DOI:10.3969/j.issn.1005-202X.2025.03.012

医学影像物理

基于沙漏阶梯残差模型的胸部影像多标签分类

方俊泽,邢素霞,郭正,李珂娴,王瑜
北京工商大学人工智能学院,北京 100048

【摘要】提出一种基于沙漏阶梯残差模型(SLRN),用于胸部影像疾病的多标签分类,提高临床诊断的准确性。SLRN的设计包括3个关键模块,首先采用沙漏卷积模块同时提取通道间信息与空间信息;然后使用阶梯自注意力模块,通过移位操作实现不同窗口划分,扩大感受野,提取并融合多尺度特征;在多标签分类阶段,使用多头残差注意力,捕捉到不同标签之间的相关性和特征间的重要性,通过调整不同特征的权重实现更精准的分类。本研究在印第安纳大学收集的胸部X光数据集(IU X-Ray)和美国国立卫生研究院收集并公开的胸部X射线数据集(Chest X-Ray14)中进行验证,实验证明SLRN结合了卷积神经网络和视觉转换器的优点,可以捕捉影像中的局部特征和全局关联,更好地处理长距离依赖关系,辅助医生进行临床诊断。

【关键词】胸部影像;多标签分类;卷积神经网络;视觉转换器;沙漏卷积
【中图分类号】R318;TP391.41 【文献标志码】A 【文章编号】1005-202X(2025)03-0360-09

Multi-label chest X-ray classification using sandglass ladder residual network

FANG Junze, XING Suxia, GUO Zheng, LI Kexian, WANG Yu
School of Artificial Intelligence, Beijng Technology and Business University, Beijng 100048, China

Abstract: A sandglass ladder residual network (SLRN) is proposed for multi-label chest X-ray classification, thereby improving the accuracy of clinical diagnosis. SLRN consists of 3 key modules: (1) a sandglass convolutional module to simultaneously extract inter-channel and spatial information; (2) a ladder self attention block to achieve different window divisions through shift operations, expand the receptive field, and realize multi-scale feature extraction and fusion; (3) class specific residual attention in the multi-label classification stage to capture the correlation between different labels and the importance of features for accomplishing more accurate classification by adjusting the weights of different features. The proposed model is validated using the IU X-Ray dataset collected by Indiana University and the publicly available Chest X-Ray14 dataset collected by the National Institutes of Health in the United States; and the results demonstrate that SLRN which combines the advantages of convolutional neural network and vision transformer enables the capture of local features and global correlations in images, better handles long-distance dependencies, and assists doctors in clinical diagnosis.

Keywords: chest X-ray; multi-label classification; convolutional neural network; vision transformer; sandglass convolution

前 言

胸部疾病因其多样性和高发病率对健康构成严重威胁,及时准确的诊断对患者具有重要意义。然而,一张胸部影像涵盖多种人体组织,各组织出现疾病的病理特征各不同,且往往有一定的关联性,给疾病的诊断工作带来巨大挑战,时有误诊和漏诊的发生[1-3]。随着深度学习的兴起,基于医学影像的计算机辅助诊断技术快速发展,可以有效提升医生诊断的准确率,减轻工作负担[4-6]。

卷积神经网络(Convolutional Neural Network, CNN)在处理图像数据时擅长捕捉局部特征,被广泛应用于解决胸部影像疾病分类任务,可以准确提取医学影像中的细节信息[7-9]。Cheng等[10]提出一种基于特征融合的CNN肺部疾病分类方法,并结合图像增强技术,提高肺部疾病的分类准确性。Chen等[11]采用图神经网络捕捉不同疾病之间的语义关联性,将图像嵌入向量映射到语义相似性图中,改善胸部X光影像的多标签分类性能。Jin等[12]将CNN和语义向量相结合,提出双加权度量损失函数,从图像级别

【收稿日期】2024-12-11
【基金项目】北京市自然科学基金(KZ202110011015)
【作者简介】方俊泽,硕士研究生,研究方向:医学图像处理、深度学习,E-mail: 2605574541@qq.com
【通信作者】邢素霞,博士,副教授,研究方向:医学图像处理、信息融合,E-mail: xingsuxia@163.com

和疾病类别分别考虑图像和标签之间的度量关系，使得14种疾病分类的平均分类准确率达到0.826。然而，CNN在一次卷积操作中只能捕捉到图像中的局部特征，对于大范围或者长距离依赖关系的病变，无法很好地捕捉到相关信息，难以对病变进行全面理解，进而影响模型对于疾病的准确判断。

视觉转换器（Vision Transformer）通过自注意力机制实现对输入序列的编码和表示学习，能捕捉序列中不同位置之间的依赖关系，在全局特征提取方面具有明显优势，越来越多的研究将Vision Transformer应用于不同的医学影像任务[13-14]，例如肿瘤与周围组织的关系、血管与组织的关联等，并且能较好地处理医学影像中的长距离依赖关系[15-16]。Xie等[17]首次将Transformer应用于三维医学图像分割，弥补卷积网络在处理局部性和权重共享时可能存在的归纳偏差。Wu等[18]提出一种采用Transformer辅助CNN进行特征提取的特征自适应网络，能有效捕捉长距离依赖和全局信息，对皮肤损伤进行精准分割。Tao等[19]提出一种Transformer网络以解决脊柱CT中椎骨的自动检测和定位问题，解释不同椎骨之间的关系，并行输出所有椎骨的全局信息。以上研究成果表明Vision Transformer在医学影像任务中表现出色，能有效提取全局信息，并处理长距离依赖关系。

本研究提出一种沙漏阶梯残差模型（Sandglass Ladder Residual Network, SLRN），该模型充分结合CNN和Vision Transformer在特征提取上的优势，同时关注并提取病灶的局部和全局信息，增强图像特征提取能力，为计算机辅助诊断提供更全面的信息。

1 SLRN模型整体结构

SLRN模型结构如图1所示，包括特征提取和多标签分类两部分。特征提取模块通过沙漏卷积模块和阶梯自注意力模块提取医学影像中的影像特征，多标签分类模块将影像特征输入多头残差注意力模块（Class Specific Residual Attention, CSRA），重点关注特征标签之间的关联性并生成对应标签概率[20-21]。

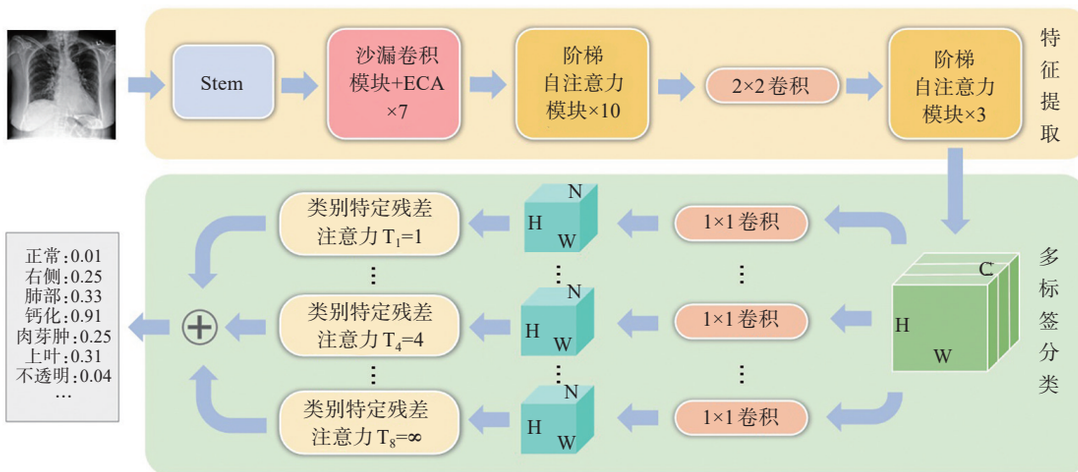


图1 SLRN模型整体结构图

Figure 1 Overall architecture of sandglass ladder residual network

1.1 基于沙漏阶梯的特征提取模块

特征提取模块由沙漏卷积模块和阶梯自注意力模块组成。沙漏卷积模块中的深度卷积和逐点卷积在局部特征提取方面发挥作用，而阶梯自注意力模块通过多尺度融合和移位操作实现全局信息的捕捉和长距离依赖的有效处理，使SLRN能更全面地理解图象中的特征和结构，提高模型对复杂病变的识别性能。特征提取模块的结构如表1所示，其中，“沙漏卷积模块+ECA↓”表示步长为2的下采样操作，ECA（Efficient Channel Attention）为高效通道注意力模块。

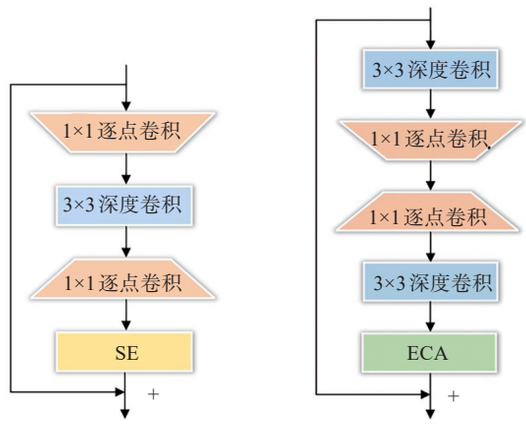
1.1.1 沙漏卷积模块　沙漏卷积模块参考倒残差模块的设计思想，使用逐点卷积和深度卷积分别提取通道间信息和空间信息，并在模块末端引入ECA代替挤压激励模块（Squeeze-and-Excitation, SE），以增强局部特征的提取能力[22-23]。两种卷积模块结构对比如图2所示。

沙漏卷积模块采用逐点卷积编码通道间信息，使用深度卷积编码空间信息。为避免逐点卷积可能导致部分空间信息丢失，从而影响特征的有效提取，将深度卷积置于逐点卷积之外，使其可以对更丰富的空间信息进行编码。同时，在残差路径的末端再引入一层深度卷积，鼓励模型学习到更丰富的空间

表1 特征提取模块结构

Table 1 Structure of feature extraction module

阶段	输入	模块	输出	步长
Stem	$224^2\times3$	3×3卷积↓	36	2
	$112^2\times36$	3×3卷积×2	36	1
沙漏卷积模块+ECA ×7	$112^2\times36$	沙漏卷积模块+ECA↓	72	2
	$56^2\times72$	沙漏卷积模块+ECA ×2	72	1
	$56^2\times72$	沙漏卷积模块+ECA↓	144	2
	$28^2\times144$	沙漏卷积模块+ECA ×2	144	1
	$28^2\times144$	沙漏卷积模块+ECA↓	288	2
阶梯自注意力模块×13	$14^2\times288$	阶梯自注意力模块×10	288	1
	$14^2\times288$	2×2卷积↓	576	2
	$7^2\times576$	阶梯自注意力模块×3	576	1



a: 倒残差模块+SE b: 沙漏卷积模块+ECA

图2 两种卷积模块结构对比

Figure 2 Comparison of two convolutional module structures

信息，提升模型的性能。

ECA是一种超轻量级模块，结构如图3所示。ECA根据通道内特征的重要性为每个通道分配不同的注意力值，使模型能有效地捕捉不同通道中的关键信息，减少冗余信息的计算，有助于网络合理地分配计算资源，提升模型准确率。该模块使用大小为k的一维卷积获取局部通道之间的相关性，并用S型函数（Sigmoid）根据通道间的相关性计算出各通道的注意力值，其中，一维卷积的核大小k由通道维数的函数来自适应确定，计算公式如式（1）所示：

$$k=\psi(C)=\left|\frac{\log_2(C)}{\gamma}+\frac{b}{\gamma}\right|_{\text{odd}} \tag{1}$$

其中，C为通道数，odd表示向上取最近的奇数，本研究取γ=2，b=1。

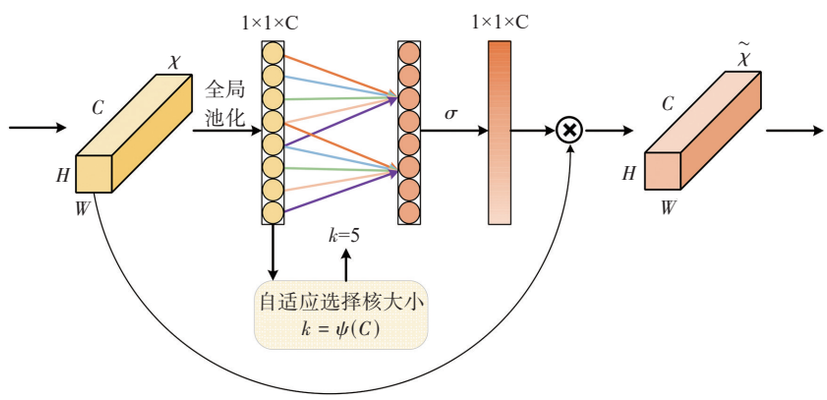


图3 ECA模块

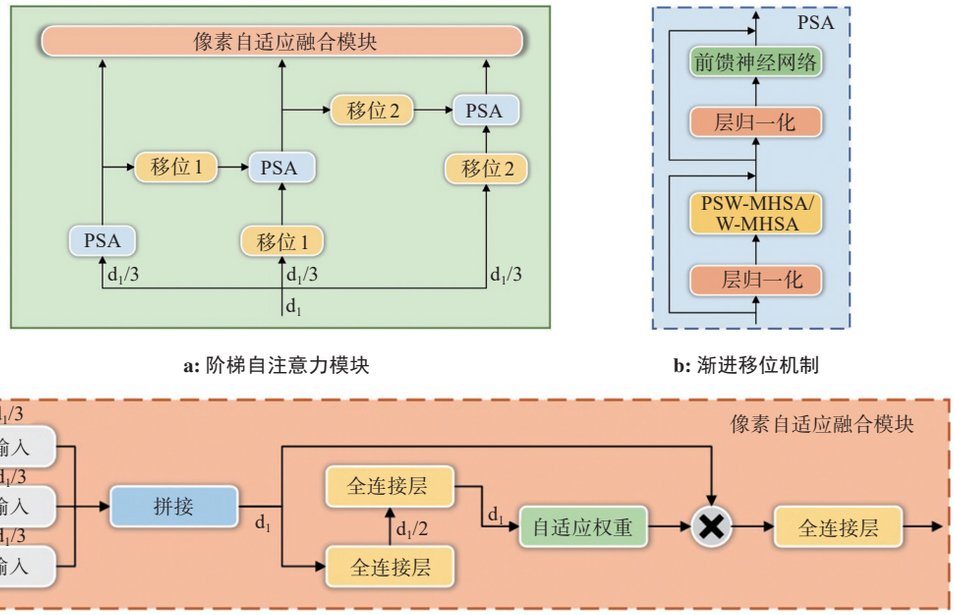
Figure 3 Efficient channel attention module

1.1.2 阶梯自注意力模块 医学影像蕴含复杂的结构关联和位置信息。为扩大模型的感受野，捕捉并利用影像中的全局关联，本研究使用阶梯自注意力模块，通过多尺度融合的渐进移位机制（Progressive Shift Attention, PSA）提取医学影像的全局特征[20]。

阶梯自注意力模块结构如图4a所示。阶梯自注意力模块由多个分支组成，通过在不同分支上均匀切分输入特征，并分别应用PSA提取特征，PSA内部结构如图4b所示。阶梯自注意力模块引入了移位操作。通过调整每个支路移位操作的幅度大小，实现

不同窗口的划分。具体操作为将上一个分支提取的特征通过移位操作调整至与本分支相同的幅度大小，作为本分支的输入用于PSA的计算。移位操作的引入扩大了感受野，实现了特征的多尺度融合，加强了模型对疾病复杂的结构关联和位置信息判别能力。

像素自适应融合模块（Pixel-Adaptive Fusion Module, PAFM）结构如图4c所示。将阶梯自注意力模块中所有分支的输出特征经过拼接后，输入到两个全连接层生成每个特征的权重。然后将得到的权重与特征相乘，并将结果输入到最后的全连接层，沿通道维度进行特征融合。PAFM能在空间和通道维度上进行自适应权值融合，使模型充分利用不同分支的信息，提高对复杂特征的提取能力。



a: 阶梯自注意力模块　　**b:** 渐进移位机制

c: 像素自适应融合模块

图4 阶梯自注意力块结构

Figure 4 Structure of ladder self attention block

阶梯自注意力模块将具有d个通道的输入特征均匀分配到每个分支中进行PSA计算。第一条分支由于不涉及移位操作，采用多头自注意力（Windows Multi-head Self-attention, W-MHSA）直接提取特征，W-MHSA的结构如图5a所示。其余分支因包含移位操作，使用渐进移位的多头自注意力（Progressive Shift Windows Multi-head Self-attention, PSW-MHSA）来提取特征，PSW-MHSA的结构如图5b所示。PSW-MHSA对当前分支的输入特征和前一个分支的输出特征应用相同的移位和窗口划分操作。首先，对输入特征进行1×1的卷积，得到查询矩阵（Query, $\boldsymbol{Q}$）和键矩阵（Key, $\boldsymbol{K}$），然后计算特征点之间的相似度。不同于W-MHSA，PSW-MHSA将前一个分支的输出作为自注意力计算中的值（Value, $\boldsymbol{V}$），以此实现多分支特征的交互，并扩大感受野，提高模型对长距离依赖关系的处理能力。阶梯自注意力模块计算过程如式（2）~式（5）所示：

$$\hat{\boldsymbol{O}}_t = \text{PSW-MHSA}(\boldsymbol{I}_t, \boldsymbol{O}_{t-1}) \tag{2}$$

$$\text{PSW-MHSA}(\boldsymbol{I}_t, \boldsymbol{O}_{t-1}) = \text{Soft max}\left(\frac{\boldsymbol{Q}_{I_t}\boldsymbol{K}_{I_t}}{\sqrt{d}}\right)\boldsymbol{O}_{t-1} + \boldsymbol{I}_t \tag{3}$$

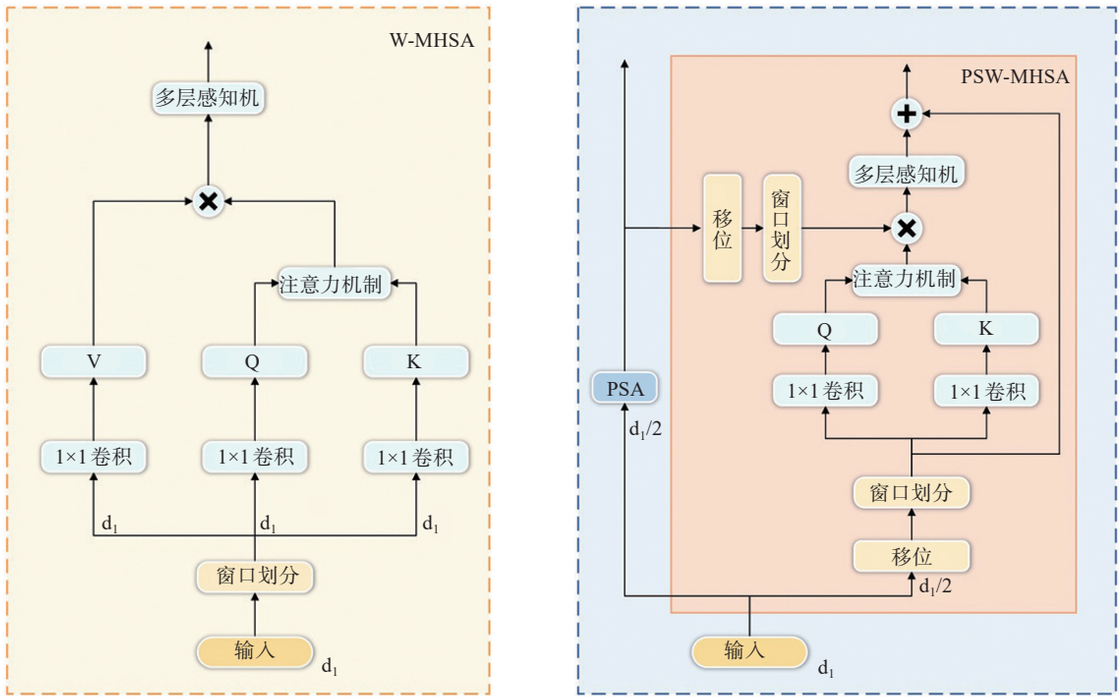
$$\boldsymbol{O}_t = \text{FFN}(\text{LN}(\hat{\boldsymbol{O}}_t)) \tag{4}$$

$$\overline{\boldsymbol{O}} = \text{PAFM}(\boldsymbol{O}_t) \tag{5}$$

其中，$t \epsilon \{0,1,2\ldots\}$为所在分支，Softamax作为归一化指数函数。PSW-MHSA第t支路输出特征$\hat{\boldsymbol{O}}_t$通过t支路的输入特征$\boldsymbol{I}_t$和t-1的输出特征$\boldsymbol{O}_{t-1}$计算得到，其中$\boldsymbol{I}_t$被用作$\boldsymbol{K}$和$\boldsymbol{Q}$，而$\boldsymbol{O}_{t-1}$被用作$\boldsymbol{V}$。计算得到的结果$\hat{\boldsymbol{O}}_t$经过层归一化和前馈神经网络处理后生成第t支路的输出$\boldsymbol{O}_t$。最后，通过PAFM将所有分支的输出特征进行融合，生成阶梯自注意力模块的最终输出特征$\overline{\boldsymbol{O}}$。

1.2 基于CSRA的多标签分类模块

医学影像通常涉及多个病变区域，且伴有许多冗余的背景信息。CSRA模块能动态调整不同特征的权重，使模型能更专注于最重要的特征，减少无关背景因素的干扰，提高多标签分类的准确率和泛化能力。CSRA的结构如图6所示。



a: W-MHSA　　b: PSW-MHSA

图5 W-MHSA和PSW-MHSA的结构

Figure 5 Structures of W-MHSA and PSW-MHSA

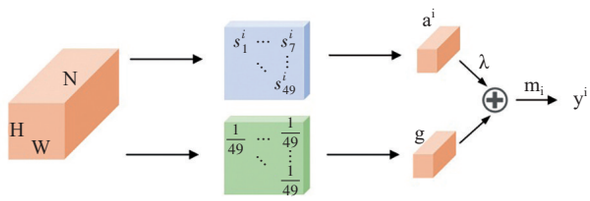


图6 CSRA结构图

Figure 6 Structure of class specific residual attention

多标签分类模块首先通过1×1的卷积降维到$H\times W\times N$的大小，其中N表示疾病类别的数量。随后送入到CSRA模块并解耦为特征张量$x_1,x_2,x_3,\cdots,x_{HW}$（$x_j\in R^{1\times 1\times N}$），通过空间池化，得到第$j$个位置特定类别的空间注意得分，并将其作为权重值，得到描述某一类空间特征信息注意力分值的残差分类向量。使用全局平均池化向量作为主要分类向量，与残差分类向量相加，得到标签预测向量$y\in R^{1\times 1\times N}$。CSRA的计算过程如式(6)、式(7)所示：

$$u_j^i=\frac{\exp(Tx_j^Tm_i)}{\sum_{k=1}^{HW}\exp(Tx_k^Tm_i)} \tag{6}$$

$$y=\frac{1}{HW}\sum_{k=1}^{HW}x_k+\lambda\sum_{k=1}^{HW}\mu_k^ix_k \tag{7}$$

其中，μ_j^i是特定类别的空间注意力得分，表示第i个类别出现在位置j的概率，m_i是第i类的分类器，λ是控制残差分类向量权值的超参数，T是温度系数。T值的不同能够调节归一化指数函数Softmax预测分布的尖锐程度，当T值趋于∞时，最大概率处的值趋近于1，其它值趋近于0。本研究中，多标签分类模块由8个不同温度系数的CSRA组成，$T_{1:8}$={1,2,3,4,5,6,7,∞}。因为不同的T值会导致不同程度的置信度量化，一些T值会更倾向于高置信度的预测，而其他T值可能更平衡置信度和准确性，所以综合考虑多个T值的组合以平衡置信度和提高模型的预测准确性。最后，将8个CSRA模块输出的预测向量相加，联合判断标签出现的概率：

$$y_{sum}=y_1+y_2+\cdots+y_8 \tag{8}$$

2 数据处理和评价指标

2.1 数据与预处理

2.1.1 IU X-Ray 来自印第安纳大学（Indiana University，IU）收集的胸部X光（X-Ray）数据集（IU X-Ray）包含3 955份诊断报告和7 470张正侧面胸腔X线影像，所有影像均被人工或自动标注[24]。参考Alfarghaly等[25]采用的标签提取方法，保留出现频次大于25的标签，得到包含疾病的名称、位置、严重程度和患病器官等信息的105个标签。

2.1.2 Chest X-Ray14 包含14种疾病标签的胸部X射线数据集（Chest X-Ray14）由美国国立卫生研究院收集并公开，包含30 805位患者的112 120幅正面胸

腔X线影像，其中60 316幅影像标注为“不患病”标签，其余51 804幅影像每幅对应一个或多个14种常见的病理标签[26]。数据集中所有标签采用自然语言处理的方法，通过诊断报告自动标注。

2.2 评价指标

选用接受者操作特征曲线（Receiver Operating Characteristics, ROC）下面积（Area Under ROC Curve, AUC）评价图像编码器的多标签分类性能，ROC曲线横坐标为假阳性率（False Positive Rate, FPR），表示所有负类样本中被错误判断成正类的比例，纵坐标为真阳性率（Ture Positive Rate, TPR），表示所有正类样本中被正确判断为正类的比例。使用所有标签的平均AUC得分作为评价指标，AUC得分越接近1，模型分类性能越好。

3 实验结果与分析

3.1 实施细节

在训练过程中，所有图像被调整至512×512大小，数据增强方式采用随机裁剪、随机旋转、灰度变换。验证和测试过程则将图像缩小至256×256并居中裁剪出224×224大小输入模型。使用二元交叉熵（Binary Cross-Entropy, BCE）损失函数和自适应矩估计优化器（Adaptive Moment Estimation, Adam）优化模型，初始学习率为1e-3，权重衰减为5e-5，并采用余弦退火学习率衰减[27]，批大小（Batch size）设置为16。

IU X-Ray数据集使用105种标签对影像进行标注，更利于展示模型性能和缺陷，为后续诊断报告生成的研究提供理论参考，然而IU X-Ray数据量相对较少，缺少对比实验。因此使用IU X-Ray数据集训练模型，并进行超参数调整和相关实验，最后将调整好的超参数直接用于Chest X-Ray14数据集的训练和测试，与其他方法对比，验证模型有效性。将IU X-Ray数据集按8∶1∶1随机划分训练集、验证集和测试集，训练集迭代次数为50轮。Chest X-Ray14数据集则随机选取1 000幅影像作为测试集，其余影像按9∶1划分为训练集和验证集。模型使用Python3.7以及PyTorch1.8搭建，并在RTX3090上进行训练。

3.2 实验结果

图7展示了SLRN在IU X-Ray数据集中训练的过程，包括所有标签的平均准确率（mean aucroc）、正常（normal）以及6种随机选择的疾病标签，分别是不透明（opacity）、间隙的（interstitial）、肺不张（pulmonary atelectasis）、扩散的（diffuse）、退化的（degenerative）和慢性的（chronic），以测试集平均AUC得分作为评价指标。由图7可知，SLRN在40轮时趋于收敛，最终模型的平均AUC得分达到0.843，证明了模型的收敛性。

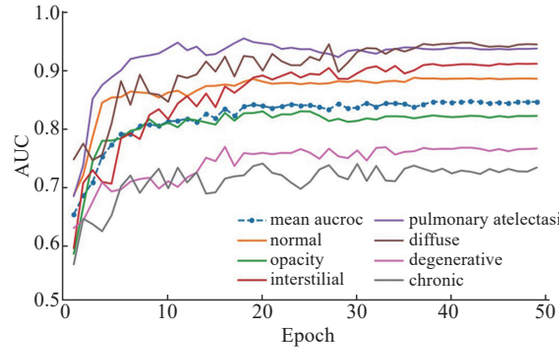


图7 SLRN训练曲线

Figure 7 SLRN training curve

3.3 消融实验

为验证多标签分类模块中CSRA不同头数对AUC的影响，在IU X-ray数据集上进行对比实验。结果如表2所示。当头数（head，H）=8时，模型的平均AUC最高，当H<8或H>8时，模型的平均AUC都在下降，说明适当增加CSRA头数，有助于捕捉到不同标签之间的相关性和每个特征的重要性。由于不同的病变可能具有不同的形状、大小、位置等特征，当分类头太少，模型无法捕捉到所有的关键信息，导致特征提取能力较差。分类头过多时，模型也不能提取到更多的特征，反而会导致参数量增加，从而降低模型的性能。因此，在本文模型结构中，CSRA分类头数选择为8。

表2 CSRA不同头数对AUC的影响

Table 2 Effects of different head counts in CSRA on AUC

CSRA头数H/个	AUC
0	0.805
2	0.823
4	0.828
6	0.833
8	0.843
10	0.832

为验证沙漏卷积模块与ECA的有效性，在IU X-ray数据集上进行消融实验，所有消融实验中，阶梯自注意力模块与CSRA均保持相同参数，结果如表3所示。实验1的基础模型采用倒残差模块+SE的组合，AUC得分为0.827。实验3在保留SE的前提下，使用沙漏卷积模块代替倒残差模块，AUC得分提升0.013，证明沙漏卷积模块对深度卷积位置的改变能

够提取更丰富的特征。如实验1和本文模型SLRN所示，将SE替换为ECA使模型AUC得分提升0.012，证明ECA根据通道内特征的重要性重新分配注意力值的操作相比SE提取特征的能力更强。其次，通过实验2可知，单独使用沙漏卷积模块比实验1中倒残差模块+ECA的组合AUC得分更高，说明沙漏卷积模块对模型分类准确性的提升更为重要。消融实验证明本文模型（沙漏卷积模块+ECA）的组合能最大程度增强模型对疾病特征的提取能力。

表3 沙漏卷积模块的消融实验

Table 3 Ablation study of sandglass convolutional module

消融实验	倒残差模块	沙漏卷积模块	SE	ECA	AUC
基础模型	√	×	√	×	0.827
实验1	√	×	×	√	0.831
实验2	×	√	×	×	0.836
实验3	×	√	√	×	0.840
SLRN	×	√	×	√	0.843

3.4 对比实验

为研究不同主干模型对医学影像多标签分类的影响，在IU X-Ray数据集中进行不同主干模型的对比实验。分别使用两种CNN模型（DenseNet121、ResNet50）以及两种视觉Transformer模型（Vision Transformer、Swin Transformer）代替本文的特征提取模型，并保持其它参数不变[28-30]。实验结果如表4所示，SLRN在多标签分类任务中取得最佳的分类效果，比Swin Transformer提升0.023。

表4 不同主干模型对比实验

Table 4 Comparative experiments on different backbone models

模型	AUC
DenseNet121[28]	0.787
ResNet50[29]	0.779
Vision Transformer[14]	0.807
Swin Transformer[30]	0.820
SLRN	0.843

医学影像包括复杂的位置信息和多样的疾病特征，DenseNet121和ResNet50两种CNN模型通过卷积操作能有效捕捉图像的局部特征，但其感受野受限，无法有效捕捉医学影像中的全局信息，导致多标签分类准确率相对较低。相比之下，Vision Transformer、Swin Transformer拥有完整的视野，能同时关注影像各个位置的状态，并专注于异常区域，取得更好的分类效果。本文模型SLRN充分结合CNN和Transformer的优势，同时捕捉图像中的局部特征和全局信息，使得模型更有效地处理长距离依赖关系，且阶梯自注意力模块能在不同尺度上提取特征信息并进行传递和融合，善于发现医学影中复杂的结构关联和位置信息，使模型获得更高的性能。

为全面展示SLRN的性能，在Chest X-Ray14数据集中与其他代表性模型进行对比测试，包括经典的CNN网络多标签分类方法（Dense Squeeze-and-Excitation Network, SE Net）、教师网络和半监督训练方法（Self-Supervised Mean Teacher for Semi-Supervised, SMTS）、图神经网络方法（Semantic Similarity Graph Embedding, SSGE）以及将CNN网络和语义向量相结合并提出双加权度量损失函数的方法（Semantic Vectors and Dual-Weighted Metric Loss, DWM）[31-32]。对比结果如表5所示。

本文提出的SLRN在11种疾病标签的AUC得分中均达到最优，尤其在“心脏肥大”和“肺实变”的识别准确率比其他方法都有所提升，而“肺气肿”和“胸膜增厚”也接近最佳模型的分类效果。14种胸部疾病标签的平均AUC得分为0.845，证明多尺度的特征融合为模型提供丰富的信息并增强对疾病的判别能力。

“肺浸润”、“肺结节”和“胸膜增厚”的识别准确率相对较低。其中，对于“肺侵润”的判断主要取决于细微的纹理变化，而这些特征较为复杂；而“肺结节”和“胸膜增厚”由于患病区域较小，容易受到无关特征的干扰，同时由于训练样本较少，模型无法充分学习到足够的特征。

3.5 模型可视化

在Chest X-Ray14数据集上进行模型可视化实验。使用Grad-CAM[33]生成热力图来验证模型的可解释性。红色突出显示模型中最受关注的部分，通过肉眼可以快速定位病变发生的区域。将分类概率大于0.5的标签作为最终预测，红色代表真实标签与预测标签一致，蓝色代表误诊，绿色代表漏诊。如图8所示，SLRN对于“肺气肿”、“疝气”、“胸腔积液”等特征明显的标签能够准确预测，但对于“肺侵润”、“胸膜增厚”等患病区域较小的标签识别效果较差，容易出现漏诊与误诊，这与表5中的结果一致。

4 结 论

本研究提出一种基于SLRN模型的胸部影像多标签分类方法，主要工作包括以下3个方面：在特征提取模块中，利用沙漏卷积模块增强局部特征的提取能力，避免特征信息在传输过程中的丢失；同时，

表5 Chest X-Ray14数据集上AUC对比

Table 5 Comparison of AUC on Chest X-Ray14 dataset

标签名称	标签占比/%	SE Net[31]	SMTS[32]	SSGE[11]	DWM[12]	SLRN
肺不张	10.31	0.785	0.787	0.792	0.797	0.799
心脏肥大	2.48	0.877	0.874	0.892	0.911	0.942
胸腔积液	11.88	0.863	0.838	0.840	0.844	0.885
肺浸润	17.74	0.673	0.709	0.714	0.724	0.727
肿块	5.16	0.804	0.833	0.848	0.836	0.854
肺结节	5.64	0.729	0.799	0.812	0.802	0.787
肺炎	1.28	0.742	0.739	0.733	0.739	0.759
气胸	4.73	0.842	0.871	0.885	0.869	0.898
肺实变	4.16	0.785	0.759	0.753	0.725	0.829
水肿	2.05	0.873	0.845	0.848	0.860	0.879
肺气肿	2.24	0.858	0.937	0.948	0.933	0.936
纤维化	1.50	0.775	0.834	0.827	0.849	0.868
胸膜增厚	3.01	0.756	0.793	0.795	0.753	0.789
疝气	0.20	0.865	0.933	0.932	0.916	0.940
平均AUC得分	–	0.802	0.825	0.830	0.826	0.845

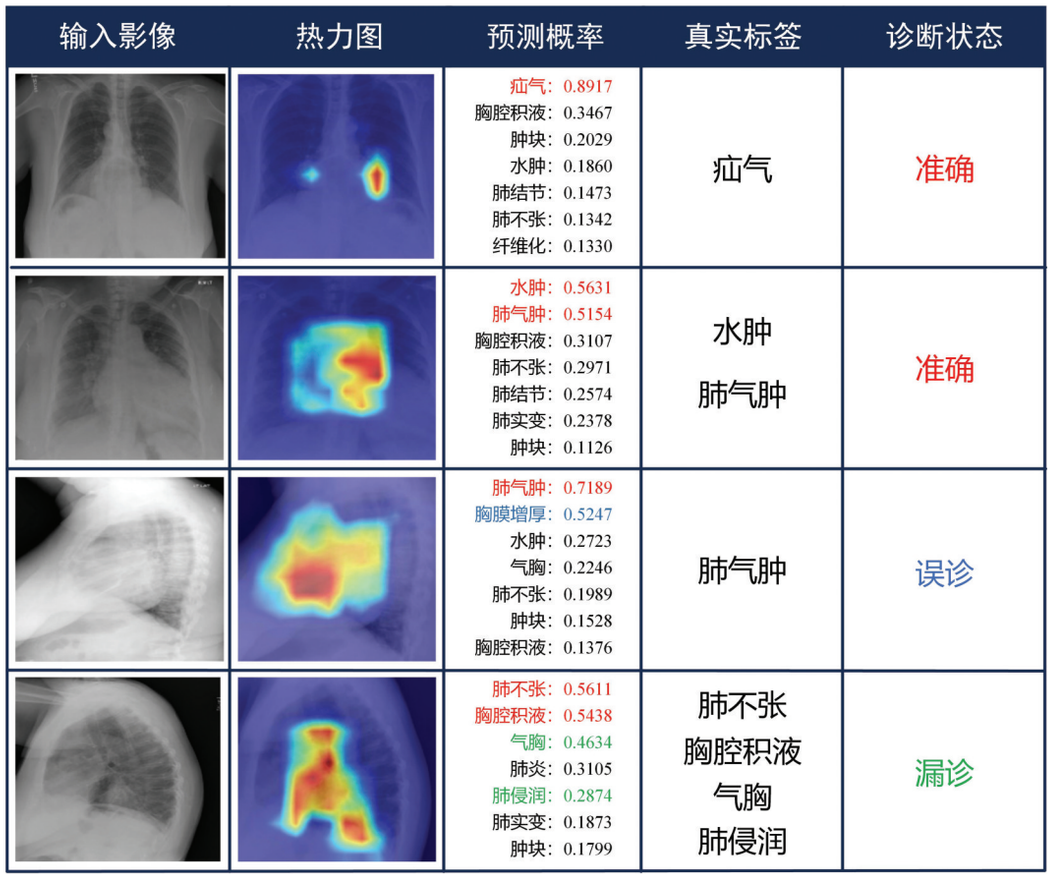


图8 SLRN模型预测热力图

Figure 8 SLRN model prediction heat maps

利用阶梯自注意力模块扩大感受野，使模型能更全面地提取多尺度的特征信息。在多标签分类模块中引入CSRA，通过实验验证CSRA的必要性，并最终确定CSRA头数为8；CSRA使模型能关注到疾病特征的空间位置信息，从而提高模型对多标签分类任务的性能。使用IU X-Ray和Chest X-Ray14数据集对SLRN模型进行实验验证，分别取得了0.843和0.845的AUC得分，在胸部影像多标签分类任务中表现出色，通过生成的热力图反映了模型关注的疾病所在位置，增加模型的可解释性。

本文模型仅对数据集中标签出现次数大于20次的数据进行处理，对于小样本数据还没有进一步验证，未来的研究将进一步展开相关实验，以提升模型的性能，进一步提高模型的泛化能力和适应性，为模型的临床应用提供更好服务。

【参考文献】

[1] Kumar SV, Gunasundari R. Computational intelligence in eye disease diagnosis: a comparative study[J]. Med Biol Eng Comput, 2023, 61(3): 593-615.

[2] Zhao ZY, Chopra K, Zeng Z, et al. Sea-net: squeeze-and-excitation attention net for diabetic retinopathy grading[C]//2020 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2020: 2496-2500.

[3] Han Y, Qi HG, Wang L, et al. Pulmonary nodules detection assistant platform: an effective computer aided system for early pulmonary nodules detection in physical examination[J]. Comput Methods Programs Biomed, 2022, 217: 106680.

[4] Kumar P, Kumar A, Srivastava S, et al. A novel bi-modal extended Huber loss function based refined mask RCNN approach for automatic multi instance detection and localization of breast cancer [J]. Proc Inst Mech Eng H, 2022, 236(7): 1036-1053.

[5] 胡晓阳, 李哲. 基于卷积神经网络和Transformer的肝脏CT图像分割方法[J]. 中国医学物理学杂志, 2023, 40(4): 423-428.
Hu XY, Li Z. Liver CT image segmentation method based on CNN and Transformer[J]. Chinese Journal of Medical Physics, 2023, 40 (4): 423-428.

[6] Shiraishi J, Li Q, Appelbaum D, et al. Computer-aided diagnosis and artificial intelligence in clinical imaging[J]. Semin Nucl Med, 2011, 41(6): 449-462.

[7] Chen BZ, Li JX, Guo XB, et al. DualCheXNet: dual asymmetric feature learning for thoracic disease classification in chest X-rays [J]. Biomed Signal Process Control, 2019, 53: 101554.

[8] Chen BZ, Li JX, Lu GM, et al. Label co-occurrence learning with graph convolutional networks for multi-label chest X-ray image classification[J]. IEEE J Biomed Health Inform, 2020, 24(8): 2292-2302.

[9] Tao R, Liu W, Zheng G. Spine-transformers: vertebra labeling and segmentation in arbitrary field-of-view spine CTs *via* 3D transformers [J]. Medical Image Analysis, 2022, 75: 102258.

[10] Cheng Y, Feng JC, Jia KB. A lung disease classification based on feature fusion convolutional neural network with X-ray image enhancement[C]//2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). Piscataway, NJ, USA: IEEE, 2018: 2032-2035.

[11] Chen BZ, Zhang Z, Li YJ, et al. Multi-label chest X-ray image classification *via* semantic similarity graph embedding[J]. IEEE Trans Circuits Syst Video Technol, 2022, 32(4): 2455-2468.

[12] Jin YF, Lu HJ, Zhu WJ, et al. Deep learning based classification of multi-label chest X-ray images *via* dual-weighted metric loss[J]. Comput Biol Med, 2023, 157: 106683.

[13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2017: 6000-6010.

[14] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[C]// Proceedings of the 9th International Conference on Learning Representations. Virtual Event, Austria: OpenReview.net, 2021: 1-21.

[15] He KL, Gan C, Li ZY, et al. Transformers in medical image analysis [J]. Intell Med, 2023, 3(1): 59-78.

[16] Al-Hammuri K, Gebali F, Kanan A, et al. Vision transformer architecture and applications in digital health: a tutorial and survey [J]. Vis Comput Ind Biomed Art, 2023, 6(1): 14.

[17] Xie YT, Zhang JP, Shen CH, et al. CoTr: efficiently bridging CNN and transformer for 3D medical image segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer International Publishing, 2021: 171-180.

[18] Wu HS, Chen SH, Chen GL, et al. FAT-Net: feature adaptive transformers for automated skin lesion segmentation[J]. Med Image Anal, 2022, 76: 102327.

[19] Tao R, Zheng GY. Spine-transformers: vertebra detection and localization in arbitrary field-of-view spine CT with transformers [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer International Publishing, 2021: 93-103.

[20] Wu GJ, Zheng WS, Lu YT, et al. PSLT: a light-weight vision transformer with ladder self-attention and progressive shift[J]. IEEE Trans Pattern Anal Mach Intell, 2023, 45(9): 11120-11135.

[21] Zhu K, Wu JX. Residual attention: a simple but effective method for multi-label recognition[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 184-193.

[22] Sandler M, Howard A, Zhu ML, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 4510-4520.

[23] Wang QL, Wu BG, Zhu PF, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 11531-11539.

[24] Demner-Fushman D, Kohli MD, Rosenman MB, et al. Preparing a collection of radiology examinations for distribution and retrieval [J]. J Am Med Inform Assoc, 2016, 23(2): 304-310.

[25] Alfarghaly O, Khaled R, Elkorany A, et al. Automated radiology report generation using conditioned transformers[J]. Inform Med Unlocked, 2021, 24: 100557.

[26] Wang XS, Peng YF, 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]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 3462-3471.

[27] He T, Zhang Z, Zhang H, et al. Bag of tricks for image classification with convolutional neural networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 558-567.

[28] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 2261-2269.

[29] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.

[30] Liu Z, Lin YT, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 9992-10002.

[31] 张智睿, 李锵, 关欣. 密集挤压激励网络的多标签胸部X光片疾病分类[J]. 中国图象图形学报, 2020, 25(10): 2238-2248.
Zhang ZR, Li Q, Guan X. Multilabel chest X-ray disease classification based on a dense squeeze-and-excitation network[J]. Journal of Image and Graphics, 2020, 25(10): 2238-2248.

[32] Liu FB, Tian Y, Cordeiro FR, et al. Self-supervised mean teacher for semi-supervised chest X-ray classification[C]//Proceedings of the Machine Learning in Medical Imaging. Cham: Springer International Publishing, 2021: 426-436.

[33] Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks *via* gradient-based localization [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 618-626.

（编辑：谭斯允）