

## 结合自注意力与卷积的胸部 X 光片疾病分类研究

关欣, 耿晶晶, 李强\*

天津大学微电子学院, 天津 300072

**摘要** 胸部 X 光片可用于诊断多种胸部疾病。由于胸部疾病特征复杂多样, 现有的胸部 X 光片疾病分类算法难以学习胸部疾病复杂的鉴别表征以及未关注不同疾病之间的相关性信息。针对以上问题, 提出一种结合自注意力与卷积的疾病分类算法, 该算法采用全维度动态卷积替换残差网络的标准卷积, 从而提高网络对多尺度信息的特征提取能力。此外, 在卷积神经网络中引入自注意力模块, 可以提供捕获多种疾病之间相关性的全局感受野。最后, 提出高效的双路注意力, 使神经网络更加关注病灶区域、自动捕捉病变位置变化。在 ChestX-ray14 数据集上, 对所提模型进行评估, 实验结果表明: 所提算法对 14 种胸部疾病的平均受试者工作特性曲线下的面积 (AUC) 达到 0.839, 检测结果与目前其他 7 种先进算法相比在准确率和效率上有所提升。

**关键词** 胸部 X 光片; 全维度动态卷积; 自注意力; 双路注意力; 疾病分类

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP231180

## Research on Combining Self-Attention and Convolution for Chest X-Ray Disease Classification

Guan Xin, Geng Jingjing, Li Qiang\*

School of Microelectronics, Tianjin University, Tianjin 300072, China

**Abstract** Chest X-rays are used to diagnose a wide range of chest conditions. However, due to the complicated and diverse features of thoracic diseases, existing disease classification algorithms for chest radiographs have difficulty in learning the complex discriminating features of thoracic diseases and do not fully consider correlation information between different diseases. This study proposes a disease classification algorithm that combines self-attention and convolution to address these problems. This study employs omni-dimensional dynamic convolution to replace the standard convolution of the residual network to enhance the feature extraction capabilities of the network for multi-scale information. In addition, a self-attention module is introduced into the convolutional neural network to provide global receptive fields that capture correlations between multiple diseases. Finally, an efficient double path attention is proposed that allows the network to give greater attention to the focal area and automatic capturing of changes in lesion locations. The proposed model is evaluated on the ChestX-ray14 dataset. Experimental results show that the accuracy of the algorithm and the efficiency of diagnosis for the classification of 14 chest diseases is improved over those of the seven current state-of-the-art algorithms, with an average area under receiver operating characteristic curve (AUC) value of 0.839.

**Key words** chest X-ray; omni-dimensional dynamic convolution; self-attention; double path attention; disease classification

## 1 引言

胸部 X 光片 (Chest X-ray, CXR) 是筛查胸部疾病最常用的影像学检查方式之一, 但针对 CXR 图像的临床分析是一个复杂的过程, 主要依靠专业放射科医生

的人工观察, 需耗费大量时间和精力<sup>[1]</sup>。随着患者数量的增加, 胸部疾病的胸片数量也随之增加, 且大多数 CXR 图像都存在一些不可避免的问题, 如背景复杂、病理表现存在多样性、疾病之间相互作用等<sup>[2]</sup>, 这些问题可能会降低医生诊断胸部疾病的效率与准确性。因

收稿日期: 2023-04-26; 修回日期: 2023-05-14; 录用日期: 2023-05-30; 网络首发日期: 2023-06-10

基金项目: 国家自然科学基金 (61471263, 61872267, U21B2024)、天津市自然科学基金 (16JCZDJC31100)、天津市研究生科研创新项目 (2021YJSS023)、天津大学创新基金会 (2021XZC-0024)

通信作者: liqiang@tju.edu.com

此,开发能够辅助诊断的自动算法具有重要意义。

近年来,深度卷积神经网络(Deep convolutional neural networks, DCNN)在计算机视觉方面取得许多突破,可以从原始图像中探索高级复杂的特征表示<sup>[3]</sup>。这促使许多研究人员将DCNN应用于胸部疾病的辅助诊断任务,如疾病分类<sup>[4-5]</sup>、病灶分割或检测<sup>[6-9]</sup>和图像配准<sup>[10-12]</sup>。在多标签胸部疾病分类领域,Wang等<sup>[13]</sup>提出ChestX-ray14数据集,使用多种DCNN模型实现胸部疾病分类和患病区域定位功能,其中ResNet取得的分类精度最高。Chen等<sup>[2]</sup>提出一种非对称特征学习网络,结合ResNet和DenseNet两种网络,从原始CXr图像的异常中自适应捕获更多的鉴别特征。Guan等<sup>[14]</sup>提出ConsultNet双分支架构来学习判别特征,该网络可以自动聚焦于疾病关键区域,并自适应捕捉不同疾病特征之间的内在联系。

此外,许多研究人员将注意力机制应用于胸部疾病检测领域。Ma等<sup>[15]</sup>提出一种以ResNet101为主干网络的多注意分类网络,该网络添加了挤压激励模块以有效构建通道之间的依赖关系,同时引入空间注意力模块以实现全局和局部信息的融合。Guan等<sup>[16]</sup>提出一个分类式剩余注意力学习框架,采用分类剩余注意力机制为不同的特征空间区域分配不同的权重,以增强特定病理的相关特征并抑制不相关特征,探究标签之间的依赖关系。Wang等<sup>[17]</sup>提出三重注意学习A<sup>3</sup>Net模型,该模型使用经过预训练的DenseNet121作为骨干网络进行特征提取,并将3个注意模块集成在

统一的框架中进行通道级、元素级和尺度级注意学习。

由以上研究可知,以DCNN为基础的深度学习方法可以高效地执行胸部疾病分类任务。但由于CXr图像中不同疾病的纹理、位置等特征复杂多样,普通网络难以学习复杂病理的鉴别特征,从而难以准确识别。不同的胸部疾病存在明显的尺寸差异,DCNN可能无法适应尺度变化。此外,CXr图像中通常包含一种或多种病理,不同病理之间会受到内在复杂关系的影响。针对以上问题,提出一种基于全维度动态卷积(Omni-dimensional dynamic convolution, ODConv)的自注意力分类网络(DA-Net),该网络能够提取病灶的多尺度信息,自动捕捉疾病位置变化以及挖掘疾病之间的潜在语义关系,从而提高疾病分类的准确性。

## 2 基本原理

所提的DA-Net整体框架如图1所示,该网络在ResNet50基础上进行改进,包含S<sub>0</sub>到S<sub>4</sub>5个阶段,各阶段空间分辨率逐渐下降。首先,CXr图像经过预处理后进入该网络并被输入到S<sub>0</sub>阶段。S<sub>0</sub>通过卷积核大小为7×7、步长为2的卷积层对输入图像进行降维,输出112×112大小的特征图,然后经过3×3最大池化操作,输出56×56大小的特征图。随后,特征图依次经过S<sub>1</sub>至S<sub>4</sub>阶段,其中S<sub>1</sub>、S<sub>2</sub>、S<sub>3</sub>使用全维度动态卷积注意模块(Omni-dimensional dynamic convolutional attention block, DA-Block)充分提取病灶特征,该模块包含多个1×1卷积、(ODConv)和高效双路注意力

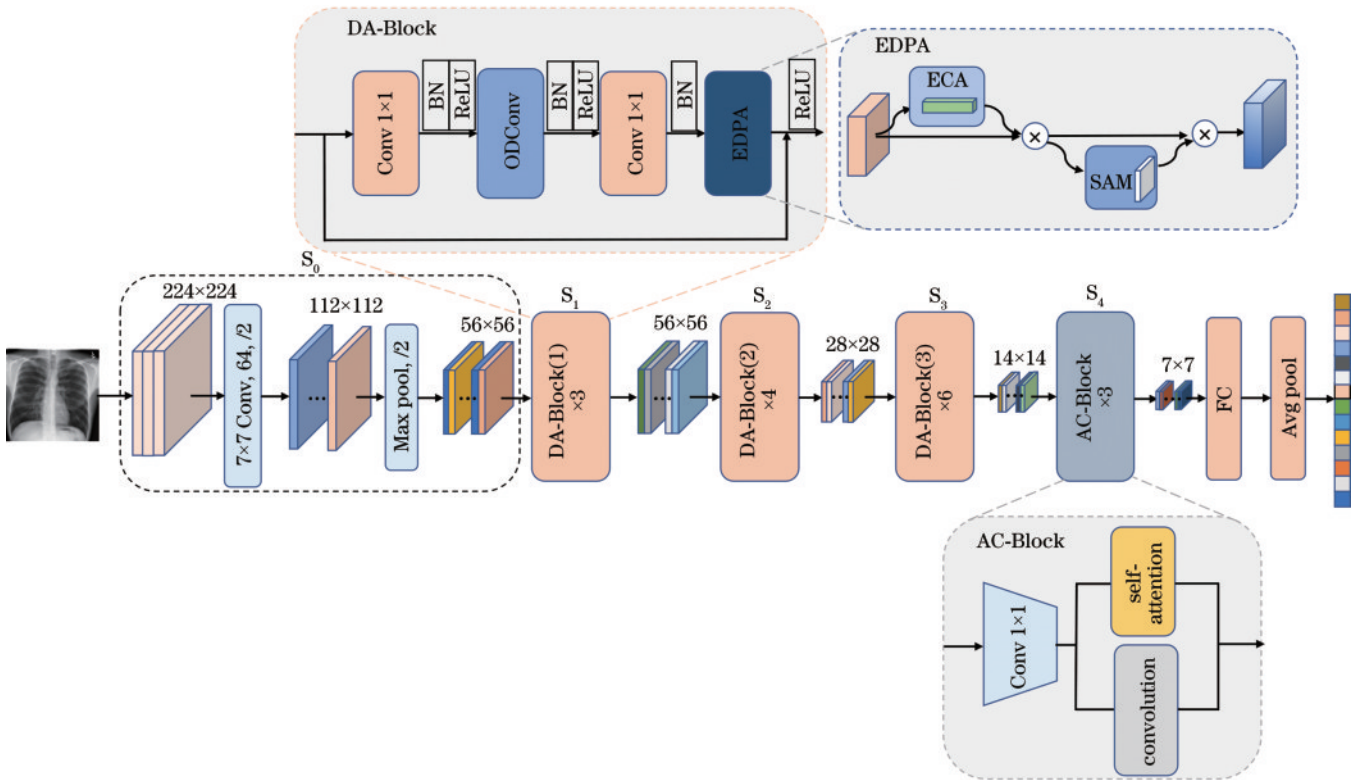


图1 DA-Net整体框架

Fig. 1 Overall framework of DA-Net

(Efficient double path attention, EDPA) 机制, 其中 BN 为批归一化操作, ReLU 为修正线性单元。最后, 将自注意力卷积融合模块 (Self-attention and convolution block, AC-Block) 置于最后一个阶段  $S_4$ , 该模块将卷积与自注意力结合, 使所提网络能够最大程度保留全局和局部特征。特征图在经过 AC-Block 输出  $7 \times 7$  大小特征图后, 进入全连接 (Fully connected, FC) 层输出 14 维特征向量, 最后利用特征向量中各疾病标签实现胸部疾病的分类。各模块的基本原理和情况如图 1 所示。

### 2.1 ODConv 模块

卷积核相应的核空间有 4 个维度: 核空间大小、每

个卷积核的输入通道数和输出通道数, 以及卷积核数。ODConv 通过并行策略采用多维注意力机制对卷积核的 4 个维度进行互补注意力的学习, 通过渐进方式将卷积操作沿空间、输入通道、输出通道以及卷积核数维度乘以不同的注意力, 可以使卷积操作对于输入数据在各个维度上具有不同的敏感性, 从而显著增强 CNN 基本卷积运算的特征提取能力<sup>[18]</sup>, 使所提模型能够捕获丰富的上下文信息。每个卷积层中采用多个卷积核, 使网络能提取到不同尺度的特征信息, 充分学习具有复杂病理、尺寸多变的病灶图像的鉴别表征, 从而区分不同疾病。

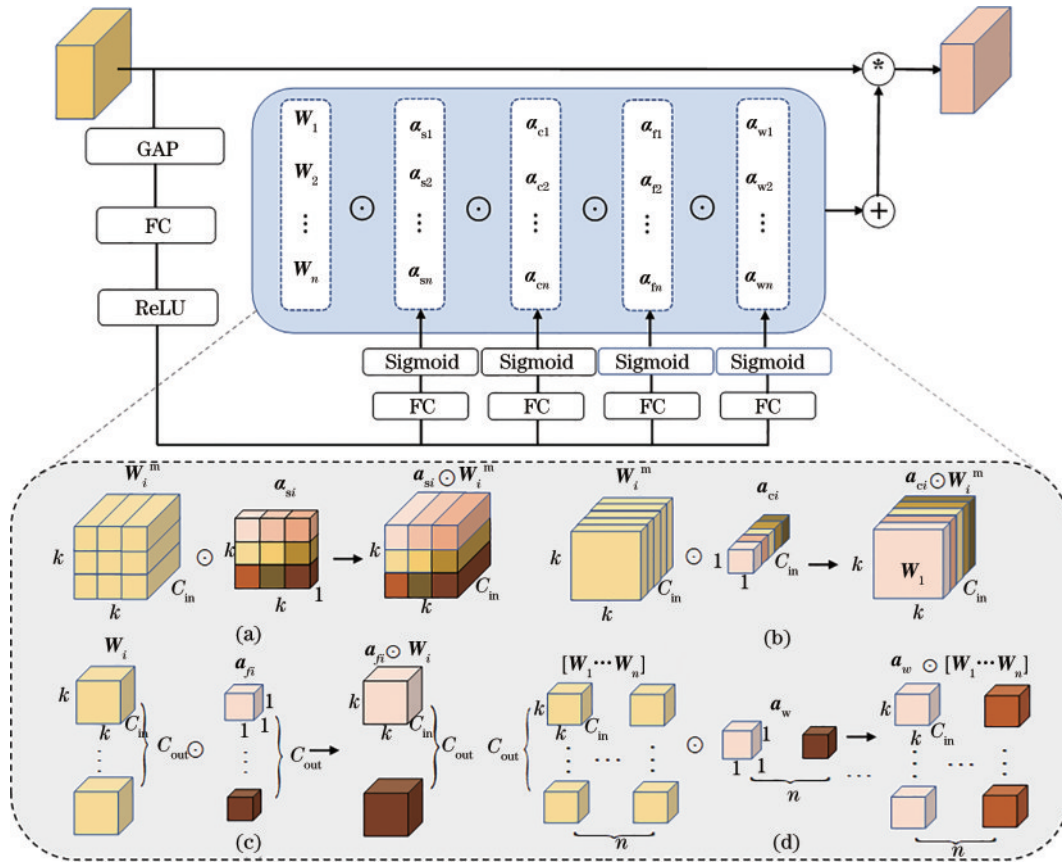


图 2 ODConv 的计算流程

Fig. 2 Workflow for ODConv

该模块的计算流程如图 2 所示, 对输入特征图  $X$  而言, 其输出特征  $Y$  为

$$Y = (\alpha_{w1} \odot \alpha_{f1} \odot \alpha_{c1} \odot \alpha_{s1} \odot W_1 + \dots + \alpha_{wn} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) * X, \quad (1)$$

式中:  $\alpha_{si} \in \mathbf{R}^{k \times k}$ ,  $\alpha_{ci} \in \mathbf{R}^{C_{in}}$ ,  $\alpha_{fi} \in \mathbf{R}^{C_{out}}$ ,  $\alpha_{wi} \in \mathbf{R}$  分别为卷积核  $W_i (i = 1, 2, 3, \dots, n)$  的空间、输入通道、输出通道以及整个卷积核维度的注意力权重;  $\odot$  为沿卷积核不同维度进行的乘法操作。

这 4 类权重由多头注意模块计算得到, 首先通过全局平均池化 (Global average pooling, GAP) 操作将输入  $X$  压缩成长度为输入通道数  $C_{in}$  的特征向量。随后, 将特征向量通过一个 FC 层和 4 个头部分支。其

中, FC 层将压缩后的特征向量以降维比  $r$  映射到低维空间 (设置  $r = 1/16$ )。4 个头分支中的每个分支都有一个 FC 层, 其输出大小分别为  $k \times k$ ,  $C_{in} \times 1$ ,  $C_{out} \times 1$  和  $n \times 1$ , 通过 Sigmoid 函数分别生成归一化注意力权重, 并共享给所有的卷积核, 其中  $C_{out}$  为输出通道数。为尽可能地增强卷积层对特征信息的提取能力, 使用  $n = 4$  个卷积核, 将 ResNet50 模型中  $S_1$  到  $S_3$  阶段的  $3 \times 3$  卷积全部替换为 ODConv, 以提高网络对疾病复杂特征的学习鉴别能力。

### 2.2 EDPA 机制

为使网络更加关注病灶区域以及捕捉病灶位置变化<sup>[19]</sup>, 引入 EDPA 机制并将此模块嵌入网络中作为卷

积的补充。该模块能够分别提取通道域和空间域的注意力信息,其中高效通道级注意(Efficient channel attention, ECA)模块能够强化特征图的信息通道,排除无关区域的干扰,空间级注意模块(Spatial attention

module, SAM)使网络聚焦于病灶相关区域,更加高效地提取病灶位置像素特征。双路注意力模块可以增强模型的判别能力,进一步提高模型诊断性能,其结构如图 3 所示。

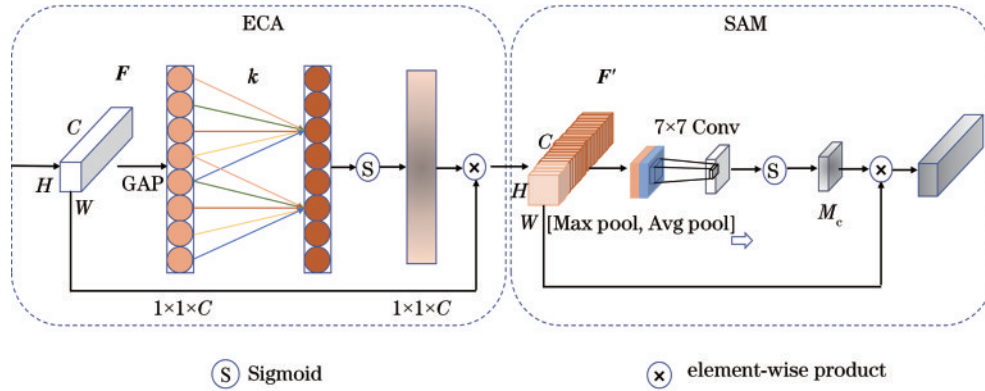


图 3 EDPA 机制  
Fig. 3 EDPA mechanism

定义输入特征图  $F \in \mathbf{R}^{H \times W \times C}$ , 其中  $H$ 、 $W$  为通道高度、宽度,  $C$  为通道大小, 在进行通道注意时, 该模块同时考虑每个通道及其  $k$  个相邻通道, 通道的权值可以使用一维卷积进行局部跨通道交互信息, 在所有通道之间共享权重来得到。其中, 卷积核大小  $k$  代表局部跨信道交互的覆盖率, 聚合特征  $y \in \mathbf{R}^C$ , 注意力通道权值  $M_c$  可表示为

$$M_c = \sigma[D_k(y)], \quad (2)$$

式中:  $D_k(\cdot)$  为卷积核大小为  $k$  的一维卷积;  $\sigma(\cdot)$  为 Sigmoid 激活函数。将  $M_c$  与输入  $F$  相乘, 得到经 ECA 模块提取后的输出  $F'$ 。

随后, 将经过通道注意的输出  $F'$  作为 SAM 的输入, 对同一通道内所有位置的元素分别进行最大池化和平均池化操作并完成通道拼接, 有效突出显示信息区域, 得到尺度为  $H \times W \times 2$  的空间掩模, 经  $7 \times 7$  卷积降维成单通道, 最后经 Sigmoid 激活函数完成归一化, 得到各位置注意力权重  $M_s$ ,  $M_s$  可表示为

$$M_s = \sigma[f^{7 \times 7}(F'_{\text{avg}}; F'_{\text{max}})], \quad (3)$$

式中:  $(\cdot; \cdot)$  为通道拼接;  $F'_{\text{avg}}$ 、 $F'_{\text{max}}$  分别为  $F'$  经过平均池化和最大池化的结果。将  $M_s$  与输入  $F'$  相乘, 得到经 SAM 提取后的输出。

### 2.3 AC-Block

为了捕获不同病理之间的内在复杂关系, 将自注意力模块引入 CNN。自注意力模块能够提供全局感受野, 可获得更多的语义信息<sup>[20]</sup>, 同时该模块可以动态地依赖输入的特征, 能够获取不同空间位置的相互关系<sup>[21]</sup>, 捕捉多标签胸部疾病之间的相关性信息。将自注意力与卷积结合, 构成 AC-Block, 并将该模块垂直堆叠在卷积层之后, 先通过卷积层降低特征图空间大小获得局部感受野, 再采用自注意力模块提取特征图

中的全局信息, 从而最大程度保留全局和局部特征, 挖掘多种疾病之间的潜在相关性信息, 提升模型表达能力。

AC-Block 包括两个阶段, 假设卷积核  $K \in \mathbf{R}^{C_m \times C_m \times k \times k}$ , 其中  $k$  为卷积核大小。张量  $F \in \mathbf{R}^{C_m \times H \times W}$ 、 $G \in \mathbf{R}^{C_m \times H \times W}$  分别为输入、输出的特征映射,  $f_{ij} \in \mathbf{R}^{C_m}$ 、 $g_{ij} \in \mathbf{R}^{C_m}$  分别为  $F$ 、 $G$  对应像素  $(i, j)$  的特征张量。

在第 1 阶段, 输入特征通过 3 个  $1 \times 1$  卷积后可以获得  $3 \times N$  个特征图。在第 2 阶段, 分别执行自注意力和卷积操作。自注意力模块将产生的特征图分为  $N$  组, 其中每组均包含来自之前  $1 \times 1$  卷积的 3 个特征, 对应的 3 个特征映射分别为查询、键和值  $(q, k, v)$ , 自注意力模块计算过程可表示为

$$g_{ij} = \parallel \left[ \sum_{l=1}^N \mathbf{A}(q_{ij}^l, k_{ab}^l) v_{ab}^l \right], \quad (4)$$

式中:  $\parallel$  为  $N$  个注意头输出的拼接;  $N_k(i, j)$  为以  $(i, j)$  为中心、空间范围为  $k$  的局部像素区域;  $\mathbf{A}(q_{ij}^l, k_{ab}^l)$  为  $N_k(i, j)$  内特征对应的注意权重。

对核大小为  $k$  的卷积模块而言, 可以采用 FC 层生成  $k^2$  个特征映射, 所使用卷积核大小  $k=3$ , 通过移动和聚合生成的特征, 从局部接受域收集信息, 计算过程可表示为

$$g_{ij}^{(p, q)} = K_{p, q} f_{i+p-\lfloor k/2 \rfloor, j+q-\lfloor k/2 \rfloor} = \text{Shift}(g_{ij}^{\sim(p, q)}, p - \lfloor k/2 \rfloor, q - \lfloor k/2 \rfloor), \quad (5)$$

$$g_{ij} = \sum_{p, q} g_{ij}^{(p, q)}, \quad (6)$$

式中: Shift 为移位操作;  $K_{p, q} \in \mathbf{R}^{C_m \times C_m}$ ;  $p, q \in \{0, 1, \dots, k-1\}$  为关于核位置  $(p, q)$  的核权值;  $\lfloor \cdot \rfloor$  为向下取整符号。

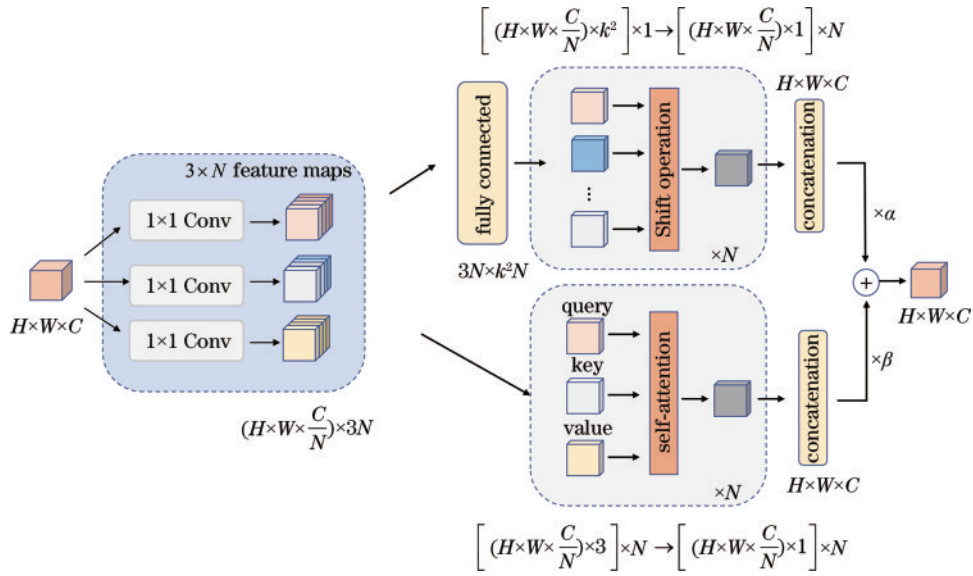


图 4 AC-Block 的结构

Fig. 4 Structure of the AC-Block

最后,将两个模块的输出相加,并由两个可学习标度  $\alpha$ 、 $\beta$  控制强度得到最后的输出  $F_{out}$ ,  $F_{out}$  可表示为

$$F_{out} = \alpha F_{att} + \beta F_{Conv}, \quad (7)$$

式中,  $F_{Conv}$  为卷积模块输出;  $F_{att}$  为自注意力模块输出。

### 2.4 多标签损失函数

使用 ChestX-ray14 数据集,该数据集中每幅 CXR 图像的标签可表示为一个 14 维向量  $Y_0 = [y_1, y_2, \dots, y_n]$ , 其中  $n = 14$  表示 14 种疾病。该数据集中许多病理在外观上相似度较高且样本数目存在类不平衡等问题,因此采用锚损失函数<sup>[22]</sup>来计算每种疾病的损失,即

$$L_i = -y_i \ln \hat{y}_i - (1 - y_i) (1 + \hat{y}_i - y_i)^\gamma \ln (1 - \hat{y}_i), \quad (8)$$

式中:  $L_i$  为第  $i$  类疾病的损失函数;  $y_i$  为第  $i$  种疾病的真实标签;  $\hat{y}_i$  为网络预测第  $i$  种疾病患病的概率(即预测值);  $(1 + \hat{y}_i - y_i)^\gamma$  为调制器,  $\gamma$  为超参数, 设置  $\gamma = 2$ ,  $y_i$  为正确类别的预测值, 若同时患有多种疾病, 则取最大值。通过引入锚损失函数, 可以提高对较难分类疾病的权重, 降低对较易分类疾病的权重, 这种调节可使网络根据疾病分类的难易程度更有效地进行优化, 以得到对疾病分类能力更强的网络模型。

## 3 实验结果分析

### 3.1 数据集和训练参数设置

本实验采用大型多标签数据集 ChestX-ray14, 此数据集共包含患有 14 种不同疾病的 112120 幅胸片前视图。选取其中 60631 幅图像标记为未患病的 CXR 图像, 其余图像均被标记为含有一种或多种胸部疾病。其中, 疝气、肺炎、纤维化等疾病样本较少, 而浸润、积液等疾病的样本较多, 这种样本数量的不平衡加大了模型分类难度。图 5 为病变区域的标注图, 其中矩形

框根据数据集中医生标注数据描绘而成, 可见不同疾病病灶位置、大小、识别难易度均有不同。

将数据集按照 7:1:2 的比例划分为训练集、验证集和测试集, 避免同一样本重复出现。在训练过程中, 将每幅 CXR 原图尺寸为 1024 pixel  $\times$  1024 pixel 的灰度图按双线性插值法调整为 256 pixel  $\times$  256 pixel, 为降低计算复杂度, 再随机选取中心点剪裁为 224 pixel  $\times$  224 pixel, 并通过执行随机旋转和水平翻转操作来实现数据增强。

实验采用的系统是 Ubuntu 16.04 操作系统, 中央处理器 (CPU) 配置为 Intel<sup>®</sup> Core i9-9900X 3.5 GHz, 且使用 4 张 Nvidia RTX2080Ti (11 GB) 图形处理器 (GPU) 进行并行模型训练, DA-Net 模型在 PyTorch 深度学习框架上实现, 使用 Adam 优化器优化网络参数。在参数设置上, 最大训练 epoch 通过验证集来确定, 在验证集达到损失稳定且实现最低损失时停止训练, 通过实验可得, epoch 设置为 40 轮。初始学习率设置为 0.001, 每经过 10 轮 epoch 训练, 将学习率乘以 0.9 进行衰减。为提高实验效率、增大显存利用率, 将批处理量大小设置为 32。

### 3.2 评价指标

将胸部疾病分类定义为多标签分类问题的 14 维二元分类任务, 即每幅 CXR 图像对每一种疾病都只有含该标签与不含该标签两种情况, 采用受试者操作特征 (Receiver operating characteristic, ROC) 曲线以及 ROC 曲线下的面积 (AUC) 来评估所提算法对每种疾病的识别能力。ROC 曲线的横坐标为假阳性率 (False positive rate, FPR)、纵坐标为真阳性率 (True positive rate, TPR), FPR 为真实标签“0”和预测“1”的样本所占比例, TPR 为真实标签“1”和预测“1”的样本所占比例, 具体计算可表示为

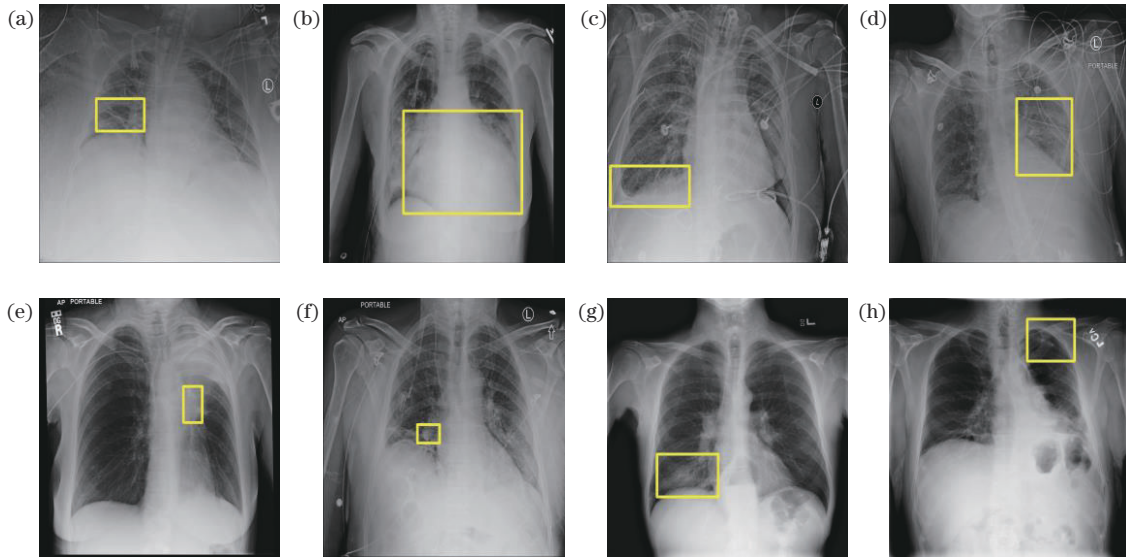


图5 ChestX-ray14数据集中的胸部疾病X射线标注图像。(a)肺不张;(b)心脏肿大;(c)积液;(d)浸润;(e)肿块;(f)结节;(g)肺炎;(h)气胸

Fig. 5 X-ray annotated images of thoracic disease in the ChestX-ray14 dataset. (a) Atelectasis; (b) cardiomegaly; (c) effusion; (d) infiltration; (e) mass; (f) nodule; (g) pneumonia; (h) pneumothorax

$$R_{FP} = \frac{f_{FP}}{f_{FP} + f_{TN}}, \quad (9)$$

$$R_{TP} = \frac{f_{TP}}{f_{TP} + f_{FN}}, \quad (10)$$

式中:  $R_{FP}$ 、 $R_{TP}$  分别为 FPR、TPR 的值;  $f_{TP}$ 、 $f_{TN}$  为正确预测的阳性和阴性样本数;  $f_{FP}$  和  $f_{FN}$  为错误预测的阳性和阴性样本数。因此, ROC 曲线越接近左上角, 网络的分类性能越佳。在 ROC 曲线中, 函数  $f(x) = x$  表示随机结果, 代表分类器性能最低。AUC 与模型对该疾病的分类准确度呈正相关, AUC 越大, 对应疾病的分类准确度越高, 即模型的分类能力越强。

### 3.3 结果与分析

所提算法在 ChestX-ray14 数据集上的实验结果如图 6 所示, 图 6 分别展示每种疾病的分类 ROC 曲线与 14 种疾病的平均 ROC 曲线, 右下角给出每种疾病具体的 AUC。由图 6 可知, 所有曲线皆位于函数  $f(x) = x$  上方, 且整体位于整个图的左上角位置, 表明所提算法对胸部疾病的整体分类准确率较高, 该网络模型具有良好的分类性能。且 14 条曲线在坐标轴上的分布较为密集和紧凑, 这表明所提算法可实现对每种疾病分类准确度的平衡。

在通过 ROC 曲线整体分析算法性能的基础上, 将所提算法与对比算法 (Ma 等<sup>[15]</sup>、Zhang 等<sup>[23]</sup>、Guan 等<sup>[14, 16]</sup>、Chen 等<sup>[2]</sup>、Shao 等<sup>[24]</sup>、Wang 等<sup>[17]</sup>) 进行比较, 实验结果如表 1 所示, 其中每种病理的最大 AUC 加粗显示。Wang 等<sup>[17]</sup> 的方法是当前最先进的, 首先, 从算法整体的分类性能来看, 所提方法对 14 种胸部疾病准确率的平均值超过其他 7 种方法, 平均 AUC 达到 0.839, 其与目前最先进的, 方法相比提升 0.013。其

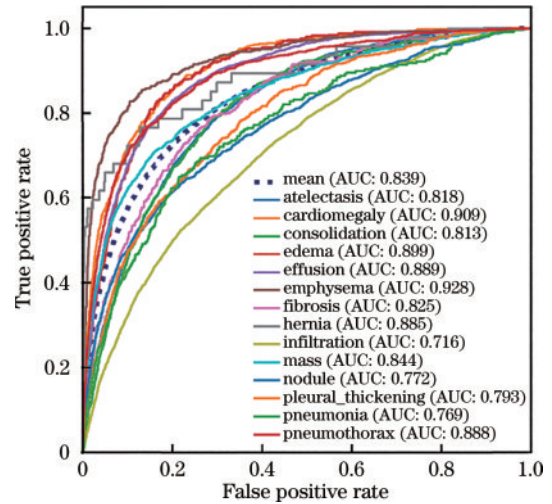


图6 ChestX-ray14数据集上的ROC曲线和AUC

Fig. 6 ROC curve and AUC value of 14 diseases on ChestX-ray14 dataset

次, 从每种疾病的表现结果来看, 与 Wang 等<sup>[17]</sup> 的方法相比, 所提方法在肺结节、肺气肿、纤维化以及疝气等 4 种胸部疾病上的分类准确率较低, 这种情况可能是数据集样本数量过少, 导致所提网络没有充分地学习相关的判别特征 (数据集中只有 227 个疝气样本, 1686 个纤维化样本)。与前 7 种对比方法相比, 所提方法对剩余的 10 种胸部疾病的分类准确度均有不同程度的提升。其中, 对具有复杂病理的疾病, 如肺炎, 其病理面积较大, 病灶片状模糊, 边界不清晰, 所提方法的平均 AUC 为 0.769, 检测结果得到提升; 水肿是在病发初期难以检查出的疾病, 所提方法的平均 AUC 为 0.899, 检测结果得到提升。同时, 对通常与其他病理关联的疾病如肺不张、肺实变、水肿等, 所提方法的检测结果

表 1 不同胸部疾病分类算法在 ChestX-ray14 数据集上的比较

Table 1 Comparison of different thoracic disease classification algorithms on the ChestX-ray14 dataset

Disease	Average AUC							
	Ma et al. <sup>[15]</sup>	Zhang et al. <sup>[23]</sup>	Guan et al. <sup>[14]</sup>	Guan et al. <sup>[16]</sup>	Chen et al. <sup>[2]</sup>	Shao et al. <sup>[24]</sup>	Wang et al. <sup>[17]</sup>	Ours
Atelectasis	0.763	0.785	0.781	0.785	0.785	0.816	0.779	<b>0.818</b>
Cardiomegaly	0.884	0.876	0.880	0.899	0.887	0.866	0.895	<b>0.909</b>
Effusion	0.816	0.863	0.829	0.835	0.831	0.870	0.836	<b>0.889</b>
Infiltration	0.679	0.673	0.702	0.699	0.703	0.695	0.710	<b>0.716</b>
Mass	0.801	0.804	0.834	0.838	0.833	0.835	0.834	<b>0.844</b>
Nodule	0.729	0.730	0.773	0.775	<b>0.798</b>	0.768	0.777	0.772
Pneumonia	0.710	0.742	0.729	0.738	0.731	0.755	0.737	<b>0.769</b>
Pneumothorax	0.837	0.843	0.857	0.871	0.881	0.869	0.878	<b>0.888</b>
Consolidation	0.744	0.785	0.754	0.763	0.754	0.795	0.759	<b>0.813</b>
Edema	0.841	0.873	0.850	0.850	0.849	0.865	0.855	<b>0.899</b>
Emphysema	0.884	0.858	0.908	0.924	0.930	0.894	<b>0.933</b>	0.928
Fibrosis	0.801	0.775	0.830	0.831	0.833	0.814	<b>0.838</b>	0.825
Pleural thickening	0.754	0.756	0.778	0.776	0.782	0.787	0.791	<b>0.793</b>
Hernia	0.876	0.865	0.917	0.922	0.921	0.903	<b>0.938</b>	0.885
Mean	0.794	0.802	0.816	0.822	0.823	0.824	0.826	<b>0.839</b>

得到改善,且所提方法对具有不同尺度大小的疾病分类均可适应。

### 3.4 消融实验分析

为了评估网络中每个模块的有效性,在 ChestX-ray14 数据集上进行消融实验。在保持基本网络结构完整的前提下,分别移除 ODConv/ECA/SAM/AC-Block 中的一个模块,激活剩下的模块,设置 4 组实验,具体操作分别为:1) 移除 ODConv 模块,采用标准  $3 \times 3$  卷积;2) 移除 AC-Block,采用不包含自注意力的全卷积网络;3) 移除 ECA 模块;4) 移除 SAM,测得每个模型对 14 种胸部疾病的分类 ROC 曲线及 AUC,如图 7 所示。

图 7(a)~图 7(d)分别为移除每个模块之后的实验结果,4 个模块对算法的性能均有提升,当替换 ODConv 采用普通卷积时,所提算法的 AUC 下降 0.029,表明 ODConv 的应用可以获得疾病更详细的特征表达,通过提取图像中病灶的多尺度信息,能有效提升分类准确度;移除 AC-Block 后,AUC 下降 0.024,这表明该模块可以提供全局感受野,通过构造自注意力结合卷积的连接架构,并进行特征互补融合以捕获全局特征和局部特征,从而提高模型性能;移除 ECA 模块和 SAM 后,AUC 分别下降 0.011 和 0.009,表明使用双路注意力模块确实能起到提升特征通道的信息表达能力以及将信息聚焦于疾病区域的作用。实验结果表明,采用的 4 个模块均有助于提升网络对胸部疾病的分类准确度。

### 3.5 复杂度分析

在 ResNet50 的基础上,依次增加每个模块后再与

基准网络进行比较,结果表明每个模块均对该网络产生正向作用,同时采用参数量和浮点运算数(Floating point operations, FLOPs)来评估模型复杂度,如表 2 所示。当 4 个模块组合时 DA-Net 网络的提升效果最为显著,与原始的基准网络相比,该网络平均 AUC 提升 0.075。为了提高卷积运算的特征提取能力,引入 ODConv,平均 AUC 提升 0.034,网络性能得到极大提升,而参数量和计算量仅略微提升。同时,采用 EDPA 机制(包含 ECA 模块和 SAM)将更多的处理资源动态地分配给重要信息,使网络更加关注病灶区域,并且加入该机制后对模型复杂度影响不大。最后将自注意力引入卷积网络,通过扩大感受野来深度挖掘出疾病的潜在语义关系,提高分类准确度。此外,当采用不同批处理大小(Batch size)训练数据时,所使用的训练时间是不一样的。在实验中,当 Batch size 设置为 16 时,训练该网络大约需要 22.00 h;当 Batch Size 设置为 32 时,训练时间减少为原来的 1/3,大约需要 7.10 h。

### 3.6 CheXpert 数据集实验结果

为评估模型的泛化能力,验证 DA-Net 网络在公开数据集 CheXpert 上的性能,实验结果如表 3 所示。在 CheXpert 数据集上,由于训练标签有不确定的设置,可以借助多标签分类中使用的两种常见处理不确定标签的方法<sup>[25]</sup>来合并不确定性标签:1) 将所有不确定的标签替换为 0;2) 将所有不确定的标签替换为 1。在实验中使用 DenseNet121、ResNet50 基本网络作为对比网络,在 0 和 1 策略中,ResNet50 对 5 种病理的平均 AUC 分别为 0.880 和 0.886,使用 DA-Net 时,平均 AUC 分别为 0.901 和 0.908,两种策略下都至少提高

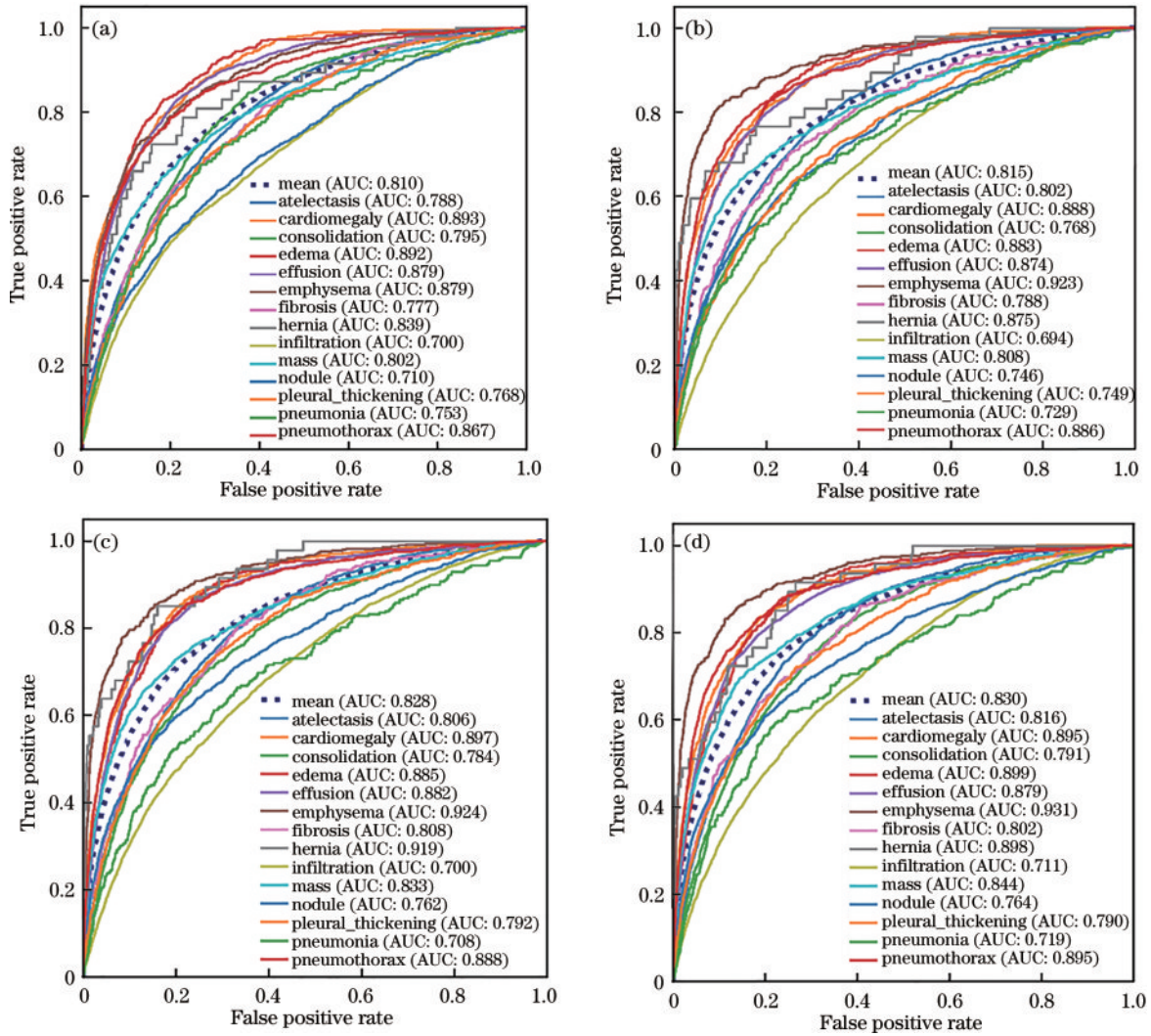


图 7 消融实验结果。(a)移除 ODConv 模块;(b)移除 AC-Block;(c)移除 ECA 模块;(d)移除 SAM

Fig. 7 Results of ablation experiments. (a) Remove ODConv module; (b) remove AC-Block; (c) remove ECA module; (d) remove SAM

表 2 模块组合的实验结果

Table 2 Experimental results for module combinations

Network	Average AUC	Parameter quantity /10 <sup>6</sup>	FLOPs /10 <sup>9</sup>	Training time /h	
				Batch size is 16	Batch size is 32
ResNet50	<b>0.764</b>	25.4	3.82	16.10	4.00
ResNet50+ODConv	0.798	27.9	3.99	18.00	5.20
ResNet50+ODConv+ECA	0.810	28.8	4.01	18.60	5.50
ResNet50+ODConv+EDPA	0.815	29.4	4.12	18.70	5.50
ResNet50+ODConv+EDPA+AC-Block	<b>0.839</b>	32.9	6.94	22.00	7.10

0.02, 且当不确定标签设置为 1 时, DA-Net 的表现更好。

从每种疾病上来看, 当不确定标签设置为 0 时, 心脏肿大的表现显著改善, 提升 0.056; 当不确定标签设

置为 1 时, 心脏肿大和水肿的 AUC 提升相对明显, 分别提升 0.036 和 0.025。DA-Net 相比 DenseNet121 和 ResNet50 均有不同程度的提升, 这表明该网络对胸部疾病的分类任务具有较好的泛化能力和鲁棒性。



表 3 DA-Net 在 CheXpert 数据集上的实验结果  
Table 3 Experimental results of DA-Net on CheXpert dataset

Strategy	Method	Atelectasis	Cardiomegaly	Consolidation	Edema	Pleural effusion	Mean
0	ResNet50	0.806	0.833	0.929	0.913	0.921	0.880
	DenseNet121	0.799	0.832	0.927	0.897	0.923	0.875
	Ours	<b>0.818</b>	<b>0.889</b>	<b>0.941</b>	<b>0.927</b>	<b>0.930</b>	<b>0.901</b>
1	ResNet50	0.805	0.855	0.937	0.910	0.923	0.886
	DenseNet121	0.772	0.844	0.942	0.906	0.901	0.873
	Ours	<b>0.827</b>	<b>0.891</b>	<b>0.948</b>	<b>0.935</b>	<b>0.939</b>	<b>0.908</b>

## 4 结 论

提出了一种基于全维度动态卷积的自注意力分类算法用于提取疾病特征,主要有3个方面的创新:加入全维度动态卷积模块,增强了以ResNet50为骨干的网络对病理异常的多尺度判别能力以及对复杂病理特征的表达能力,从而能够识别出不同疾病,有效地提高了分类准确度;将自注意力模块引入CNN,以提供全局感受野,捕获长距离表达,学习多标签疾病之间具有的潜在相关特征;提出高效的通道空间注意力模块,将更多的处理资源分配给重要信息,使网络聚焦于病灶区域,捕捉胸部异常区域的位置变化。最后,通过锚损失函数来缓解数据集的不平衡问题。实验结果表明,所提算法对CXR疾病分类具有较高准确度,平均AUC达到0.839,同时由消融实验可知,各模块的加入均有助于提升模型的性能,所提方法可以有效执行CXR的疾病分类任务。

## 参 考 文 献

- [1] Ouyang X, Karanam S, Wu Z Y, et al. Learning hierarchical attention for weakly-supervised chest X-ray abnormality localization and diagnosis[J]. IEEE Transactions on Medical Imaging, 2021, 40(10): 2698-2710.
- [2] Chen B Z, Zhang Z, Lin J Y, et al. Two-stream collaborative network for multi-label chest X-ray Image classification with lung segmentation[J]. Pattern Recognition Letters, 2020, 135: 221-227.
- [3] 邱云飞, 张家欣, 兰海, 等. 融合张量合成注意力的改进ResNet图像分类模型[J]. 激光与光电子学进展, 2023, 60(6): 0610008.  
Qiu Y F, Zhang J X, Lan H, et al. Improved ResNet image classification model based on tensor synthesis attention[J]. Laser & Optoelectronics Progress, 2023, 60(6): 0610008.
- [4] Anthimopoulos M, Christodoulidis S, Ebner L, et al. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1207-1216.
- [5] Kumar P, Grewal M, Srivastava M M. Boosted cascaded convnets for multilabel classification of thoracic diseases in chest radiographs[M]//Campilho A, Karray

F, ter Haar R B. Image analysis and recognition. Lecture notes in computer science. Cham: Springer, 2018, 10882: 546-552.

- [6] 赵清一, 孔平, 闵建中, 等. 肺结节检测与分类的深度学习方法综述[J]. 生物医学工程学杂志, 2019, 36(6): 1060-1068.  
Zhao Q Y, Kong P, Min J Z, et al. A review of deep learning methods for the detection and classification of pulmonary nodules[J]. Journal of Biomedical Engineering, 2019, 36(6): 1060-1068.
- [7] Gilanie G, Attique M, Naweed S, et al. Object extraction from T2 weighted brain MR image using histogram based gradient calculation[J]. Pattern Recognition Letters, 2013, 34(12): 1356-1363.
- [8] Liskowski P, Krawiec K. Segmenting retinal blood vessels with deep neural networks[J]. IEEE Transactions on Medical Imaging, 2016, 35(11): 2369-2380.
- [9] Rajinikanth V, Satapathy S C, Fernandes S L, et al. Entropy based segmentation of tumor from brain MR images-a study with teaching learning based optimization [J]. Pattern Recognition Letters, 2017, 94: 87-95.
- [10] Bülow H, Dooley L, Wermser D. Application of principal axes for registration of NMR image sequences [J]. Pattern Recognition Letters, 2000, 21(4): 329-336.
- [11] 赵卫东, 刘俊德, 王慢慢, 等. 基于改进AKAZE算法的快速图像配准方法[J]. 激光与光电子学进展, 2023, 60(6): 0610007.  
Zhao W D, Liu J D, Wang M M, et al. Fast image registration method based on improved AKAZE algorithm [J]. Laser & Optoelectronics Progress, 2023, 60(6): 0610007.
- [12] Liao R, Miao S, de Tournemire P, et al. An artificial agent for robust image registration[C]//Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA. New York: ACM Press, 2017: 4168-4175.
- [13] Wang X S, Peng Y F, 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), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3462-3471.
- [14] Guan Q J, Huang Y P, Luo Y W, et al. Discriminative feature learning for thorax disease classification in chest X-ray images[J]. IEEE Transactions on Image Processing,

- 2021, 30: 2476-2487.
- [15] Ma Y B, Zhou Q H, Chen X S, et al. Multi-attention network for thoracic disease classification and localization [C]//ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 12-17, 2019, Brighton, UK. New York: IEEE Press, 2019: 1378-1382.
- [16] Guan Q J, Huang Y P. Multi-label chest X-ray image classification via category-wise residual attention learning [J]. Pattern Recognition Letters, 2020, 130: 259-266.
- [17] Wang H Y, Wang S S, Qin Z B, et al. Triple attention learning for classification of 14 thoracic diseases using chest radiography[J]. Medical Image Analysis, 2021, 67: 101846.
- [18] Tan H, Dong S J. Pixel-level concrete crack segmentation using pyramidal residual network with omni-dimensional dynamic convolution[J]. Processes, 2023, 11(2): 546.
- [19] 李家昱, 陈明惠, 杨瑞君, 等. 糖尿病视网膜病变眼底图像筛查研究[J]. 中国激光, 2022, 49(11): 1107001.  
Li J Y, Chen M H, Yang R J, et al. Study on fundus image screening of diabetic retinopathy[J]. Chinese Journal of Lasers, 2022, 49(11): 1107001.
- [20] Pan X R, Ge C J, Lu R, et al. On the integration of self-attention and convolution[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 805-815.
- [21] 周涛, 叶鑫宇, 刘凤珍, 等. 基于轻量化方向 Transformer 模型的肺炎 X 光片辅助诊断[J]. 光学学报, 2023, 43(13): 1310001.
- Zhou T, Ye X Y, Liu F J, et al. Aided diagnosis of pneumonia radiographs based on the lightweight direction Transformer model[J]. Acta Optica Sinica, 2023, 43(13): 1310001.
- [22] Ryou S, Jeong S G, Perona P. Anchor loss: modulating loss scale based on prediction difficulty[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2020: 5991-6000.
- [23] 张智睿, 李镔, 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 Z R, 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.
- [24] 邵凌云, 李镔, 关欣, 等. 基于高效通道注意力的胸部 X 光片疾病分类算法[J]. 激光与光电子学进展, 2023, 60(12): 1217001.  
Shao L Y, Li Q, Guan X, et al. Efficient channel attention-based disease classification algorithm for chest radiographs [J]. Laser & Optoelectronics Progress, 2023, 60(12): 1217001.
- [25] Irvin J, Rajpurkar P, Ko M, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 590-597.