让模型可以捕获同类实例之间的相似性信息,以便在决策时可以获得更好的结果。利用式(1,2)可以获得模态内的实例感知损失。

$$\ell_i^m = \frac{-1}{|P(i)|} \sum_{q \in P(i)} \log \frac{\exp(p_i^m \cdot p_q^m / \tau)}{\sum_{j \in A(i)} \exp(p_i^m \cdot p_j^m / \tau)} \quad (1)$$

$$L_{\text{intra}} = \sum_{m=1}^M \sum_{i=1}^N \ell_i^m \quad (2)$$

式中: i 表示模态 m 中的样本, $P(i)$ 和 $A(i)$ 分别为模态中与 i 有相同标签和不同标签的样本集合, $|P(i)|$ 为集合 $P(i)$ 的大小, ℓ_i^m 表示模态 m 中样本 i 的实例感知损失。

在多模态超声中,同一个实例的各模态信息都能对肝癌进行初步诊断,即超声模态之间存在语义一致性,在高维特征空间中联系比较密切。同样地,受样本对齐^[17]和对比学习的启发,本文将同一实例

在某个超声模态下的数据与其他超声模态下的数据进行交互,这种处理方式让同一实例的不同超声模态之间更为接近,而与其他实例的不同模态之间的距离则更远,引入了模态间的语义一致性。利用式(3,4)可以获得模态间的实例感知损失。

$$\delta_i^m = - \sum_{r=1, r \neq m}^M \log \frac{\exp(p_i^m \cdot p_i^r / \tau)}{\sum_{j=1, j \neq i}^N \exp(p_i^m \cdot p_j^r / \tau)} \quad (3)$$

$$L_{\text{inter}} = \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^N \delta_i^m \quad (4)$$

式中: j 和 r 分别为样本和模态, δ_i^m 表示模态 m 中样本 i 的模态间实例感知损失。

2.3 不确定性度量模块

传统的神经网络在获得输出之前大多需要 softmax 操作,这种做法仅获得了分类的概率^[21],但无法量化分类结果的不确定度。为了引入不确定度量,本文利用激活函数(如 Softplus)替换 softmax 层,这样网络最终可以得到非负输出,也被称为证据^[34]。

主观逻辑(Subjective logic, SL)^[32]定义了一个理论框架,允许通过证据来量化信念质量和不确定质量。具体而言,给定模态 m 中样本 i 的证据 $e_i^m = \{e_{i,1}^m, e_{i,2}^m, \dots, e_{i,K}^m\}$,主观逻辑理论可以为每一个类别 k 提供信念质量 $b_{i,k}^m$,还会提供一个总体不确定度(也称不确定质量) u_i^m ,信念质量和不确定质量满足

$$\sum_{k=1}^K b_{i,k}^m + u_i^m = 1 \quad (5)$$

式中:不确定度 $u_i^m \geq 0$,信念质量 $b_{i,k}^m \geq 0$ 。

对于模态 m ,主观逻辑将样本 i 的证据 $e_i^m = \{e_{i,1}^m, e_{i,2}^m, \dots, e_{i,K}^m\}$ 与狄利克雷分布 $D(p_i^m | \alpha_i^m)$ 的参数 $\alpha_i^m = \{\alpha_{i,1}^m, \alpha_{i,2}^m, \dots, \alpha_{i,K}^m\}$ 联系起来,简单来讲,两者之间满足: $e_{i,1}^m = \alpha_{i,1}^m - 1$ 。然后,信念质量 $b_{i,k}^m$ 和不确定度 u_i^m 可以通过式(6)计算得到。

$$b_{i,k}^m = \frac{e_{i,k}^m}{S_i^m}, u_i^m = \frac{K}{S_i^m} \quad (6)$$

式中: $S_i^m = \sum_{k=1}^K (e_{i,k}^m + 1) = \sum_{k=1}^K \alpha_{i,k}^m$,被称为狄利克雷强度。

由狄利克雷分布导出的分类概率 $\hat{p}_{i,k}^m = \frac{\alpha_{i,k}^m}{S_i^m} = \frac{e_{i,k}^m + 1}{S_i^m}$,不难看出,当样本 i 观察到第 k 类的证据越多,最终预测为第 k 个类别的概率越高;当样本 i 观察到分类的证据越少,模态 m 预测的不确定度 u_i^m 越高。

2.4 多模态融合模块

之前的模块已经可以获得模态 m 中样本 i 的证据和不确定度 u_i^m ,现在考虑如何实现多个模态信息的融合。Dempster-Shafer 证据理论允许将来自不同模态的证据组合在一起,从而得到模态综合考虑后的分类结果。该理论将 M 个模态的信念质量和不确定度 U_i^m 结合起来,获得联合信念质量 $U_i = \left\{ \{b_{i,k}\}_{k=1}^K, u_i \right\}$,其中 $U_i^m = \left\{ \{b_{i,k}^m\}_{k=1}^K, u_i^m \right\}$ 。两个模态的组合规则如式(7,8)所示。

$$U_i = U_i^1 \oplus U_i^2 \quad (7)$$

$$b_{i,k} = \frac{1}{1-C} (b_{i,k}^1 b_{i,k}^2 + b_{i,k}^1 u_i^2 + b_{i,k}^2 u_i^1), u_i = \frac{1}{1-C} u_i^1 u_i^2 \quad (8)$$

式中: \oplus 表明组合规则,具体过程如式(8)所示, $C = \sum_{s \neq t} b_{i,s}^1 b_{i,t}^2$ 一定程度上反映了类别之间的冲突, $1-C$ 用来对融合以后的结果做归一化。

直观来看,该组合规则具有以下特点:(1)当两个模态都有较高的不确定度时,最终的分分类结果会有较低的置信度;(2)当两个模态都有低不确定度时,最终的分分类结果会有较低不确定度;(3)当有且仅有一个模态有低不确定度时,最终的分分类结果会依靠该模态。

当然,该组合规则也可以应用到更多模态。给定 M 个模态的信念质量和不确定度,可以连续使用上述组合规则获得最终的联合信念质量 $U_i = \left\{ \{b_{i,k}\}_{k=1}^K, u_i \right\}$,即 $U_i = U_i^1 \oplus U_i^2 \oplus \dots \oplus U_i^M$ 。

在获得多模态融合以后的联合信念质量 U_i 后,根据式(6),多模态融合以后的联合证据 e_i 和相应的狄利克雷分布的参数 α_i 可以通过式(9)计算得到。然后,可以获得最终用于分类决策的概率 $\hat{p}_{i,k} =$

$$\frac{\alpha_{i,k}}{S_i} = \frac{e_{i,k} + 1}{S_i} \quad (9)$$

$$e_i = b_{i,k} \cdot S_i, \alpha_i = e_i + 1$$

式中: $S_i = \frac{K}{u_i}$ 为狄利克雷分布强度。

2.5 总损失

训练过程中模型的损失目标为

$$L_{\text{train}} = L_{\text{intra}} + L_{\text{inter}} + L_{\text{ce}} \quad (10)$$

式中模态内的实例感知损失 L_{intra} 和模态间的实例感知损失 L_{inter} 分别由式(2)和式(4)给出。实例感知模块是约束模型捕获到模态内的样本相似性信息和模态间的语义一致性信息,仅存在于训练期间,在验证和测试时只有 L_{ce} 损失。具体来讲, L_{ce} 损失由式(11,12)计算得到。

$$L_{(\alpha_i)} = \mathbb{E}_{p_i \sim \text{Dirh}(p_i|\alpha_i)} \left[\sum_{k=1}^K -y_{i,k} \log(p_{i,k}) \right] = \sum_{k=1}^K y_k (\psi(S_i) - \psi(\alpha_{i,k})) \quad (11)$$

$$L_{\text{ce}} = \sum_{i=1}^N \left(\sum_{m=1}^M L_{(\alpha_i^m)} + L_{(\alpha_i)} \right) \quad (12)$$

式中: ψ 表示双伽马函数, y 表示分类的标签(0-1表示形式)。

3 实验验证

3.1 数据集和实验设置

实验中使用了来自不同设备的多模态超声影像数据,包括超声数据(BUS)、造影增强数据(CEUS)和参量成像数据(PARA),这些数据共涵盖了238例病例,其中136例为肝细胞癌(HCC),102例为胆管癌(ICC)。在造影增强模态数据中,本文选择了造影灌注达到峰值的时刻^[35]作为CEUS模态的肝癌诊断依据,并且利用分割结果分别从这3种模态的影像中提取特征向量^[36]作为神经网络的输入。所有的实验均在Pytorch环境下完成,并在Nvidia RTX3060 GPU上进行训练。

为更好地评估本文提出的方法,实验中选择了近几年取得较优效果的方法进行比较,包含了自适应多模态融合、可信多模态融合以及动态多模态融合:

(1)MvNNcor^[14],自适应加权不同模态的表示进行联合决策;

- (2)TMC^[18],基于证据理论使用Dempster-Shafer理论集成不同模态的证据;
- (3)TMDLO^[21],基于证据理论通过证据积累聚合各个模态的不确定性进行融合;
- (4)ETMC^[19],添加了一个伪模态,是对TMC的改进和优化;
- (5)MMDynamics^[15],利用特征级和模态级动态性进行可信多模态融合。

实验中基于多模态超声的HCC和ICC分类是一个二分类问题,而ROC曲线下与坐标轴围成的面积(Area under curve, AUC)是二分类中一个重要和广泛使用的指标,因此本文选择AUC作为评估的一个重要标准。为了更好地评估模型的有效性,在实验中还选择了ACC(准确率)、Macro- F_1 (宏平均 F_1 -score)和Weighted- F_1 (加权平均 F_1 -score)作为评估标准。

3.2 实验结果

3.2.1 方法有效性

将多模态融合和单个模态的预测误差进行对比,结果如图2所示。从图2可以观察到,多模态融合的预测误差折线几乎始终低于任何单独模态,这表明多模态融合降低了分类误差,说明提出的多模态方法是有效的。当然,可以从折线图中看出不同模态的超声对于HCC和ICC分类任务存在一定差异,整体来看,BUS和CEUS的分类效果优于PARA的分类效果。

除此以外,本文还对比了单独模态以及任意模态融合的结果,10次的平均结果如表1所示。从表1可以看出,随着模态的增加,准确率、AUC等评价指标整体上呈上升趋势,并且3个模态融合之后的分类准确率达到82.92%,表明本文方法可以很好地实现多模态融合。

为了评估方法的临床可行性,还与各多模态融合方法中提出的模型从参数量、训练时间以及推理时间角度进行了比较,对比结果如表2所示,其中参数量最小、时间最短的结果用加粗字体标出。从表2可以看出,本文提出的模型具有较大的参数量,主要是因为实例感知模块引入了高维的全连接层将图像特征投影到高维空间,以便可以更好地捕获模态内的样本相似性信息和模态间的语义一致性信息。相应的,训练时间也会有所增加,相较于参数最小的TMC方法,虽然200轮的训练时间相差接近5s,但模型在推理方面还有小幅领先。这表明复杂的实例感知模块确实起到了作用,并且极少的推理时间也表明该方法构建的模型具有临床可行性。

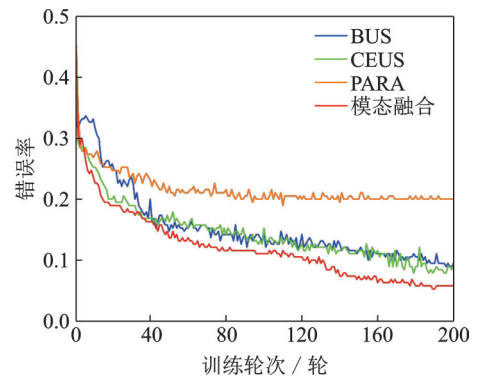


图2 训练时各模态以及融合模态的错误率
Fig.2 Error rate of each modality and fusion during training

表1 单独模态以及任意模态融合的分类准确率

Table 1 Classification accuracy of modalities and fusion of any modalities

| 模态 | ACC | AUC | Macro- F_1 | Weighted- F_1 | % |
|---------------|------------|------------|--------------|-----------------|---|
| BUS | 75.42±3.20 | 82.48±4.96 | 74.07±3.45 | 75.38±3.12 | |
| CEUS | 79.58±4.55 | 86.10±2.77 | 78.33±4.47 | 79.51±4.48 | |
| PARA | 73.96±7.47 | 75.25±5.59 | 72.83±7.55 | 74.03±7.58 | |
| BUS+CEUS | 81.67±3.93 | 89.48±4.12 | 80.68±4.35 | 81.71±3.96 | |
| BUS+PARA | 79.38±4.70 | 85.20±5.47 | 78.56±4.69 | 79.47±4.83 | |
| CEUS+PARA | 81.25±4.56 | 85.68±4.15 | 80.15±4.97 | 81.25±4.48 | |
| BUS+PARA+CEUS | 82.92±2.92 | 89.54±4.73 | 81.74±3.65 | 82.83±3.04 | |

表2 各多模态融合方法中模型的参数量和成本时间比较

Table 2 Comparison of model parameters and time cost in various multimodal fusion methods

| 方法 | 参数量 | 训练时间/s | 推理时间/s |
|------------|-----------|----------|---------|
| MvNNcor | 2 387 718 | 14.276 8 | 0.011 0 |
| TMC | 700 | 10.854 2 | 0.009 4 |
| TMDLO | 700 | 18.905 1 | 0.017 1 |
| ETMC | 1 396 | 12.794 1 | 0.010 4 |
| MMDynamics | 93 899 | 7.196 4 | 0.006 3 |
| Ours | 441 478 | 15.496 1 | 0.007 1 |

本文还记录了训练过程中不确定质量和准确率的变化折线图,如图3所示。可以看出,在训练过程中,该多模态融合方法对HCC和ICC分类的准确率逐渐上升的同时,不确定度在逐渐下降,再次证明了方法的有效性。

3.2.2 对比试验

为了更好地评估提出的多模态融合方法,还与近年来其他多模态方法在多模态超声上的表现进行了比较,对比结果表3所示。从表3可以看出,本文提出的多模态融合方法在各个分类指标上均优于其他多模态融合方法,比起效果第二好的模型,本文方法的分类准确率提高超过3%,且具有更小的标准差,表明本文方法具有更好的融合稳定性。所有这些结果再次验证了该方法的优势。

通过对比表1和表3的结果可以发现,本文提出的融合方法在仅使用BUS和CEUS两个模态的数据时已经展示出优异的表现,对于肝癌诊断的分类准确率已经达到MvNNcor、TMDLO在所有模态上的表现,更加印证本文方法在多模态融合时的优异表现。

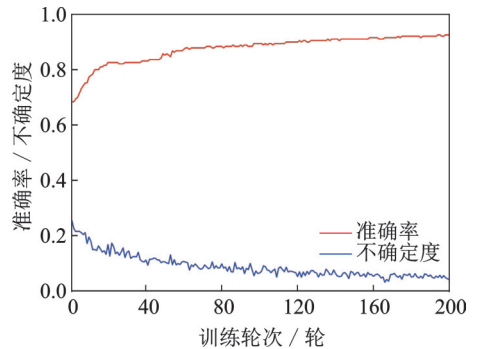


图3 训练过程中准确率和不确定度的变化
Fig.3 Changes in accuracy and uncertainty during training

表3 各多模态融合方法在多模态超声上的对比实验结果

Table 3 Comparative experimental results of various multi-modal fusion methods on multi-modal ultrasound

| 方法 | ACC | AUC | Macro- F_1 | Weighted- F_1 |
|------------|------------|------------|--------------|-----------------|
| MvNNcor | 81.02±2.75 | 88.81±4.34 | 81.31±3.50 | 82.50±2.82 |
| TMC | 81.04±7.12 | 86.49±6.90 | 79.50±7.59 | 80.75±7.17 |
| TMDLO | 81.25±6.11 | 85.19±6.27 | 79.79±6.14 | 80.98±6.18 |
| ETMC | 81.67±5.73 | 86.43±6.26 | 80.07±6.40 | 81.39±5.76 |
| MMDynamics | 82.08±5.76 | 88.09±5.06 | 80.81±6.23 | 81.95±5.82 |
| Ours | 85.21±2.37 | 90.85±4.24 | 84.12±2.94 | 85.12±2.35 |

3.2.3 特征分析

本文还对各个模态的特征进行分析,B模态超声中HCC和ICC之间的差异如图4所示(CEUS和PARA模态中HCC和ICC之间的差异见分别见图5和图6),其中第1行在确诊为HCC的BUS影像,第2行是确诊为ICC的BUS影像。从图中可以看出,HCC和ICC之间的特征差异比较明显,从Spicul特征(反映病灶毛刺信息)来看,HCC相对来说具有较少的毛刺,而ICC则具有较多的毛刺。对比BUS、

CEUS和PARA三个模态中的特征差异,可以发现,BUS、CEUS模态中HCC和ICC之间的特征差异更加显著,对肝癌的诊断具有更高的区分度,这与表1中BUS和CEUS模态的诊断准确率优于PARA模态的结果一致。

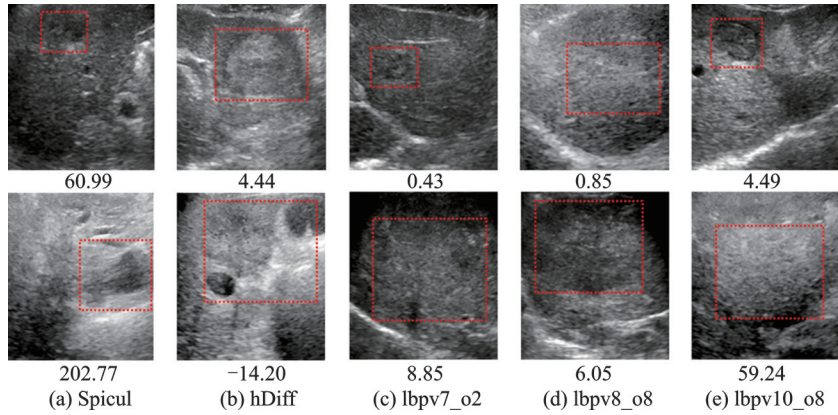


图4 BUS模态中HCC和ICC之间的特征差异

Fig.4 Feature differences between HCC and ICC in BUS modality

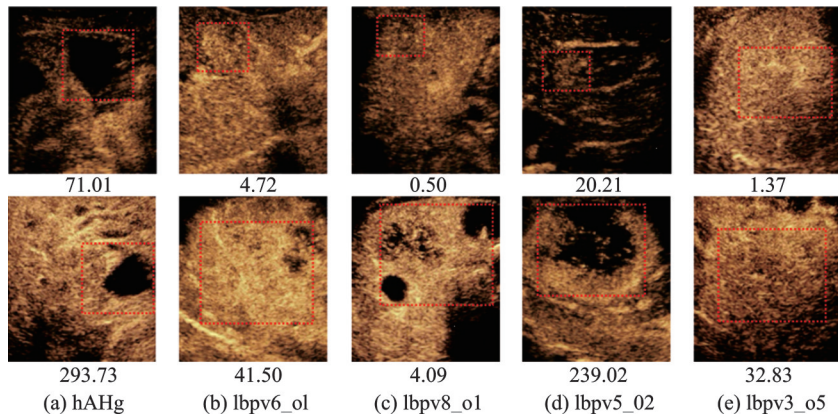


图5 CEUS模态中HCC和ICC之间的特征差异

Fig.5 Feature differences between HCC and ICC in CEUS modality

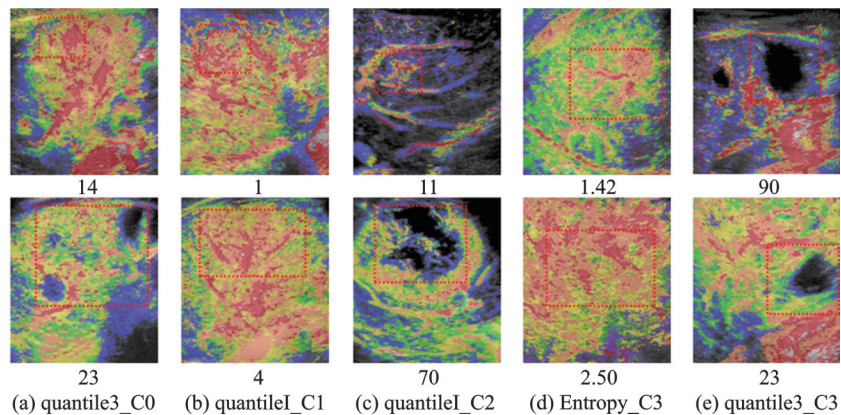


图6 PARA模态中HCC和ICC之间的特征差异

Fig.6 Feature differences between HCC and ICC in PARA modality

3.3 消融实验

本文还进行了详细的消融实验来验证方法中各个关键模块的有效性,包括对多模态融合策略、模态内实例感知损失以及模态间实例感知损失的评估。HCC 和 ICC 的分类准确率如表 4 所示,随着融合策略、模态内实例感知损失以及模态间实例感知的引入,多模态融合策略对于肝癌的分类准确率逐渐提高,表明实例感知模块确实起到作用,可以捕获超声模态内的样本相似性信息和超声模态间的语义一致性信息。消融实验的结果证明本文方法中各个关键模块的有效性。

表 4 消融实验结果

Table 4 Results of ablation experiments

| 融合策略 | 模态内实例感知损失 | 模态间实例感知损失 | 准确率/% |
|------|-----------|-----------|------------|
| — | — | — | 79.79±4.07 |
| ▲ | — | — | 80.21±5.91 |
| √ | — | — | 82.92±2.92 |
| √ | √ | — | 84.58±2.50 |
| √ | — | √ | 84.79±2.47 |
| √ | √ | √ | 85.21±2.37 |

注:“√”表示采用该模块,“—”表示不使用该模块,“▲”表示基于注意力机制。

4 结束语

本文提出的基于多模态超声对比学习的肝癌诊断方法,利用监督对比学习的思想深入挖掘各超声模态特征,捕获超声模态内同类样本之间的相似性信息和超声模态间样本的语义一致性信息,以提高肝癌诊断的准确率。此外,该方法基于主观逻辑为每一个超声模态引入了不确定度的度量,可以更好地衡量超声模态的置信度,通过证据组合规则实现多个模态的动态融合。在多模态超声数据上验证了方法的有效性,对比其他多模态融合方法,肝癌诊断准确率有一定的提升。在后续的研究工作中,将考虑利用多模态超声数据的动态过程提高肝癌诊断的准确率和模型的泛化性能。

参考文献:

- [1] TORRE L A, BRAY F, SIEGEL R L, et al. Global cancer statistics, 2012[J]. CA: A Cancer Journal for Clinicians, 2015, 65(2): 87-108.
- [2] CHEN Wangqing, ZHENG Rongshou, BAADE P D, et al. Cancer statistics in China, 2015[J]. CA: A Cancer Journal for Clinicians, 2016, 66(2): 115-132.
- [3] TSENG H, XIONG Wei, BADETI S, et al. Efficacy of anti-CD147 chimeric antigen receptors targeting hepatocellular carcinoma[J]. Nature Communications, 2020, 11(1): 4810.
- [4] WANG Qiyuan, WANG Zhongmin, SUN Yu, et al. SCCNN: A diagnosis method for hepatocellular carcinoma and intrahepatic cholangiocarcinoma based on siamese cross contrast neural network[J]. IEEE Access, 2020, 8: 85271-85283.
- [5] WANG Anqiang, WU Liangcai, LIN Jianzhen, et al. Whole-exome sequencing reveals the origin and evolution of hepatocellular carcinoma[J]. Nature Communications, 2018, 9(1): 894.
- [6] BALTRUŠAITIS T, AHUJA C, MORENCY L P. Multimodal machine learning: A survey and taxonomy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(2): 423-443.
- [7] HOTELLING H. Relations between two sets of variates[M]. New York: Springer, 1992: 162-190.
- [8] ANDREW G, ARORA R, BILMES J, et al. Deep canonical correlation analysis[C]//Proceedings of the 30th International Conference on Machine Learning. Atlanta, USA: PMLR, 2013: 1247-1255.
- [9] WANG Weiran, ARORA R, LIVESCU K, et al. On deep multi-view representation learning[C]//Proceedings of the 32th

- International Conference on Machine Learning. Lille, France: PMLR, 2015: 1083-1092.
- [10] 杨杨, 詹德川, 姜远, 等. 可靠多模态学习综述[J]. 软件学报, 2021, 32(4): 1067-1081.
YANG Yang, ZHAN Dechuan, JIANG Yuan, et al. Reliable multi-modal learning: A survey[J]. *Journal of Software*, 2021, 32(4): 1067-1081.
- [11] FEICHTENHOFER C, PINZ A, ZISSERMAN A. Convolutional two-stream network fusion for video action recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 1933-1941.
- [12] LIU Xinwang, ZHU Xinzong, LI Miaomiao, et al. Late fusion incomplete multi-view clustering[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 41(10): 2410-2423.
- [13] ILSE M, TOMCZAK J, WELLING M. Attention-based deep multiple instance learning[C]//Proceedings of International Conference on Machine Learning. Stockholm, Sweden: PMLR, 2018: 2127-2136.
- [14] XU Jinglin, LI Wenbin, LIU Xinwang, et al. Deep embedded complementary and interactive information for multi-view classification[C]// Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA: AAAI, 2020, 34(4): 6494-6501.
- [15] HAN Zongbo, YANG Fan, HUANG Junzhou, et al. Multimodal dynamics: Dynamical fusion for trustworthy multimodal classification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 20707-20717.
- [16] KHOSLA P, TETERWAK P, WANG C, et al. Supervised contrastive learning[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 18661-18673.
- [17] JIN Jiaqi, WANG Siwei, DONG Zhibin, et al. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 11600-11609.
- [18] HAN Zongbo, ZHANG Changqing, FU Huazhu, et al. Trusted multi-view classification[EB/OL]. (2021-02-03)[2024-04-30]. <https://arXiv.org/pdf/2102.02051>.
- [19] HAN Zongbo, ZHANG Changqing, FU Huazhu, et al. Trusted multi-view classification with dynamic evidential fusion[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(2): 2551-2566.
- [20] YANG Yang, ZHANG Jingshuai, GAO Fan, et al. DOMFN: A divergence-orientated multi-modal fusion network for resume assessment[C]//Proceedings of the 30th ACM International Conference on Multimedia. New York: ACM, 2022: 1612-1620.
- [21] LIU Wei, YUE Xiaodong, CHEN Yufei, et al. Trusted multi-view deep learning with opinion aggregation[C]//Proceedings of the AAAI Conference on Artificial Intelligence.[S.l.]: AAAI, 2022, 36(7): 7585-7593.
- [22] XUE Zihui, MARCULESCU R. Dynamic multimodal fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 2574-2583.
- [23] ZHANG Qingyang, WU Haitao, ZHANG Changqing, et al. Provable dynamic fusion for low-quality multimodal data[C]// Proceedings of International Conference on Machine Learning. Honolulu: PMLR, 2023: 41753-41769.
- [24] DEMPSTER A P. Upper and lower probabilities induced by a multivalued mapping[M]//Classic works of the Dempster-Shafer Theory of Belief Functions. Berlin: Springer, 2008: 57-72.
- [25] ZHONG Huasong, WU Jianlong, CHEN Chong, et al. Graph contrastive clustering[C]//Proceedings of the IEEE/CVF International Conference on computer Vision. Piscataway: IEEE, 2021: 9224-9233.
- [26] YOU Yuning, CHEN Tianlong, SUI Yongduo, et al. Graph contrastive learning with augmentations[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 5812-5823.
- [27] XIAO Teng, ZHU Huaisheng, CHEN Zhengyu, et al. Simple and asymmetric graph contrastive learning without augmentations[EB/OL]. (2023-10-29)[2024-02-24]. <https://doi.org/10.48550/arXiv.2310.18884>.
- [28] LI Yunfan, HU Peng, LIU Zitao, et al. Contrastive clustering[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2021: 8547-8555.
- [29] HAGER P, MENTEN M J, RUECKERT D. Best of both worlds: Multimodal contrastive learning with tabular and imaging data[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023:

23924-23935.

- [30] SHAN Ziyu, ZHANG Yujie, YANG Qi, et al. Contrastive pre-training with multi-view fusion for no-reference point cloud quality assessment[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 25942-25951.
- [31] SUNG C, KIM W, AN J, et al. Contextrust: Contextual contrastive learning for semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 3732-3742.
- [32] JSANG A. Subjective Logic: A formalism for reasoning under uncertainty[M],[S.l.]: Springer Publishing Company, 2018.
- [33] JOSANG A, HANKIN R. Interpretation and fusion of hyper opinions in subjective logic[C]//Proceedings of 2012 15th International Conference on Information Fusion. Singapore: IEEE, 2012: 1225-1232.
- [34] SENSOY M, KAPLAN L, KANDEMIR M. Evidential deep learning to quantify classification uncertainty[J]. Advances in Neural Information Processing Systems, 2018, 31: 3183-3193.
- [35] 万鹏,刘晗,赵俊勇,等.基于机器学习的超声造影分析综述[J].数据采集与处理,2023,38(4):741-758.
WAN Peng, LIU Han, ZHAO Junyong, et al. Contrast-enhanced ultrasound analysis based on machine learning: A survey[J]. Journal of Data Acquisition and Processing, 2023, 38(4): 741-758.
- [36] RODRÍGUEZ-CRISTERNA A, GÓMEZ-FLORES W, DE ALBUQUERQUE-PEREIRA W C. BUSAT: A MATLAB toolbox for breast ultrasound image analysis[C]//Proceedings of Pattern Recognition: 9th Mexican Conference. [S.l.]: Springer, 2017: 268-277.

作者简介:



杨印凯(2001-),男,硕士研究生,研究方向:多模态融合,E-mail:1595154470@qq.com。



万鹏(1993-),男,讲师,研究方向:医学图像智能计算。



石航(1999-),男,硕士研究生,研究方向:生物信息学。



薛海燕(1985-),女,硕士研究生,研究方向:肝脏超声造影与介入治疗,E-mail:xhainiao@163.com。



邵伟(1986-),通信作者,男,副教授,研究方向:机器学习、医学图像处理,E-mail:shaowei20022005@nuaa.edu.cn。

(编辑:夏道家)