



自监督对比特征学习的多模态乳腺超声诊断

丁维昌, 施俊, 王骏

引用本文:

丁维昌,施俊,王骏. 自监督对比特征学习的多模态乳腺超声诊断[J]. 智能系统学报, 2023, 18(1): 66–74.

DING Weichang,SHI Jun,WANG Jun. Multi-modality ultrasound diagnosis of the breast with self-supervised contrastive feature learning[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(1): 66–74.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202111052>

您可能感兴趣的其他文章

面向不完整多视图聚类的深度互信息最大化方法

Deep mutual information maximization method for incomplete multi-view clustering

智能系统学报. 2023, 18(1): 12–22 <https://dx.doi.org/10.11992/tis.202203051>

融合VAE和StackGAN的零样本图像分类方法

Zero-shot image classification method combining VAE and StackGAN

智能系统学报. 2022, 17(3): 593–601 <https://dx.doi.org/10.11992/tis.202107012>

面向自闭症辅助诊断的无监督模糊特征学习新方法

A novel unsupervised fuzzy feature learning method for computer-aided diagnosis of autism

智能系统学报. 2019, 14(5): 882–888 <https://dx.doi.org/10.11992/tis.201808005>

基于宽度学习方法的多模态信息融合

Multi-modal information fusion based on broad learning method

智能系统学报. 2019, 14(1): 150–157 <https://dx.doi.org/10.11992/tis.201803022>

利用多模态U形网络的CT图像前列腺分割

Prostate segmentation in CT images with multimodal U-net

智能系统学报. 2018, 13(6): 981–988 <https://dx.doi.org/10.11992/tis.201806012>

DOI: 10.11992/tis.202111052

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20221213.1433.001.html>

自监督对比特征学习的多模态乳腺超声诊断

丁维昌, 施俊, 王骏

(上海大学 通信与信息工程学院, 上海 200444)

摘要: 超声图像的乳腺癌自动诊断具有重要的临床价值。然而, 由于缺乏大量人工标注数据, 构建高精度的自动诊断方法极具挑战。近年来, 自监督对比学习在利用无标签自然图像产生具有辨别性和高度泛化性的特征方面展现出巨大潜力。然而, 采用自然图像构建正负样本的方法在乳腺超声领域并不适用。为此, 本文引入超声弹性图像 (elastography ultrasound, EUS), 利用超声图像的多模态特性, 提出一种融合多模态信息的自监督对比学习方法。该方法采用同一病人的多模态超声图像构造正样本; 采用不同病人的多模态超声图像构建负样本; 基于模态一致性、旋转不变性和样本分离性来构建对比学习的目标学习准则。通过在嵌入空间中学习两种模态的统一特征表示, 从而将 EUS 信息融入模型, 提高了模型在下游 B 型超声分类任务中的表现。实验结果表明本文提出的方法能够在无标签的情况下充分挖掘多模态乳腺超声图像中的高阶语义特征, 有效提高乳腺癌的诊断正确率。

关键词: 自监督学习; 对比学习; 超声图像; 弹性超声; B 型超声; 多模态; 乳腺癌; 计算机辅助诊断; 深度学习
中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2023)01-0066-09

中文引用格式: 丁维昌, 施俊, 王骏. 自监督对比特征学习的多模态乳腺超声诊断 [J]. 智能系统学报, 2023, 18(1): 66-74.

英文引用格式: DING Weichang, SHI Jun, WANG Jun. Multi-modality ultrasound diagnosis of the breast with self-supervised contrastive feature learning[J]. CAAI transactions on intelligent systems, 2023, 18(1): 66-74.

Multi-modality ultrasound diagnosis of the breast with self-supervised contrastive feature learning

DING Weichang, SHI Jun, WANG Jun

(School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China)

Abstract: An automatic ultrasound-based diagnosis of breast cancer has important clinical value. However, high-precision automatic diagnosis methods are very difficult to construct because many labeled data are missing. Recently, self-supervised contrastive learning has shown great potential in using unlabeled natural images to generate discriminative and highly generalized features. However, this approach is not applicable to using natural images to construct positive and negative samples in the field of breast ultrasound. To this end, this work introduces the elastography ultrasound (EUS) image and proposes a self-supervised contrastive learning method integrating multimodal information based on the multimodal features of an EUS image. Specifically, positive and negative samples are constructed using multi-modality ultrasound images collected from the same and different patients, respectively. We construct the object learning criterion of contrastive learning based on modal consistency, rotation invariance, and sample separation. The EUS information is integrated into the model by learning the unified feature representation for both modalities in the embedding space, which improves model performance in the downstream B-mode ultrasound classification task. The experimental results show that our method can fully mine the high-level semantic features from unlabeled multimodal breast ultrasound images, thereby effectively improving the diagnosis accuracy of breast cancer.

Keywords: self-supervised learning; contrastive learning; ultrasound image; flexible ultrasound; B-mode ultrasound; multi-modality; breast cancer; computer-aided diagnosis; deep learning

B 型超声 (B-mode ultrasound, BUS) 是乳腺癌早期诊断的重要工具, 它具有价格低、方便快捷

等优点^[1-2]。目前, 基于 BUS 的计算机辅助诊断模型已广泛应用于乳腺癌的早期诊断中^[3-4]。虽然这些模型可以达到早期诊断的目的, 但是它们在训练过程中仅使用了 BUS 单一模态。由于 BUS

收稿日期: 2021-11-29. 网络出版日期: 2022-12-14.

基金项目: 上海市自然科学基金项目 (20ZR1419900).

通信作者: 王骏, E-mail: wangjun_shu@shu.edu.cn.

仅能提供病灶的结构形态和内部回声信息,这在一定程度上限制了模型的性能。近年来,超声弹性成像(elastography ultrasound, EUS)在临床诊断中得到了越来越多的应用。EUS可以提供病灶生物力学和功能特性的相关信息,在乳腺癌早期诊断过程中可作为BUS信息的有益补充,结合二者能够有效提高乳腺癌的诊断准确率^[5-6]。但是,基层医院大多没有配置高性能的EUS采集设备,只能通过采集BUS图像进行乳腺癌早期诊断。因此,开发融合EUS诊断信息的BUS计算机辅助诊断模型具有重要意义。

为了得到令人满意的BUS计算机辅助诊断模型,通常需要使用大量带标注的超声图像来训练模型^[7-8]。但是,获取大量有标注的多模态超声图像难度很大。一方面,对超声图像进行标注专业性很强,要求标注者具有专业的医学背景知识;另一方面,还需要耗费大量的时间和人力成本。自监督学习利用无标签数据,从数据本身寻找监督信息来引导模型的训练,是克服标注样本不足的有效方法^[9]。对比学习作为一种重要的自监督学习方法,在计算机视觉领域已经得到了初步的应用。它基于数据自身构建正、负样本实例,利用正样本的聚集性和负样本的分离性从训练数据本身学习到相应的特征嵌入表示。代表性工作包括MoCo^[10]、SimCLR^[11]、PIRL^[12]和BYOL^[13]等。常用的正负样本构造方法大多基于遮挡、去色等随机数据增强方法。但是这些方法大多仅适用于自然图像。与自然图像相比,超声图像灰度不均匀,对比度低。因此,将这些正负样本构造方法应用于超声图像不能取得理想的对比学习效果,从而影响下游分类任务的表现。

在乳腺癌的超声计算机辅助诊断中,EUS通过描述感兴趣区域的硬度,能够很好地体现组织内部的变化。相较于BUS,它从另一个角度体现了病灶信息。由于采集自同一病灶的两种模态的超声图像描述了相同的疾病情况,它们体现一致的疾病信息,本文将其称为模态一致性;同时,对于下游疾病分类任务,患者的疾病分类结果不会因图像旋转而改变,本文称其为旋转不变性。另一方面,因存在个体差异,来自不同患者的样本病灶信息应相互区别,本文将其称为样本分离性。本文从多模态超声图像的模态一致性、旋转不变性和样本分离性出发,提出一种新型的正负样本实例构建方法;在此基础上,利用正样本的聚集性和负样本的分离性,提出一种融合多模态信息的自监督对比学习算法用于乳腺癌肿瘤良恶

性诊断。

本文创新如下:

1) 针对乳腺超声图像训练样本有限的问题,本文提出一种新颖的自监督对比学习算法用于乳腺癌超声诊断;

2) 基于超声图像的多模态特性,提出一种基于模态一致性、旋转不变性和样本分离性的跨模态正负样本实例构建方法;

3) 针对单模态BUS包含诊断信息有限这一问题,通过优化对比损失函数,使模型学习到BUS和EUS这两种模态的公共特征表示,隐性的将多模态信息融入模型,提升了模型在下游单模态BUS分类任务中的性能。

1 相关工作

自监督学习通过构建辅助任务从无监督数据中挖掘监督信息,并利用构建的监督信息来训练网络,从而可以学到对下游任务有效的特征表示,已成为当前机器学习领域的一个研究热点^[9]。比较著名的工作有:Gidaris等^[14]通过训练模型识别输入图像的旋转角度,Pathak等^[15]使用神经网络填补图像中大量缺失的区域,使得神经网络学到图片的深层特征,Noroozi等^[16]通过预测图像的拼接顺序构建辅助任务,达到特征学习的目的。在医学图像领域中,自监督学习也得到较为成功的应用,例如:Zhuang等^[17]分别从CT和MR图像中提出了用于脑出血分类和肿瘤分割的立方体恢复任务;Bai等^[18]提出了一种通过预测解剖位置来学习自监督特征的心脏MR图像分割方法;Blendowski等^[19]设计图像内在空间偏移关系任务学习自监督特征;Spitzer等^[20]利用从同一个大脑中两个patches之间的三维距离的预测作为辅助任务。但是,这些辅助任务大都依赖于空间位置、图像颜色等单一维度的特征,学到的特征缺乏泛化性,难以迁移到下游任务中。

近年来,对比学习作为一种有效的辅助任务,在自监督特征表示学习中取得了很好的效果。对比学习通过设计适合的对比损失函数^[21]使得相似样本的特征表示尽可能接近,不相似样本的特征表示尽可能远离。相似性往往以一种无监督的方式定义,通常情况下使用同一图像的不同变换作为相似样本。例如,Dosovitskiy等^[22]通过一致性假设获取每个训练样本的参数特征表示,Wu等^[23]在Dosovitskiy等^[22]的基础上使用memory bank去存储数据样本的高阶特征表示。Chen等^[11]提出一种不需要特定网络或者memory bank

的对比学习算法,该方法使用大批量样本实例并获取每个训练实例的负样本,使得每一次迭代学到的特征质量都有所提高。这些方法大多采用单一的数据增强的方法来构建正负样本实例。由于超声图像相较于自然图像亮度不均,边界模糊,对比度低,使用单一的数据增强的方法去构建正

负样本在乳腺超声领域并不适用。

2 自监督对比学习方法

2.1 流程概述

拟提出的模型包含特征提取网络和 k-NN 分类器两个模块,算法流程如图 1 所示。

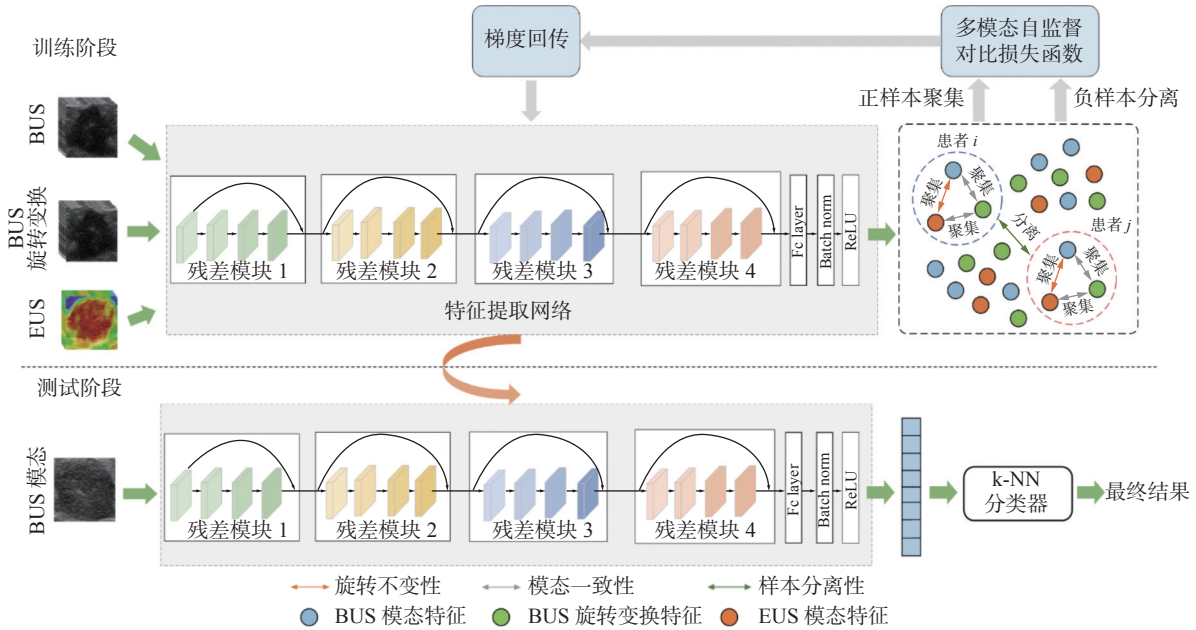


图 1 融合多模态乳腺超声数据的自监督对比特征学习框架

Fig. 1 Self supervised contrastive learning framework for multimodal ultrasound breast cancer diagnosis

考虑到残差网络可以减少深度网络中的梯度消散,具有泛化能力强等诸多优点,本文使用 ResNet18^[24] 作为特征提取网络,对最后一个残差模块的输出进行平均池化操作,然后展开成一维向量,依次送入全连接层、Batch Norm 层和 ReLU 层。在模型的训练阶段,通过 BUS 模态、BUS 旋转变换和弹性模态 EUS 构建正负样本,融合多模态超声数据进行自监督对比特征学习。在测试阶段,将 BUS 模态图像送入特征提取网络得到其嵌入特征表示,最后送入 k-NN 分类器,得到最终分类结果。

2.2 正负样本构建

为便于描述,设 $B = \{b_1, b_2, \dots, b_n\}$ 和 $E = \{e_1, e_2, \dots, e_n\}$ 分别为 BUS 和 EUS 图像集合。患者 i 的 BUS 模态图像为 b_i ; 患者 i 的 EUS 模态图像为 e_i ; b_i^T 为 b_i 的旋转图像; 令 $h_\theta(\cdot)$ 为特征嵌入网络,使用 $f_i = h_\theta(b_i)$, $g_i = h_\theta(e_i)$, $f_i^T = h_\theta(b_i^T)$ 分别代表患者 i 的 BUS 模态、EUS 模态和 BUS 旋转图像经特征提取网络处理后得到的高阶特征。本文使用 $s_{\text{im}}(z_i, z_j) = z_i^T z_j / (\tau z_i^T z_j)$ 度量两个特征向量间的相似性, τ 为温度参数,代表样本分布的聚集水平^[25]。

2.2.1 利用模态一致性和旋转不变性构建正样本

由于同一患者 i 的 BUS 模态 b_i 和 EUS 模态 e_i 具有模态一致性,所以它们经网络提取得到的高阶特征 f_i 和 g_i 应该保持一致; 另一方面,由于 BUS 模态的旋转不变性, b_i 经网络提取得到的高阶特征 f_i 与其旋转图像 b_i^T 送入网络后得到的高阶特征 f_i^T 应该具有相同的特征表示。基于模态一致性和旋转不变性,本文将同一患者的 BUS 图像、BUS 旋转图像和 EUS 图像进行组合作为正样本,即将 (b_i, e_i) 和 (b_i, b_i^T) 作为正样本。为此,本文定义:

$$m(i) = \frac{\exp(\text{sim}(f_i, g_i)/\tau)}{\sum_{k=1}^N \exp(\text{sim}(f_k, g_k)/\tau)} \quad (1)$$

$$r(i) = \frac{\exp(\text{sim}(f_i, f_i^T)/\tau)}{\sum_{k=1}^N \exp(\text{sim}(f_k, f_k^T)/\tau)} \quad (2)$$

式中: N 为同一训练批次内所有样本数量; $m(i)$ 用来衡量高阶特征 f_i 和 g_i 一致的程度,两个向量越相似, $m(i)$ 的值越大; $r(i)$ 用来衡量 f_i 与 f_i^T 一致的程度,两个向量越相似, $r(i)$ 的值越大。为了使学习到的特征保持模态一致性和旋转不变性,需要通过优化 $h_\theta(\cdot)$ 使 $m(i)$ 和 $r(i)$ 尽可能大。

2.2.2 利用样本分离性构建负样本

在缺乏监督信息的前提下, 患者 i 的 BUS 模态图像 b_i 应与患者 j 的 BUS 旋转变换图像 b_j^T 、EUS 模态图像 e_j 互相区别, 因此 b_i 经过网络提取得到的特征 f_i 与 b_j^T 、 e_j 经网络提取得到的特征 f_j^T 、 g_j 应尽可能远离。基于样本分离性, 本文将不同患者的 BUS 图像、BUS 旋转图像和 EUS 图像进行组合作为负样本, 即将 (b_i, b_j^T) 和 (b_i, e_j) 作为负样本。为此, 本文定义:

$$S(i, j) = \begin{cases} 1 - \frac{\exp(\text{sim}(f_i, g_i)/\tau)}{\sum_{k=1}^N \exp(\text{sim}(f_k, g_i)/\tau)} \\ \left(1 - \frac{\exp(\text{sim}(f_i, f_j^T)/\tau)}{\sum_{k=1}^N \exp(\text{sim}(f_k, f_j^T)/\tau)} \right), j \neq i \end{cases} \quad (3)$$

式中: N 为同一训练批次内所有样本数量, $S(i, j)$ 表示患者 i 与患者 j 之间的样本分离的程度。 $\text{sim}(f_i, g_i)/\tau$ 越小, 或者 $\text{sim}(f_i, f_j^T)/\tau$ 越小, $S(i, j)$ 值越大, 表示患者 i 与患者 j 之间的样本分离性越强。

2.2.3 对比学习损失函数构建

基于上述对正负样本的定义, 经过对比学习后, 来自同一患者的多模态图像在嵌入空间中应该尽可能聚集, 来自不同患者的多模态图像在嵌入空间中应该尽可能分离 (如图 2 所示)。为此基于正样本聚集性和负样本分离性构造目标学习准则如下:

$$L_i = -\log m(i) - \log r(i) - \sum_{j \neq i} \log S(i, j) \quad (4)$$

将 N 名患者的损失函数取平均值并加入正则化项得到最终损失函数为:

$$L = \frac{1}{N} \sum_{i=1}^N L_i + \lambda \|\theta\|_2^2 \quad (5)$$

式中 λ 为超参数。在训练过程中, 最小化公式 (5), 即可得到模型的最优参数, 从而得到两种模态间的统一特征表示。

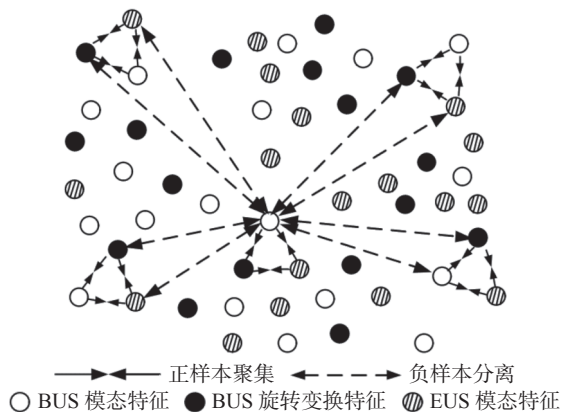


图 2 正负样本的特征嵌入空间

Fig. 2 The illustration of feature embedding space of positive-negative samples

3 实验与分析

3.1 技术细节

3.1.1 实验设置

本文所有实验数据来自南京鼓楼医院, 该数据集包括 264 对 BUS 和 EUS 图像, 其中良性肿瘤 129 例, 恶性肿瘤 135 例。所有数据均由经验丰富的放射科医生使用迈瑞 Resona7 超声扫描仪采集。为了得到更多的训练样本, 在训练集上进行数据增强, 包括随机进行缩放裁剪、水平翻转, 灰度处理, 并随机改变输入图像的亮度、对比度和饱和度等操作。实验中的所有样本均采用相同的数据增强策略。

本文选用 PyTorch 深度学习框架搭建神经网络, 并选用 NVIDIA Titan X Pascal GPU 的显卡进行加速计算。batch 大小设为 75; 优化器为 Adam; epoch 大小为 3000; 初始学习率设置为 0.0001, 每 1000 个 epoch 下降 0.1; 最终的结果采用 5 折交叉验证。

3.1.2 评估公式

本文使用的评价指标分别为准确率 A_{CC} 、查准率 P_{RE} 、查全率 R_{EC} 和 F_1 分数, 分别定义为

$$A_{CC} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$P_{RE} = \frac{TP}{TP + FP}$$

$$R_{EC} = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2P_{RE}R_{EC}}{P_{RE} + R_{EC}}$$

式中: TP 代表正确分类为恶性肿瘤患者的人数; TN 代表正确分类为良性肿瘤患者的人数; FP 代表良性肿瘤患者被分类为恶性肿瘤患者的人数; FN 代表恶性肿瘤患者被分类为良性肿瘤患者的人数。

3.2 消融实验

3.2.1 不同的模态融合方式比较

本文方法综合考虑模态一致性、旋转不变性和样本分离性, 基于自监督对比特征学习对 BUS 和 EUS 模态图像进行融合, 从而得到最佳的多模态特征嵌入表示。为了验证所得特征的有效性, 本节使用 k-NN 分类器进行分类, 并使用其他两种融合方法进行比较, 方法具体说明如表 1 所示, 所得结果如表 2 可示。

从表 2 中可以看出, 使用本文方法所得结果相较于仅使用 BUS 模态进行训练和测试的结果, 准确率提高了 0.114, 说明本文提出的方法可以有效提高乳腺癌的诊断正确率。使用本文方法进行特征学习所得到的特征, 相较于其他两种方法所

得到的特征, 进行 k-NN 分类所得结果的 AUC 分别提高 0.056 和 0.058。这说明了本文提出的多模态融合方法能够有效的提高自监督对比学习模型提取图像高阶信息的能力。

表 1 不同的模态融合方法描述

Table 1 Description about different modal fusion methods

方法	描述
仅使用BUS模态	只使用BUS模态进行训练和测试
直接整合	将EUS模态数据集和BUS模态进行直接混合
BUS+EUS	将EUS模态作为对应BUS模态的一种增强方式

表 2 不同的模态融合方法描述结果比较

Table 2 Comparison of the performances of different modal fusion methods

方法	A_{UC}	A_{CC}	P_{RE}	R_{EC}	F_1
BUS	0.734	0.735	0.741	0.734	0.737
直接整合	0.791	0.793	0.785	0.804	0.794
BUS+EUS	0.789	0.793	0.832	0.789	0.809
本文方法	0.847	0.849	0.858	0.848	0.852

通过使用 T-SNE^[26] 降维算法对获得的特征进行降维处理, 图 3 进一步可视化了 3 种不同融合方法的特征表示。随机选取一定数量样本, 每个

样本包含 BUS(蓝色)、旋转后的 BUS(绿色)、EUS(红色) 这 3 种模态。将每个样本送入网络获取最终的特征表示, 使用降维算法 T-SNE^[26] 对获得的特征进行降维处理, 然后对其进行可视化处理, 得到如图 3 所示结果。图中, 蓝色样本点与红色样本点越接近, 意味着获得了良好的模态一致性特征, 蓝色样本点与绿色样本点相接近, 意味着获得了良好的旋转不变性特征。

从图 3(a) 可以看出直接将 EUS 模态作为数据集扩充不能取得理想效果, BUS(蓝色) 和 EUS(红色) 两种模态的图像之间并没有明显相关性, 这是因为直接混合两种模态的融合方案仅考虑了旋转不变性, 无法考虑不同模态之间的一致性特征。从图 3(b) 可以看出将 EUS 作为数据增强可以将 BUS(蓝色) 和 EUS(红色) 间的距离更加接近, 然而不同患者间的特征并未分开, 这种融合方式在一定程度上学习到了不同模态间的交互信息, 但是却忽略了不同个体间的差异。从图 3(c) 可以看出, 本文方法可以最小化同一患者 BUS 模态、BUS 模态旋转变换和 EUS 模态间的距离, 同时使不同患者间的特征距离尽可能的大, 特征分布更加均匀, 保留了尽可能多的有用信息, 说明了本文提出的自监督对比学习算法学得了较好的嵌入特征表示。

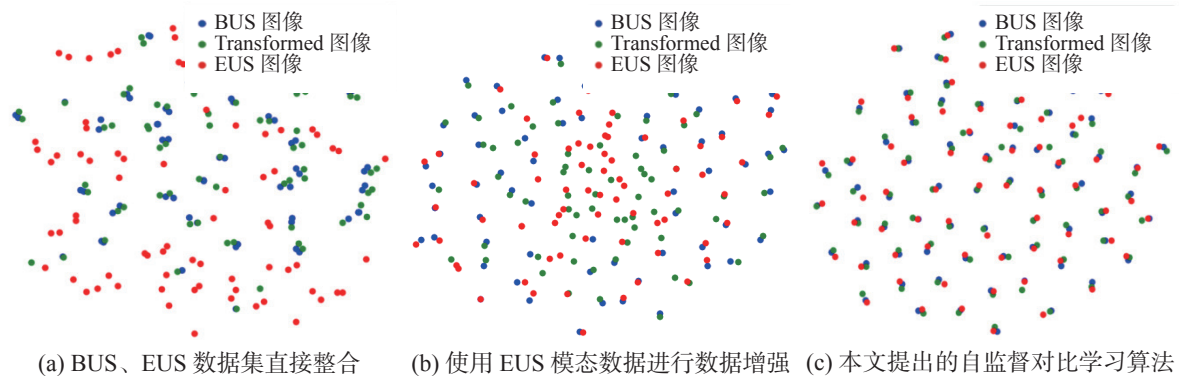


图 3 学习到的特征嵌入空间可视化

Fig. 3 The visualization of feature embedding space

3.2.2 不同深度的网络架构比较

使用不同深度的网络架构对分类结果会产生影响, 本节对此进行研究。本文选用 ResNet18、ResNet34 和 ResNet50 进行比较。实验结果如表 3 所示。从表 3 可以看出, ResNet18 取得了最好的结果, 相较于 ResNet34 和 ResNet50, AUC 分别提高了 0.094 和 0.118。该结果表明, 对于多模态乳腺超声数据, 更深层次的网络结构并没有学得更有利于分类的嵌入特征表示。这是因为过多的网络层数使原输入图像中的模态特有信息在多次卷

积后丢失, 使得网络所得到的特征中没有包含原输入图像中的模态特有信息。

表 3 不同深度的网络架构比较结果

Table 3 Comparison of the performances of network backbone in different depths

Backbone	A_{UC}	A_{CC}	P_{RE}	R_{EC}	F_1
ResNet18	0.847	0.849	0.858	0.848	0.852
ResNet34	0.753	0.755	0.764	0.753	0.752
ResNet50	0.729	0.731	0.736	0.729	0.732

3.2.3 正样本的不同构造方式对分类性能的影响

本文方法在构建正样本时同时涉及 BUS 模态、EUS 模态和 BUS 模态的旋转变换。为了考查这 3 个因素对分类效果的影响, 本节使用不同方法来构建正样本, 并修改相应的对比学习目标函数, 以验证本文的正负样本构造方法的有效性。具体构造方法一共有 3 种, 分别为方案一: 仅考虑旋转不变性; 方案二: 仅考虑模态一致性; 方案三: 同时考虑旋转不变性、模态一致性。实验结果如表 4 所示。

表 4 正样本的构造方法对实验结果的影响

Table 4 Comparison of the performances of different construction method of positive samples

正样本构造方式	A_{UC}	A_{CC}	P_{RE}	R_{EC}	F_1
方案一	0.699	0.698	0.698	0.697	0.679
方案二	0.753	0.755	0.764	0.753	0.758
方案三	0.847	0.849	0.858	0.848	0.852

从表 4 中可以看出, 同时使用旋转一致性和模态一致性去构建正样本, 相较于仅仅使用模态一致性或旋转不变性构建正样本 AUC 分别提高 0.094 和 0.148。这说明本文提出的正样本构建方法更有利于网络学习具有判别性的特征表示。

3.3 与基于 BUS 的分类算法比较

为了验证本文算法的有效性, 我们首先基于单模态 BUS, 将本文算法与 JWSSDL^[27]、FCN-Alex⁺进行比较。具体算法说明如表 5 所示, 表 6 给出了不同算法在相同数据集上的结果。

表 5 基于 BUS 的分类算法描述

Table 5 Description about classification algorithms based on BUS modality

方法	描述
JWSSDL ^[27]	基于弱监督与半监督学习的 BUS 分类方法。在训练与测试阶段均使用 BUS 模态。
FCN-Alex ^[28]	基于全卷积神经网络的端到端深度学习方法。在训练与测试阶段均使用 BUS 模态。

表 6 不同方法分类结果比较

Table 6 Comparison of the performances of different classification algorithms

方法	A_{UC}	A_{CC}	P_{RE}	R_{EC}	F_1
JWSSDL	0.828	0.830	0.843	0.828	0.835
FCN-Alex	0.772	0.773	0.780	0.772	0.771
本文方法	0.847	0.849	0.858	0.848	0.852

从表 6 中可以看出, 与 JWSSDL^[27]、FCN-Alex^[28] 算法相比, 本文算法可以得到更好的结果, 这表明在训练阶段将 EUS 信息融入模型可以

有效提高测试阶段 BUS 模型的性能。

3.4 与其他多模态融合算法比较

为了验证本文提出的算法在多模态信息融合方面的有效性, 将本文方法与 DAAN^[29]、DSAN^[30]、JDDA^[31] 这 3 种方法进行比较。这 3 种方法均通过领域自适应的方法在训练阶段融合 EUS 信息, 具体说明如表 7 所示, 表 8 为比较结果。

表 7 不同的多模态融合算法描述

Table 7 Description about different multimodal fusion algorithms

方法	描述
DAAN ^[29]	一种使用对抗模型学习域不变表达的深度域自适应网络。通过学习 BUS 和 EUS 模态间域不变表达特征使得多模态信息融入网络模型。
DSAN ^[30]	基于局部最大均值差异的深度域自适应网络。通过对齐 BUS 和 EUS 模态间的相关子域分布使得多模态信息融入网络模型。
JDDA ^[31]	一种联合域对齐和判别式特征的深度域自适应网络。通过学习 BUS 和 EUS 模态间的深度判别式特征将多模态信息融入网络模型。

表 8 不同的多模态融合算法分类结果比较

Table 8 Comparison of the performances of different multimodal fusion algorithms

方法	A_{UC}	A_{CC}	P_{RE}	R_{EC}	F_1
DAAN	0.827	0.830	0.856	0.827	0.841
DSAN	0.812	0.813	0.812	0.811	0.811
JDDA	0.837	0.839	0.850	0.837	0.843
本文方法	0.847	0.849	0.858	0.848	0.852

从表 8 中可以看出, 虽然各算法在训练阶段融合了 EUS 信息, 但是由于受样本数量的限制, DAAN^[29]、DSAN^[30]、JDDA^[31] 等方法带来的性能提升有限。本文方法通过自监督对比学习, 从有限训练样本中充分挖掘出了监督信息, 因此可以取得更好的效果。

3.5 与其他自监督学习算法比较

为了验证本文提出的自监督对比特征学习的有效性, 本文选择常见的自监督学习方法进行比较, 具体的算法说明如表 9 所示, 表 10 为不同算法在相同数据集上的结果。

从表 10 中可知, 与主流自监督学习方法相比, 本文提出的算法 A_{UC} 提高至少 0.023, A_{CC} 提高至少 0.018, P_{RE} 提高至少 0.026, R_{EC} 提高至少 0.012, F_1 得分至少提高 0.042。图 4 为不同自监督学习算法的 ROC 曲线, 其中本文提出的算法获得了最高的 A_{UC} 值 0.847, 表明了该算法具有最好的分类性能。

表 9 不同的自监督学习方法描述

Table 9 Description about different self-supervised learning methods

方法	描述
SimCLR ^[11]	随机抽取一个小批量, 给每张BUS图像实例设定两个独立的增强函数, 让相关视图互相吸引, 不相关视图互相排斥。
Invariant ^[32]	通过随机旋转数据增强BUS图像构建正负样本, 采用了孪生网络进行训练。
MoCo ^[10]	一种新颖的无监督表征学习方法, 利用对比损失函数学习图像数据的高阶特征。
PIRL ^[12]	将BUS图像进行重新拼接排列, 与原来的BUS图像构建正负样本对, 利用孪生网络和对比损失函数学习BUS图像的高阶特征表示。
CPC v2 ^[33]	将BUS图像分割成重叠的方块, 通过编码器得到最终的表示向量, 使用对比损失函数学习特征的高阶表示。

表 10 不同的自监督学习方法分类结果比较

Table 10 Comparison of the performances of different self-supervised learning methods

方法	A_{UC}	A_{CC}	P_{RE}	R_{EC}	F_1
SimCLR	0.774	0.754	0.785	0.734	0.758
Invariant	0.789	0.793	0.832	0.789	0.809
MoCo	0.824	0.831	0.786	0.836	0.810
PIRL	0.754	0.777	0.747	0.827	0.784
CPC v2	0.805	0.824	0.786	0.814	0.799
本文方法	0.847	0.849	0.858	0.848	0.852

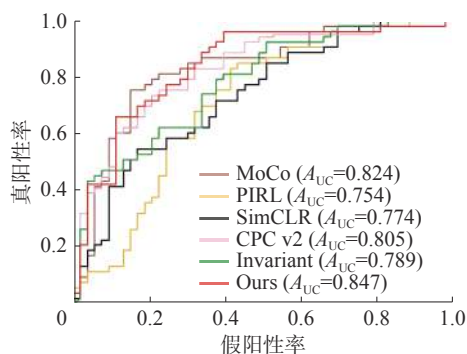


图 4 不同自监督学习算法的 ROC 曲线

Fig. 4 ROC curve of different self-supervised methods

上述实验结果验证了本文提出的多模态自监督对比学习算法相较于其他自监督算法在乳腺癌肿瘤分类上效果具有显著提升。

4 结束语

针对乳腺超声图像训练样本有限的问题, 本文提出一种融合多模态信息的自监督对比特征学习算法, 在训练阶段引入弹性超声图像, 构建

B 型超声图像和弹性超声图像的跨模态正负样本对, 通过对比学习挖掘两种模态间的一致性信息, 得到多模态数据的特征嵌入空间表示。最终实验结果表明本文提出的自监督对比学习算法相较于其他算法取得了最好的分类效果, 有效提高了乳腺癌肿瘤的良恶性诊断率。利用多模态数据间的一致性可以从超声数据自身挖掘出更好的嵌入特征表示, 未来的工作将持续探索不同超声模态特征信息对乳腺癌自动诊断算法的影响。

参考文献:

- [1] CHENG H D, SHAN Juan, JU Wen, et al. Automated breast cancer detection and classification using ultrasound images: a survey[J]. *Pattern recognition*, 2010, 43(1): 299–317.
- [2] 龚勋, 杨菲, 杜章锦, 等. 甲状腺、乳腺超声影像自动分析技术综述 [J]. *软件学报*, 2020, 31(7): 2245–2282.
GONG Xun, YANG Fei, DU Zhangjin, et al. Survey of automatic ultrasonographic analysis for thyroid and breast[J]. *Journal of software*, 2020, 31(7): 2245–2282.
- [3] HUANG Qinghua, ZHANG Fan, LI Xuelong. Machine learning in ultrasound computer-aided diagnostic systems: a survey[J]. *BioMed research international*, 2018, 2018: 5137904.
- [4] 郑光远, 刘峡壁, 韩光辉. 医学影像计算机辅助检测与诊断系统综述 [J]. *软件学报*, 2018, 29(5): 1471–1514.
ZHENG Guangyuan, LIU Xiabi, HAN Guanghui. Survey on medical image computer aided detection and diagnosis systems[J]. *Journal of software*, 2018, 29(5): 1471–1514.
- [5] SIGRIST R M S, LIAU J, KAFFAS A E, et al. Ultrasound elastography: review of techniques and clinical applications[J]. *Theranostics*, 2017, 7(5): 1303–1329.
- [6] 刘飞, 张俊然, 杨豪. 基于深度学习的医学图像识别研究进展 [J]. *中国生物医学工程学报*, 2018, 37(01): 86–94.
LIU Fei, ZHANG Junran, YANG Hao. Advances in medical image recognition based on deep learning[J]. *Journal of biomedical engineering*, 2018, 37(01): 86–94.
- [7] 高歌, 马帅, 王霄英. 计算机辅助诊断在医学影像诊断中的基本原理和应用进展 [J]. *放射学实践*, 2016, 31(12): 1127–1129.
GAO Ge, MA Shuai, WANG Xiaoying. Basic principle and application progress of computer aided diagnosis in medical image diagnosis[J]. *Radiologic practice*, 2016, 31(12): 1127–1129.

- [8] 周涛, 陆惠玲, 陈志强, 等. 多模态医学影像融合识别技术研究进展 [J]. 生物医学工程学杂志, 2013, 30(05): 1117–1122.
ZHOU Tao, LU Huiling, CHEN Zhiqiang, et al. Research progress of multimodal medical image fusion recognition technology[J]. Journal of Biomedical Engineering, 2013, 30(05): 1117–1122.
- [9] JING Longlong, TIAN Yingli. Self-supervised visual feature learning with deep neural networks: a survey[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2021, 43(11): 4037–4058.
- [10] HE Kaiming, FAN Haoqi, WU Yuxin, et al. Momentum contrast for unsupervised visual representation learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 9726–9735.
- [11] CHEN Ting, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//ICML'20: Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 1597–1607.
- [12] MISRA I, VAN DER MAATEN L. Self-supervised learning of pretext-invariant representations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 6706–6716.
- [13] GRILL J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent—a new approach to self-supervised learning[J]. *Advances in neural information processing systems*, 2020(33): 21271–21284.
- [14] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised representation learning by predicting image rotations [EB/OL]. (2018–03–21)[2021–11–29].<https://arxiv.org/abs/1803.07728>.
- [15] PATHAK D, KRÄHENBÜHL P, DONAHUE J, et al. Context encoders: feature learning by inpainting[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2536–2544.
- [16] NOROOZI M, FAVARO P. Unsupervised learning of visual representations by solving jigsaw puzzles[M]//Computer Vision – ECCV 2016. Cham: Springer International Publishing, 2016: 69–84.
- [17] ZHUANG Xinrui, LI Yuexiang, HU Yifan, et al. Self-supervised feature learning for 3D medical images by playing a rubik's cube[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019: 420–428.
- [18] BAI Wenjia, CHEN Chen, TARRONI G, et al. Self-supervised learning for cardiac MR image segmentation by anatomical position prediction[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019: 541–549.
- [19] BLENDOWSKI M, NICKISCH H, HEINRICH M P. How to learn from unlabeled volume data: self-supervised 3D context feature learning[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019: 649–657.
- [20] SPITZER H, KIWITZ K, AMUNTS K, et al. Improving cytoarchitectonic segmentation of human brain areas with self-supervised Siamese networks[M]//Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Cham: Springer International Publishing, 2018: 663–671.
- [21] HADSELL R, CHOPRA S, LECUN Y. Dimensionality reduction by learning an invariant mapping[C]//2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2006: 1735–1742.
- [22] DOSOVITSKIY A, FISCHER P, SPRINGENBERG J T, et al. Discriminative unsupervised feature learning with exemplar convolutional neural networks[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 38(9): 1734–1747.
- [23] WU Zhirong, XIONG Yuanjun, YU S X, et al. Unsupervised feature learning *via* non-parametric instance discrimination[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3733–3742.
- [24] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [25] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. (2015–03–09)[2021–11–29].<https://arxiv.org/abs/1503.02531>.
- [26] VAN der MAATEN, LAURENS, HINTON Geoffrey. Visualizing data using t-SNE[J]. *Journal of machine learning research*, 2008, 9(11).
- [27] SHIN S Y, LEE S, YUN I D, et al. Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images[J]. *IEEE transactions on medical imaging*, 2019, 38(3): 762–774.
- [28] YAP M H, GOYAL M, OSMAN F M, et al. Breast ultra-

- sound lesions recognition: end-to-end deep learning approaches[J]. *Journal of medical imaging*, 2018, 6(1): 011007.
- [29] YU Chaohui, WANG Jindong, CHEN Yiqiang, et al. Transfer learning with dynamic adversarial adaptation network[C]//2019 IEEE International Conference on Data Mining. Beijing: IEEE, 2019: 778–786.
- [30] ZHU Yongchun, ZHUANG Fuzhen, WANG Jindong, et al. Deep subdomain adaptation network for image classification[J]. *IEEE transactions on neural networks and learning systems*, 2021, 32(4): 1713–1722.
- [31] CHEN Chao, CHEN Zhihong, JIANG Boyuan, et al. Joint domain alignment and discriminative feature learning for unsupervised deep domain adaptation[J]. *Proceedings of the AAAI conference on artificial intelligence*, 2019, 33: 3296–3303.
- [32] YE Mang, ZHANG Xu, YUEN P C, et al. Unsupervised embedding learning via invariant and spreading instance feature[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 6203–6212.
- [33] HÉNAFF O J, SRINIVAS A, FAUW J, et al. Data-efficient image recognition with contrastive predictive cod-

ing[C]// Proceedings of the 37th International Conference on Machine Learning. JMLR, 2020, 119: 4182–4192

作者简介:



丁维昌, 硕士研究生, 主要研究方向为深度学习、医学图像处理。



施俊, 教授, 主要研究方向为医学图像分析与处理、模式识别。



王骏, 副教授, 主要研究方向为机器学习、医学影像智能分析。