

基于解剖结构与标签关系建模的胸片多标签分类方法

张志豪, 胡金蓉, 陈浩

(成都信息工程大学 计算机学院, 四川 成都 610200)

摘要: 胸部X光成像是临床胸部疾病筛查与诊断的重要手段,但现有方法往往仅依赖局部特征提取,难以充分利用图像的上下文信息与解剖结构先验,导致识别精度和鲁棒性受限。为此,提出了一种双引导上下文优化模块,融合了解剖掩膜引导的空间注意力与边缘引导的通道注意力,从解剖与结构层面增强特征表达能力。同时,针对现有方法在疾病标签共现关系和语义关联建模方面的不足,构建了可学习的标签共现知识库,用于捕获疾病间的共现关系与语义相似性。实验结果表明,该方法在ChestX-ray14和CheXpert两个公开胸片数据集上分别取得了85.0%和89.6%的AUC分数,验证了所提框架在多标签胸部疾病识别任务中的有效性和泛化性。

关键词: 多标签分类;解剖先验;标签共现;知识库;医学影像

DOI: 10.11907/rjtk.251853

扫描二维码阅读全文: 

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(XXXX)0XX-0001-08

Multi-Label Chest X-ray Classification Method Based on Anatomical Structure and Label Relationship Modeling

ZHANG Zhihao, HU Jinrong, CHEN Hao

(School of Computer Science, Chengdu University of Information Science Technology, Chengdu 610200, China)

Abstract: Chest X-ray imaging served as an essential tool for clinical screening and diagnosis of thoracic diseases. Existing approaches often relied solely on local feature extraction, failing to effectively utilize contextual information and anatomical priors, which limited their diagnostic accuracy and robustness. To address these issues, we proposed a dual-guided context refinement module, which integrated anatomy-guided spatial attention and edge-guided channel attention to enhance feature representation from both anatomical and structural perspectives. Meanwhile, to overcome limitations in modeling label co-occurrence and semantic dependencies, we constructed a learnable label co-occurrence knowledge base to capture co-occurrence relationships and semantic similarities among disease labels. Experimental results demonstrated that the proposed method achieved AUC scores of 85.0% and 89.6% on two public CXR datasets, ChestX-ray14 and CheXpert, respectively. These findings indicate that the proposed framework effectively improves multi-label chest disease recognition and exhibits strong generalization capability.

Key Words: multi-label classification; anatomical prior; label co-occurrence; knowledge base; medical imaging

0 引言

胸部疾病是全球范围内发病率和死亡率最高的健康威胁之一,胸部X光(Chest X-Ray, CXR)因其无创、经济、高效的特点,已成为胸部疾病筛查与诊断的首选影像学工具^[1]。然而,随着检查需求激增与放射科医生短缺矛盾的

加剧,智能辅助诊断技术的需求日益迫切^[2]。在此背景下,基于深度学习的多标签分类模型已成为辅助放射科医生实现高效、精准影像分析的研究热点。

在CXR图像分析任务中,多标签分类需要同时识别图像中可能共存的多种病理表现,对特征表达能力与语义建模能力提出了较高要求。现有基于深度学习的方法在该任务中取得了一定进展,但在复杂病理共存场景下仍面临

收稿日期: 2026-03-19

作者简介: 张志豪(1999-),男,成都信息工程大学计算机学院硕士研究生,研究方向为医学图像处理与自然语言处理;胡金蓉(1983-),女,博士,成都信息工程大学计算机学院副教授,研究方向为人工智能;陈浩(1989-),男,博士,成都信息工程大学计算机学院讲师,研究方向为人工智能、自然语言处理。

挑战。一方面,多数方法主要依赖卷积神经网络自动学习特征,缺乏对胸部解剖结构信息的显式建模,难以有效区分具有相似影像表现的病理区域;另一方面,不同疾病标签之间通常存在明显的共现关系与语义关联,但部分方法对标签关系的建模仍较为简单,限制了模型整体判别能力的提升。

1 相关工作

卷积神经网络(Convolutional Neural Network, CNN)是 CXR 图像多标签分类任务中最常用的基础模型。典型网络结构如 ResNet^[3]和 DenseNet^[4]在医学影像分析中表现稳定,其中 CheXNet^[5]基于 DenseNet-121 在 ChestX-ray14 数据集上取得了具有代表性的分类性能。后续研究是在该框架基础上进行改进,如:Rubin 等^[6]提出的 DualNet 通过融合正位与侧位影像提升病变识别能力;Li 等^[7]提出的 GL-MLL 框架通过联合建模全局与局部标签相关性缓解类别不平衡问题。总体而言,基于 CNN 的方法在局部特征提取方面具有优势,但对多标签任务中标签之间的全局依赖建模能力仍有限。

为增强模型对关键病灶区域的感知能力,部分研究引入注意力机制,通过自适应加权突出判别性区域。Gundel 等^[8]提出的定位感知网络 DNet 同时实现了胸部 X 光图像分类与病灶定位;Ma 等^[9]提出的交叉注意力网络仅依赖图像级标签即可挖掘判别性特征;Wang 等^[10]提出的 A³Net 将多种注意力机制统一嵌入 DenseNet 架构中,在多标签分类任务中取得了较好效果。然而,此类方法多依赖模型内部学习注意力权重,在缺乏结构先验约束的情况下,对复杂病理共存场景的适应性仍有不足。

针对多标签分类中疾病标签之间的语义依赖问题,已有研究尝试通过标签相关性建模提升分类一致性。Chen 等^[11]提出的 CheXGCN 利用图卷积网络刻画疾病标签之间的共现关系;Wang 等^[12]提出的 BB-GCN 通过标签共现嵌入与图结构建模进一步增强了多标签识别能力。这类方法验证了标签共现关系作为有效先验的重要性,但通常依赖静态图结构,难以刻画不同样本或数据分布下标签关系的动态变化。

综上所述,现有 CXR 图像多标签分类方法在特征提取、注意力建模和标签相关性刻画方面取得了一定进展,但在解剖结构先验利用和标签关系动态建模等方面仍存在不足。针对上述问题,本文提出了一种融合解剖结构先验与标签关系建模的 CXR 图像多标签分类方法,以进一步提升模型在复杂病理场景下的分类性能。

2 基于解剖结构与标签关系建模的分类模型

基于解剖结构与标签关系建模的多标签胸部 X 光片

分类模型整体结构如图 1 所示,其主要包括 3 个部分:解剖分割部分、特征提取部分和知识增强校正部分。

输入的胸部 X 光图像首先经过解剖分割部分,提取关键解剖区域并生成解剖掩码,为后续特征学习提供空间先验信息。随后,将原始图像与解剖掩码共同输入特征提取部分,其中嵌入的双引导上下文优化(Dual-Guided Context Refinement, DGCR)模块利用解剖先验进行空间注意力引导,并结合边缘特征增强机制,获得融合解剖信息与结构特征的多尺度表征。接着,提取的特征被输入知识增强校正部分,基于疾病标签之间的共现关系和语义关联信息,对特征表示和预测结果进行自适应优化。最终,经校正后的结果通过 Sigmoid 激活函数处理,输出各疾病标签的预测概率,实现多标签胸部疾病的分类识别。

2.1 解剖分割

为引导模型聚焦于与疾病相关的关键解剖区域,本文研究在整体框架中引入解剖分割部分。该部分采用基于 U-Net^[13]的网络结构,并在 JSRT^[14]数据集上进行预训练,该数据集提供了包括左/右肺、心脏及背景在内的精细像素级标注。与病灶级精细分割不同,该部分的主要作用在于对胸部主要解剖结构进行像素级解析,以获得稳定的解剖区域先验知识(例如肺野区域与心影区域等)。并且,该先验仅对特征响应进行重加权并与原始特征进行融合,而非对特征进行硬裁剪。

给定输入图像 $I \in \mathbb{R}^{H \times W \times C}$, 预训练分割网络 $UNet(\cdot)$ 生成包含多个解剖类别(如背景、心脏、左肺、右肺)的概率图。为了获得对疾病诊断具有普适性的感兴趣区域,并减少下游模块的计算冗余,本文使用了语义类别聚合策略。具体而言,首先通过逐像素 Argmax 操作确定每个像素所属的解剖类别索引,随后构建二值解剖掩模 $M_{anatomy} \in \mathbb{R}^{H \times W \times 1}$, 如式(1)所示。在该掩模中,将所有属于肺部类与心脏类的像素定义为前景(赋值为 1),将背景类像素定义为背景(赋值为 0)。

$$M_{anatomy} = \operatorname{argmax}(UNet(I)) \quad (1)$$

得到的二值解剖掩模 $M_{anatomy}$ 经过双线性插值下采样至与各层特征图相同的空间分辨率后,作为解剖空间先验被输入主干特征提取网络,用于在 DGCR 模块中引导空间注意力机制向关键病灶区域倾斜。

2.2 特征提取

本文研究采用 DenseNet-121 作为骨干网络,以充分利用其特征复用优势。然而,标准 DenseNet 缺乏对胸部解剖结构的显式建模,因此在每个密集块前引入 DGCR 模块,实现解剖引导的空间注意、边缘增强的通道注意以及全局上下文融合。输入 CXR 图像 $I \in \mathbb{R}^{H \times W \times C}$ 经初始卷积与池化后得到特征图 $F_0 \in \mathbb{R}^{C_0 \times H_0 \times W_0}$ 。随后,网络经过 4 个阶段,每个阶段由 1 个 DGCR 模块与 1 个密集块组成。第 i 阶段的输入特征 $F_{in}^{(i)}$ 依次经 DGCR 模块与密集块处理后,输出特征 $F_{out}^{(i)}$ 可表示为如式(2)所示。

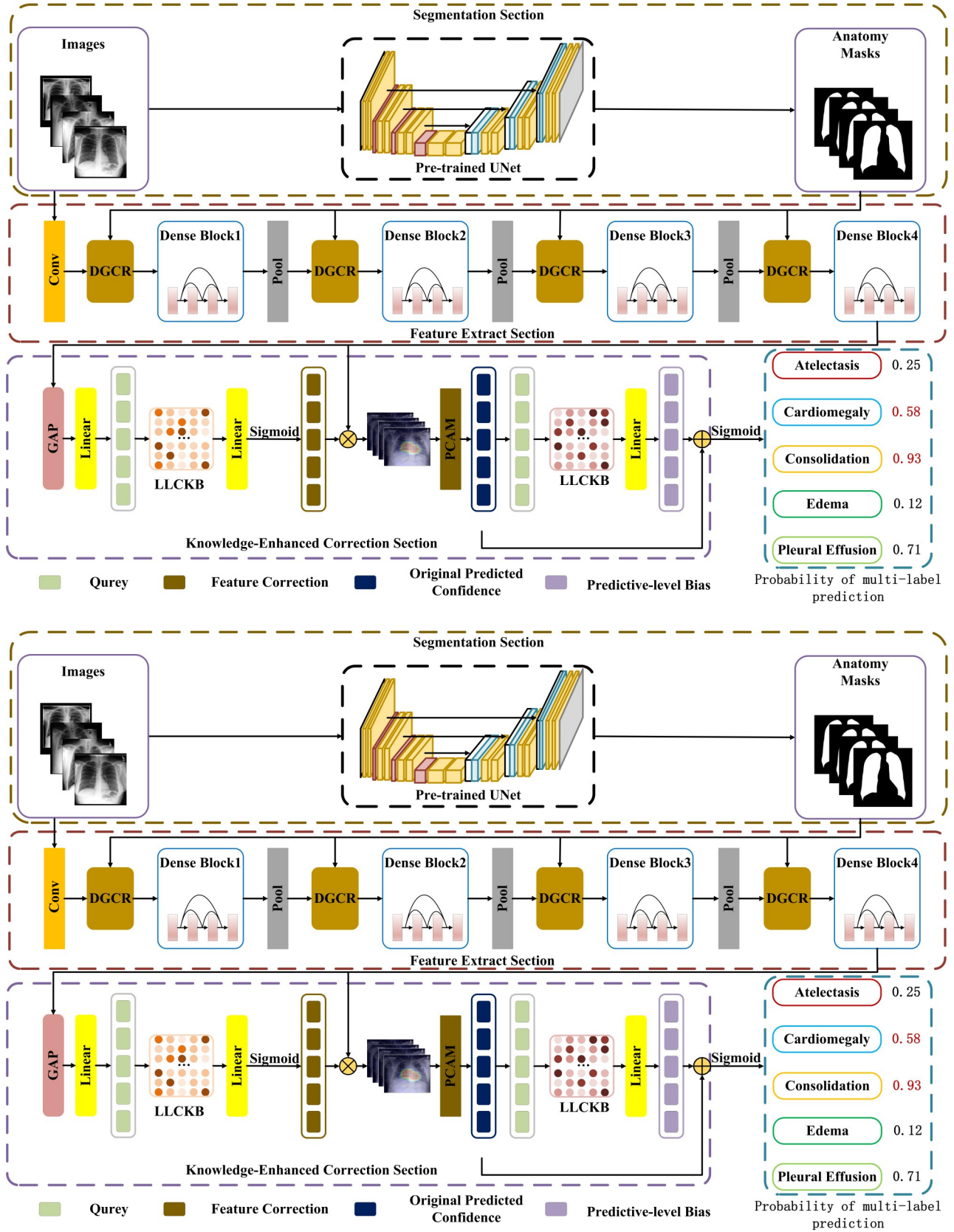


Fig. 1 Overall framework of the classification model based on anatomical structure and label relationship modeling

图 1 基于解剖结构与标签关系建模的分类模型整体框架

$$F_{out}^{(i)} = DenseBlock^{(i)}(DGCR^{(i)}(F_{in}^{(i)})) \quad (2)$$

其中, $DGCR^{(i)}(.)$ 与 $DenseBlock^{(i)}(.)$ 分别表示第 i 阶段

的 DGCR 模块和密集块。

如图 2 所示, DGCR 模块通过引入解剖先验与边缘增

强机制,对输入特征进行层次化优化。整个过程包括3个阶段:解剖引导的空间注意、边缘增强的通道注意和全局

上下文建模。

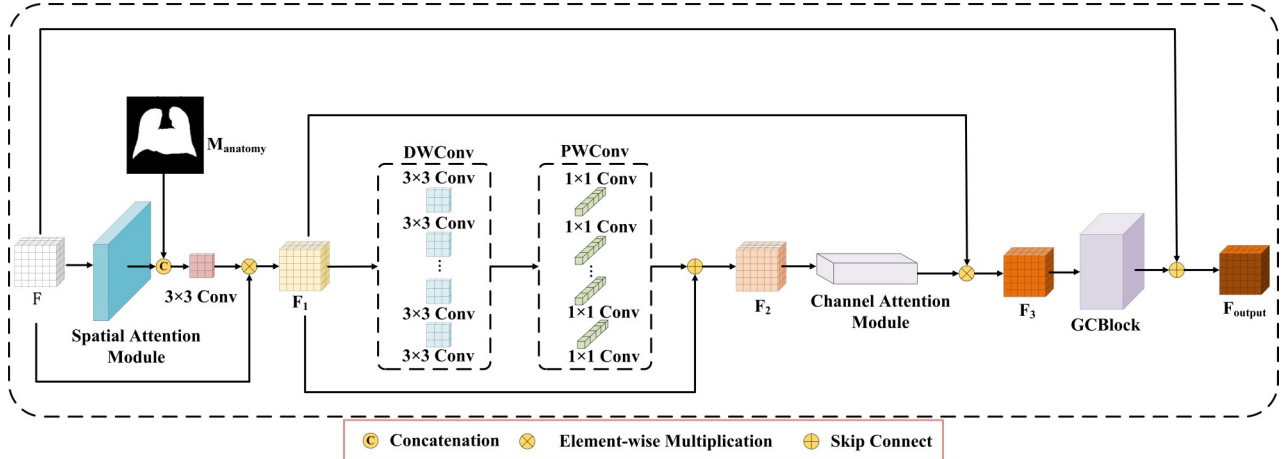


Fig. 2 Framework of the dual-guided context refinement module

图2 双引导上下文优化模块框架

首先,利用上述解剖分割部分生成的解剖掩膜作为空间先验信息,对注意力的空间分布进行引导,从而增强与疾病相关区域的特征响应。给定输入特征 F ,空间注意模块根据视觉特征自适应生成原始视觉注意力图,并将其与下采样后的二值解剖掩膜在通道维度进行拼接,随后通过可学习的卷积层进行融合,从而生成综合考虑视觉特征与解剖先验的空间权重图。将该空间权重图与输入特征逐元素相乘,得到融合了解剖结构信息的特征表示 F_1 ,通过软权重方式调节不同空间位置的特征响应强度,在抑制背景噪声的同时,保持对潜在病灶区域特征的充分表达能力。

随后,模块通过深度可分离卷积提取局部边缘与纹理特征。 F_1 首先经过逐深度卷积(Depthwise Convolution, DWConv)以捕捉细粒度的边缘变化,再通过逐点卷积(Pointwise Convolution, PWConv)实现通道融合。该部分设计在保持丰富特征表达的同时降低参数量,并通过残差连接提升特征传递的连续性与稳定性。所得特征 F_2 进一步输入通道注意力模块进行加权校准,以自适应地增强与疾病判别相关的特征通道,生成更具判别力的特征表示 F_3 。

最后,为捕捉图像中更高层次的语义依赖关系,DGCR模块引入全局上下文模块^[15](Global Context Block, GCBlock)进行长程特征建模。该模块通过全局特征聚合与通道重分配实现语义一致性增强。最终,输出特征与初始输入特征通过残差连接融合,形成DGCR模块的输出结果 F_{output} 。

虽然DGCR通过解剖与边缘引导有效提升了病变相关区域的视觉特征表示,但其主要作用仍集中于视觉层的优化。为了进一步融入高层语义推理,在框架中引入可学习标签共现知识库(Learnable Label Co-occurrence Knowledge Base, LLCKB)模块,该模块在特征层与预测层进行标签感知的上下文校准,从而对DGCR模块形成互补。

2.3 知识增强校正

在多标签胸部X光图像分类任务中,不同疾病标签之间通常存在复杂的语义依赖与共现关系。单纯依赖视觉特征进行独立分类,往往难以充分刻画疾病之间的潜在关联。为此,本文设计了一个以LLCKB模块为核心的知识增强校正部分,以在模型中显式建模标签层面的先验语义信息。考虑到语义信息在模型不同阶段所发挥的作用存在差异,本文从特征表示阶段和预测决策阶段两个层面引入LLCKB,对视觉特征和分类输出进行分阶段的语义增强与校正。

在特征表示阶段,特征提取部分输出的特征经过全局平均池化与线性投影处理,用于从LLCKB中检索标签间的共现关系与语义关联信息。LLCKB模块在融合这些知识后,通过线性变换和Sigmoid激活生成特征级的校正信息,并与原始特征逐元素融合,得到知识增强后的特征表示,从而强化与疾病标签相关的关键特征,同时抑制与目标标签无关的噪声。

知识增强后的特征被输入到概率类激活图(Probabilistic Class Activation Map, PCAM)池化模块,该模块通过加权激活图生成类别相关的空间响应,并进行特征聚合,最终得到初步预测结果。为进一步刻画疾病标签之间的共现关系与依赖结构,本文在预测决策阶段再次引入LLCKB。具体而言,初步预测结果经过线性变换后再次输入LLCKB模块,以生成预测级的偏置信息。该偏置与原始预测结果进行残差融合,最后通过Sigmoid激活得到最终的多标签预测结果。

疾病标签之间的共现关系与语义关联在辅助多标签分类中有重大意义,现有研究通过使用图神经网络将标签共现关系应用于分类中,但是静态的图结构难以刻画不同样本或数据分布下标签关系的动态变化^[12]。为此,本文设计了LLCKB模块以动态地建模疾病标签间的共现关系与

语义依赖, 为多标签胸部 X 光片分类提供可学习的知识增强支撑, 如图 3 所示。该模块本质上是一个可学习的记忆矩阵, 在训练阶段, 记忆矩阵通过动态的学习标签共现关

系和语义依赖以更新其所存储的内容; 在推理阶段, 根据输入的特征从记忆矩阵中检索相关信息, 以辅助特征表达与预测结果优化。

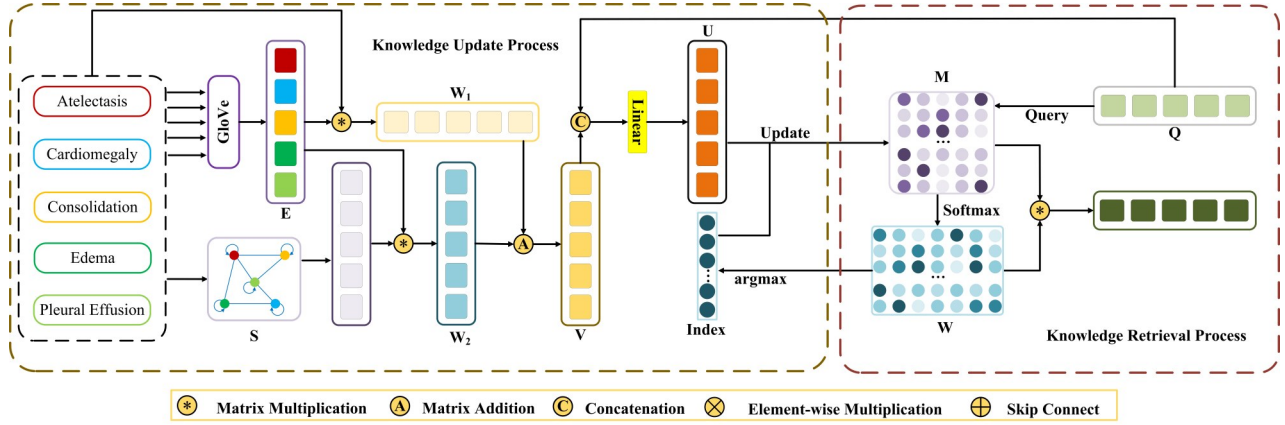


Fig. 3 Framework of the learnable label co-occurrence knowledge base module

图 3 可学习标签共现知识库模块框架

在训练阶段, LLCKB 通过融合标签共现关系与语义表示, 不断更新知识存储以积累任务相关知识。具体而言, 模块首先基于当前批次标签信息构建全局共现矩阵 S , 并通过指数移动平均 (Exponential Moving Average, EMA) 策略进行平滑更新, 以平衡历史统计与新样本信息。随后, 利用预训练的 GloVe 模型获取标签语义嵌入表示 E , 为每个疾病标签提供高维语义特征。根据当前样本的标签信息, 模块通过加权聚合获得加权语义嵌入 W_1 , 用于表征样本中显式的语义信息。同时, 结合 S 计算的标签依赖关系, 生成共现增强语义嵌入 W_2 , 以补充标签间的隐式语义关联。通过综合显式语义与隐式共现信息, 模块得到融合后的语义表示 V 。该语义表示与特征提取阶段生成的查询信息 Q 拼接后输入线性变换层, 以生成新的知识内容 U 。最后, 根据注意力权重分布得到最相关的知识记忆槽索引向量 $Index$, 并通过 EMA 策略对知识记忆槽进行渐进式更新, 从而保持知识库对任务语义的动态适应性。其更新过程可表示为如式 (3) 所示。

$$M_t = (1 - \alpha)M_{t-1} + \alpha U_t \quad (3)$$

其中, M_t 表示当前时间步的知识存储内容, M_{t-1} 为历史记忆, U_t 为当前批次生成的新知识, $\alpha \in (0, 1)$ 为平衡系数, 用于控制历史知识与新知识的融合比例。

在推理阶段, LLCKB 根据输入特征生成查询信息, 并与知识库中存储的语义与共现信息进行匹配。通过计算相似度并基于注意力机制分配权重, 模块自适应地聚合与输入最相关的知识内容, 生成知识增强后的输出表示。该表示在后续特征与预测层中被用于优化模型的判别性与一致性, 从而实现基于知识驱动的多标签分类推理。

2.4 损失函数

针对胸部 X 光图像的多标签分类任务, 本文采用二元交叉熵 (Binary Cross-Entropy, BCE) 损失函数对模型进行优化。对于每一幅输入图像, 其对应的多标签预测结果可

表示为 $\hat{y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_C]$, 其中 C 表示疾病类别数, $\hat{y}_c \in [0, 1]$ 为模型对第 c 类疾病存在概率的预测值。

给定真实标签 $y = [y_1, y_2, \dots, y_C]$, 其中 $y_c \in \{0, 1\}$, 二元交叉熵损失函数定义为如式 (4) 所示。

$$\mathcal{L}_{BCE} = -\frac{1}{C} \sum_{c=1}^C [y_c \log(\hat{y}_c) + (1 - y_c) \log(1 - \hat{y}_c)] \quad (4)$$

该损失函数将多标签分类问题视为多个相互独立的二分类任务, 能够有效处理 CXR 图像中疾病共存的情况, 并已广泛应用于多标签医学图像分析。

3 实验与结果分析

3.1 实验数据与设置

本文在两个公开的胸部 X 光图像数据集上进行了实验评估。第一个数据集为 CheXpert^[16], 由斯坦福大学发布, 包含 224 316 张胸部 X 光片, 每张图像均标注 14 类病理观察结果 (标签值 1、0、-1 分别表示阳性、阴性与不确定) 如图 4 所示 (不确定标签以蓝色标示)。第二个数据集为 ChestX-ray14^[17], 由美国国立卫生研究院发布, 包含 112 120 张正面视角胸部 X 光图像, 每张图像标注 14 种肺部疾病或“无异常”类别。本文在两个数据集中分别选取了临床中较为常见且发生概率较高的 5 类病理进行实验。

所有图像在输入网络前均统一采样并随机裁剪至 512×512 分辨率, 并施加随机仿射变换 (旋转、平移、缩放等) 以增强数据多样性。对于 CheXpert 数据集中存在的不确定标签, 本文采用 U-ones 与 U-zeros 策略进行处理^[16]: 将“水肿 (Edema)”和“肺不张 (Atelectasis)”视为正样本 (U-ones), 将“心脏肥大 (Cardiomegaly)”、“实变 (Consolidation)”和“胸腔积液 (Pleural Effusion)”视为负样本 (U-zeros)。ChestX-ray14 数据集中不存在不确定标签。

在训练过程中对验证集性能变化进行持续监控发现,

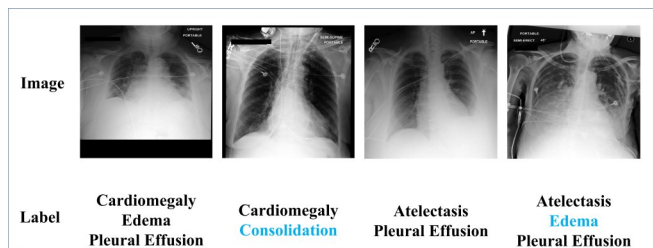


Fig. 4 Input chest X-ray images from the CheXpert dataset and their corresponding ground-truth pathology labels

图4 来自CheXpert数据集的输入胸部X光图像及其对应的真实病理标签

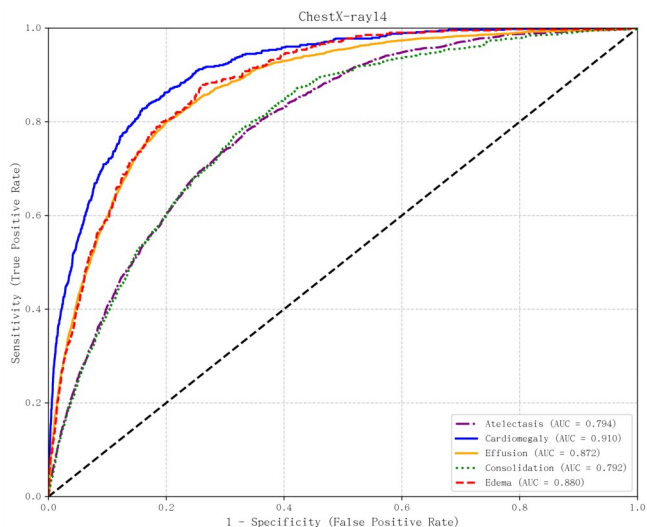
对于CheXpert和ChestX-ray14这两个数据集,模型分别经过6和8个epoch后性能提升趋于平缓,继续增加训练轮次难以带来显著性能提升。因此,本文选8和10个epoch作为最终训练轮次,批大小设为16。尽管训练轮次相对较少,但由于数据集规模较大,模型在两个数据集上分别完成了约13万次和5万次参数更新,为特征学习提供了充足的梯度更新机会。优化器采用Adam,将初始学习率设为0.0001,并采用了OneCycleLR学习率调度策略(最大学习率0.00015,上升比例为0.25),以加速模型收敛过程。同时,在推理阶段,为了从概率输出中确定最终疾病标签,采用0.5作为分类判定阈值。具体而言,当某标签的预测概率大于或等于0.5时,判定该疾病真实存在,反之则判定为不存在,从而将连续的概率值转化为明确的二值化多标签预测结果。

3.2 对比实验

考虑到胸部疾病标签分布存在明显的不平衡,本文采用AUC(Area Under the ROC Curve, ROC曲线下面积)作为评估指标,用以综合反映模型分类性能。表1和表2给出了所提方法与各对比模型的实验结果。为保证对比实验的公平性与可复现性,本文所采用的对比方法均在统一的实验设置下进行实现与评估。对于已公开实现代码的方法,优先采用其官方开源版本;对于未提供完整代码的工作,则严格依据原论文中给出的网络结构与训练策略进行复现。所有模型均使用相同的数据集划分、输入图像分辨率及评估指标进行训练与测试。对于原文中未明确给出的超参数设置,本文采用与当前主流胸部X光多标签分类研究一致的配置,以避免人为偏置。

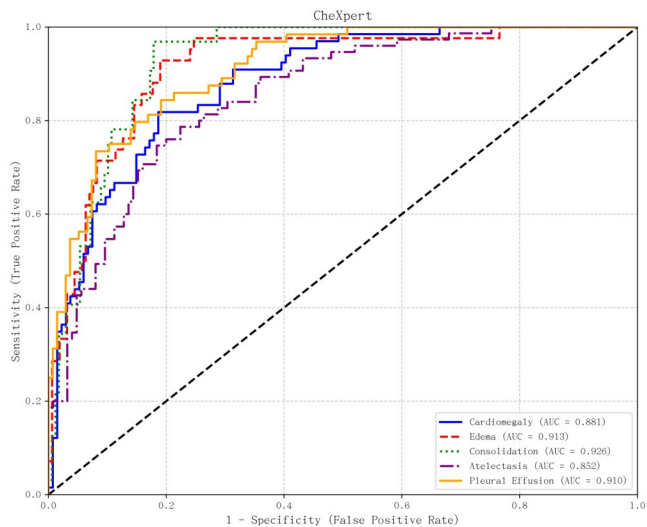
表1给出了本文方法与7种现有方法在ChestX-ray14数据集上的比较结果。本文模型在5类疾病上平均AUC为0.850,整体性能优于对比方法。其中,“Cardiomegaly”、“Edema”、“Consolidation”、“Atelectasis”和“Effusion”均取得最高AUC。图5(a)展示了5类病理的ROC曲线,其中“Cardiomegaly”最接近左上角,表现最佳;“Consolidation”和“Atelectasis”略低,但仍高于其他方法。

表2展示了在CheXpert数据集上7种方法的分类性能比较。结果表明,本文方法在5种疾病的平均AUC达到0.896,优于所有现有方法。其中,“Edema”、“Consolidation”



(a) ROC curves on the ChestX-ray14 datasets

(a) ChestX-ray14数据集上的ROC曲线



(b) ROC curves on the CheXpert datasets

(b) CheXpert数据集上的ROC曲线

Fig. 5 ROC curves of the proposed method on the ChestX-ray14 and CheXpert datasets

图5 本文方法在ChestX-ray14和CheXpert数据集上的ROC曲线

Table 1 Comparative experimental results on the ChestX-ray14 dataset

表1 在ChestX-ray14数据集上的对比实验结果

Method	Cardiomegaly	Edema	Consolidation	Atelectasis	Effusion	Mean
Xu et al[18]	0.806	0.833	0.772	0.751	0.860	0.804
MPC[19]	0.885	0.829	0.775	0.731	0.810	0.806
MAE[20]	0.876	0.862	0.745	0.770	0.824	0.815
BB-GCN[12]	0.895	0.855	0.759	0.779	0.836	0.825
PCAN[21]	0.887	0.854	0.759	0.791	0.841	0.826
ML-LGL[22]	0.904	0.866	0.771	0.782	0.835	0.832
PCAM[23]	0.864	0.833	0.732	0.772	0.825	0.805
ours	0.910	0.880	0.792	0.794	0.872	0.850

和“Pleural Effusion”的AUC均超过0.90;“Cardiomegaly”和“Atelectasis”也取得最优结果,分别为0.881和0.852。图5

(b)展示了5类病理的ROC曲线,其中“Consolidation”的曲线最接近左上角,显示最高的分类性能;“Atelectasis”相对较低,但仍优于其他方法的平均水平。

Table 2 Comparative experimental results on the CheXpert dataset
表 2 在 CheXpert 数据集上的对比实验结果

Method	Cardiomegaly	Edema	Consolidation	Atelectasis	Pleural Effusion	Mean
Xu et al[18]	0.796	0.797	0.606	0.651	0.819	0.734
BB-GCN [12]	0.866	0.858	0.720	0.732	0.877	0.811
MPC[19]	0.786	0.913	0.908	0.822	0.927	0.871
MAE[20]	0.776	0.923	0.894	0.815	0.920	0.887
ML-LGL [22]	0.853	0.912	0.943	0.811	0.928	0.889
PCAN[21]	0.859	0.906	0.911	0.840	0.937	0.891
PCAM[23]	0.853	0.905	0.916	0.825	0.896	0.879
ours	0.881	0.913	0.926	0.852	0.910	0.896

3.3 可视化分析

为进一步验证模型判别区域与病理特征的一致性,本文采用 Grad-CAM^[24]对 CXR 图像生成类别相关的注意力图,如图 6 所示。其中,Ground Truth 是官方提供的手动病变区域,Heat Map 是用 Grad-CAM 生成的热力图,Score 是本文模型预测结果。真实病理情况用红色突出显示,红色与黄色区域代表高关注度,蓝色与绿色表示低关注度。模型生成的高响应区域与真实病灶位置高度一致,同时其最高预测得分的病理标签与标注结果相符。当图像中存在多种病理时,模型对不同病理的响应强度接近,反映其多标签识别能力。

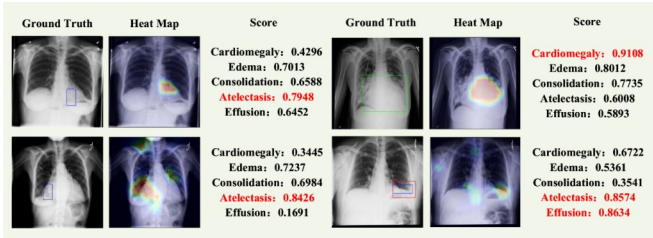


Fig. 6 Lesion region localization by the proposed model on the ChestX-ray14 dataset

图 6 所提模型在 ChestX-Ray14 数据集上的病变区域定位

3.4 消融实验

如表 3 所示,通过消融实验系统评估了 DGCR 模块与 LLCKB 模块对模型性能的影响。以文献[23]中的网络架构作为基线(Baseline),分别在其上独立引入 DGCR 模块与 LLCKB 模块,并进一步考察两者结合的效果。

实验结果表明,引入 DGCR 模块后,模型在 AUC 指标上显著提升,说明该模块通过引入解剖与边缘双引导的上下文优化,有效增强了特征的判别性与空间聚焦能力;而引入 LLCKB 模块后,模型的 AUC 同样得到明显改善,验证了标签共现知识对特征语义和预测一致性的增强作用。当同时集成 DGCR 与 LLCKB 模块时,模型表现出最高的 AUC 值,表明两者在特征表达与语义推理层面具有互补

性,能够协同提升多标签疾病分类的整体判别性能。

Table 3 Ablation experimental results of the DGCR module and LLCKB module on the ChestX-ray14 dataset

表 3 在 ChestX-ray14 数据集上 DGCR 模块与 LLCKB 模块的消融实验结果

Method	Cardiomegaly	Edema	Consolidation	Atelectasis	Effusion	Mean
Baseline	0.864	0.833	0.732	0.772	0.825	0.805
Baseline+DGCR	0.892	0.864	0.783	0.779	0.863	0.836
Baseline+LLCKB	0.899	0.873	0.788	0.782	0.868	0.842
Baseline+DGCR+LLCKB	0.910	0.880	0.792	0.794	0.872	0.850

同时,表 4 给出了 LLCKB 在不同使用阶段下的消融实验结果(其中,LLCKB1 表示在特征表示阶段引入 LLCKB,LLCKB2 表示在预测阶段引入 LLCKB)。可以观察到,仅在特征表示阶段引入 LLCKB 即可在所有疾病类别上获得显著性能提升,说明标签共现与语义知识在特征层面对视觉表示具有较强的指导作用。相比之下,仅在预测阶段使用 LLCKB 虽然也能带来一定改善,但整体提升幅度相对有限,表明在缺乏充分语义增强特征的情况下,预测级校正作用受到一定约束。进一步地,当在特征阶段与预测阶段同时引入 LLCKB 时,模型取得了最优整体性能,验证了两阶段知识增强在功能上的互补性。

Table 4 Ablation results of introducing the LLCKB module at different stages on the ChestX-ray14 dataset

表 4 在 ChestX-ray14 数据集上不同阶段引入 LLCKB 模块的消融实验结果

Method	Cardiomegaly	Edema	Consolidation	Atelectasis	Effusion	Mean
Baseline	0.864	0.833	0.732	0.772	0.825	0.805
Baseline+LLCKB1	0.887	0.860	0.782	0.781	0.845	0.831
Baseline+LLCKB2	0.874	0.846	0.784	0.774	0.842	0.817
Baseline+LLCKB	0.899	0.873	0.788	0.782	0.868	0.842

4 结语

本文提出了一种融合解剖结构与标签共现关系的胸部 X 光图像多标签分类方法。与现有仅依赖图像内容或基于模型内部注意力机制的方法不同,本方法显式引入解剖先验信息引导特征提取,并利用标签间语义关联对特征表示与预测结果进行协同优化。这一机制模拟了放射科医生的诊断逻辑:首先聚焦于与疾病相关的解剖区域,再结合临床知识对共现疾病进行联合推理。实验结果表明,本文方法在两个公开 CXR 数据集上均取得领先性能:在 ChestX-ray14 和 CheXpert 数据集上的平均 AUC 分别达到 85.0% 和 89.6%,验证了所提出模块在提升模型判别能力与临床可解释性方面的有效性。未来,将进一步探索该方法在更多医学影像任务中的迁移能力,并将其拓展至临床辅助诊断的实际应用中。

参考文献:

[1] RAOOF S, FEIGIN D, SUNG A, et al. Interpretation of plain chest roent-

- genogram[J]. Chest, 2012, 141(2): 545-558.
- [2] YU P, XU H, ZHU Y, et al. An automatic computer-aided detection scheme for pneumoconiosis on digital chest radiographs[J]. Journal of Digital Imaging, 2011, 24(3): 382-393.
- [3] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [4] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [5] RAJPURKAR P, IRVIN J, ZHU K, et al. CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning[DB/OL]. <https://arxiv.org/abs/1711.05225>.
- [6] RUBIN J, SANGHAVI D, ZHAO C, et al. Large scale automated reading of frontal and lateral chest X-Rays using dual convolutional neural networks[DB/OL]. <https://arxiv.org/abs/1804.07839>.
- [7] LI L, CAO P, YANG J, et al. Modeling global and local label correlation with graph convolutional networks for multi-label chest X-ray image classification[J]. Medical & Biological Engineering & Computing, 2022, 60(9): 2567-2588.
- [8] GUENDEL S, GRBIC S, GEORGESCU B, et al. Learning to recognize abnormalities in chest x-rays with location-aware dense networks[C]//Iberoamerican Congress on Pattern Recognition, 2018: 757-765.
- [9] MA C, WANG H, HOI S C H. Multi-label thoracic disease image classification with cross-attention networks[C]//International Conference on Medical Image Computing and Computer-assisted Intervention, 2019: 730-738.
- [10] WANG H, WANG S, QIN Z, et al. Triple attention learning for classification of 14 thoracic diseases using chest radiography[J]. Medical Image Analysis, 2021, 67: 101846.
- [11] CHEN B, LI J, LU G, et al. Label co-occurrence learning with graph convolutional networks for multi-label chest x-ray image classification[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(8): 2292-2302.
- [12] WANG G, WANG P, CONG J, et al. BB-GCN: a bi-modal bridged graph convolutional network for multi-label chest X-Ray recognition[DB/OL]. <https://arxiv.org/abs/2302.11082>.
- [13] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-assisted Intervention, 2015: 234-241.
- [14] SHIRAIISHI J, KATSURAGAWA S, IKEZOE J, et al. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules[J]. American Journal of Roentgenology, 2000, 174(1): 71-74.
- [15] CAO Y, XU J, LIN S, et al. GCNET: non-local networks meet squeeze-excitation networks and beyond[C]//Seoul: The IEEE/CVF International Conference on Computer Vision Workshops, 2019.
- [16] IRVIN J, RAJPURKAR P, KO M, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019: 590-597.
- [17] WANG X, PENG Y, LU L, et al. ChestX-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2097-2106.
- [18] XU L, NI Z, SUN H, et al. A foundation model for generalizable disease diagnosis in chest X-ray images[DB/OL]. <https://arxiv.org/abs/2410.08861>.
- [19] XIAO J, LI S, LIN T, et al. Multi-label chest X-ray image classification with single positive labels[J]. IEEE Transactions on Medical Imaging, 2024, 43(12): 4404-4418.
- [20] XIAO J, BAI Y, YUILLE A, et al. Delving into masked autoencoders for multi-label thorax disease classification[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023: 3588-3600.
- [21] ZHU X, PANG S, ZHANG X, et al. PCAN: Pixel-wise classification and attention network for thoracic disease classification and weakly supervised localization[J]. Computerized Medical Imaging and Graphics, 2022, 102: 102137.
- [22] LIU Z, CHENG Y, TAMURA S. Multi-label local to global learning: a novel learning paradigm for chest x-ray abnormality classification[J]. IEEE Journal of Biomedical and Health Informatics, 2023, 27(9): 4409-4420.
- [23] YE W, YAO J, XUE H, et al. Weakly supervised lesion localization with probabilistic-cam pooling[DB/OL]. <https://arxiv.org/abs/2005.14480>.
- [24] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 618-626.

(责任编辑:孙娟)