

基于多尺度注意力信息复用网络的 胸片图像分类

张瑞峰, 郭芳兆, 李 镨

(天津大学 微电子学院, 天津 300072)

摘 要: 针对胸部 X 射线图像的病变区域辨识度低、准确捕捉病变空间位置难等问题, 提出了一种有利于提高胸片图像分类精度的多尺度注意力信息复用网络。首先, 通过引入多路空间信息复用模块, 增强疾病部位在特征图及通道之间的位置联系; 其次, 通过多尺度融合注意力模块, 整合多尺度图像特征信息, 自动捕捉病灶位置变化, 以实现关键病理信息的灵活关注; 最后, 通过非对称移位焦点损失函数, 缓解胸部疾病样本分布不平衡的问题。在公开数据集 ChestX-ray14 和 CheXpert 上的多组实验表明: 本文网络在两个数据集上的平均 AUC 值分别达到 0.847 和 0.901, 优于近年来较为先进的网络模型, 表明该网络能有效地提高胸部疾病的分类精度。

关键词: 计算机应用技术; 胸部 X 光图像分类; 空间信息复用; 多尺度注意力; 非对称移位焦点损失

中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5497(2025)11-3686-11

DOI: 10.13229/j.cnki.jdxbgxb.20240222

Chest X-ray images classification based on multi-scale attention information multiplexing network

ZHANG Rui-feng, GUO Fang-zhao, LI Qiang

(School of Microelectronics, Tianjin University, Tianjin 300072, China)

Abstract: To address issues such as low recognition of lesion areas in chest X-ray images and the difficulty in accurately capturing the spatial positions of lesions, a multi-scale attention information multiplexing network that helps improve the classification accuracy of chest X-ray images was proposed in this paper. Firstly, by introducing multiple spatial information multiplexing blocks, the network enhances the positional connections between disease regions on feature maps and across channels; Secondly, through a multi-scale integration attention blocks, the network integrates multi-scale image feature information to automatically capture disease location variations and flexibly focus on key pathological information; Finally, the problem of imbalanced distribution of chest disease samples was alleviated by using an

收稿日期: 2024-03-05.

基金项目: 国家自然科学基金项目(62071323); 超声医学工程国家重点实验室开放课题项目(2022KFKT004); 天津市自然科学基金重点项目(22JCZDJC00220).

作者简介: 张瑞峰(1974-), 男, 副教授, 博士. 研究方向: 智能信息处理. E-mail: zhangruifeng@tju.edu.cn

通信作者: 李镨(1974-), 男, 教授, 博士. 研究方向: 智能信息处理. E-mail: liqiang@tju.edu.cn

asymmetric shift focus loss function. Multiple experiments on the publicly available datasets ChestX-ray14 and CheXpert have shown that the average area under curve (AUC) value of the proposed network on two datasets reached 0.847 and 0.901 respectively, which is superior the more advanced network models in recent years. This indicates that the network can effectively improve the classification accuracy of chest diseases.

Key words: computer application technology; chest X-ray image classification; spatial information multiplexing; multi-scale attention; asymmetric shift focus loss function

0 引言

胸部疾病十分常见且种类繁多,已成为威胁现代人健康的主要疾病之一。X射线胸片(Chest X-ray, CXR)是目前最常用的胸部疾病放射检查手段,但其诊断需要高度专业的知识,给医师带来不小的挑战。依靠人工诊疗达到 CXR 图像较高的识别精度非常困难,因此医学人工智能在疾病分类领域具有重要的研究意义^[1]。

近年来,深度学习在医学图像处理领域的研究取得了很大的突破^[2],许多优秀算法不断涌现并持续改进^[3]。2017年,美国国立卫生研究院首次公开大型 CXR 图像数据集 ChestX-ray14, Wang 等^[4]基于该数据集评估了多种卷积网络的诊断性能,如 AlexNet^[5]、ResNet^[6]等。2019年,Irvin 等^[7]发布了大型 CXR 图像数据集 CheXpert,为网络模型泛化能力的验证提供了更多依据。Jiang 等^[8]提出适用于多标签 CXR 图像分类的金字塔视觉转换器^[9]MXT,它可以通过自注意力机制捕捉 CXR 图像中的短距离和远距离视觉信息。胡锦涛等^[10]提出一种可形变 Transformer 辅助的胸部疾病诊断网络,提高了不同疾病在影像区域重叠场景下的特征区分度。

此外,许多研究人员将注意力机制应用到疾病分类领域。该机制通过调整可用的资源配置,引导网络模型聚焦于 CXR 图像中的病变区域。Wang 等^[11]提出三重注意网络 A³Net,用于通道级、元素级及规模级的注意学习。Zhu 等^[12]提出逐像素分类和注意力网络(Pixel-wise classification and attention network, PCAN),为疾病分类提供了可解释性支持。Chen 等^[13]提出基于金字塔卷积和洗牌注意力模块的胸部疾病分类和 COVID-19 检测新网络 PCSANet。

然而,胸部疾病图像分类领域的研究仍然面临着一些困难。首先, CXR 图像中病灶区域的纹

理、位置等特征多样且复杂,上述部分网络虽然采用了注意力机制,但在对 CXR 图像进行病变分析时,缺乏对不同层次、不同维度特征图信息的综合考虑和充分利用。其次,同一 CXR 图像可能同时存在多种胸部疾病,增加了胸部疾病识别和分类的复杂性,现有网络的多尺度特征提取能力有限,难以捕捉代表性特征用于胸部疾病的细粒度分类。再次, CXR 图像中的病灶区域特征辨识度低,许多深度学习网络在提取特征过程中易受到图像噪声和不相关区域的影响。最后,数据集中各类疾病的样本数目存在较为严重的类不平衡问题。

针对上述问题,本文提出多尺度注意力信息复用网络(Multi-scale attention information multiplexing network, MIM-Net)。首先,提出多路空间信息复用(Multiple spatial information Multiplexing, MSIM)模块,借助超像素操作^[14]和亚像素操作构建空间信息多路复用块,增强网络跨通道和空间的信息交流,充分结合特征图的多尺度特征进行融合学习。其次,采用融合空间注意力机制和通道注意力机制的多尺度融合注意力(Multi-scale integration attention, MIA)模块,增加病灶相关区域的权重,同时抑制不相关背景区域干扰,通过网络的深层信息和浅层信息融合提取病灶的多尺度特征。最后,使用非对称移位焦点损失函数,缓解因简单易分的无疾病样本过多导致的无效学习问题。

1 本文方法

本文提出多尺度注意力信息复用网络 MIM-Net,其结构如图 1(a)所示。

输入图像首先通过 3 个 3×3 卷积层和 1 个池化层提取初始特征,捕捉图像中的局部特征和结构信息,相较于采用 1 个 7×7 卷积,该设计在相同感受野下增强了网络的非线性能力。初始特征提

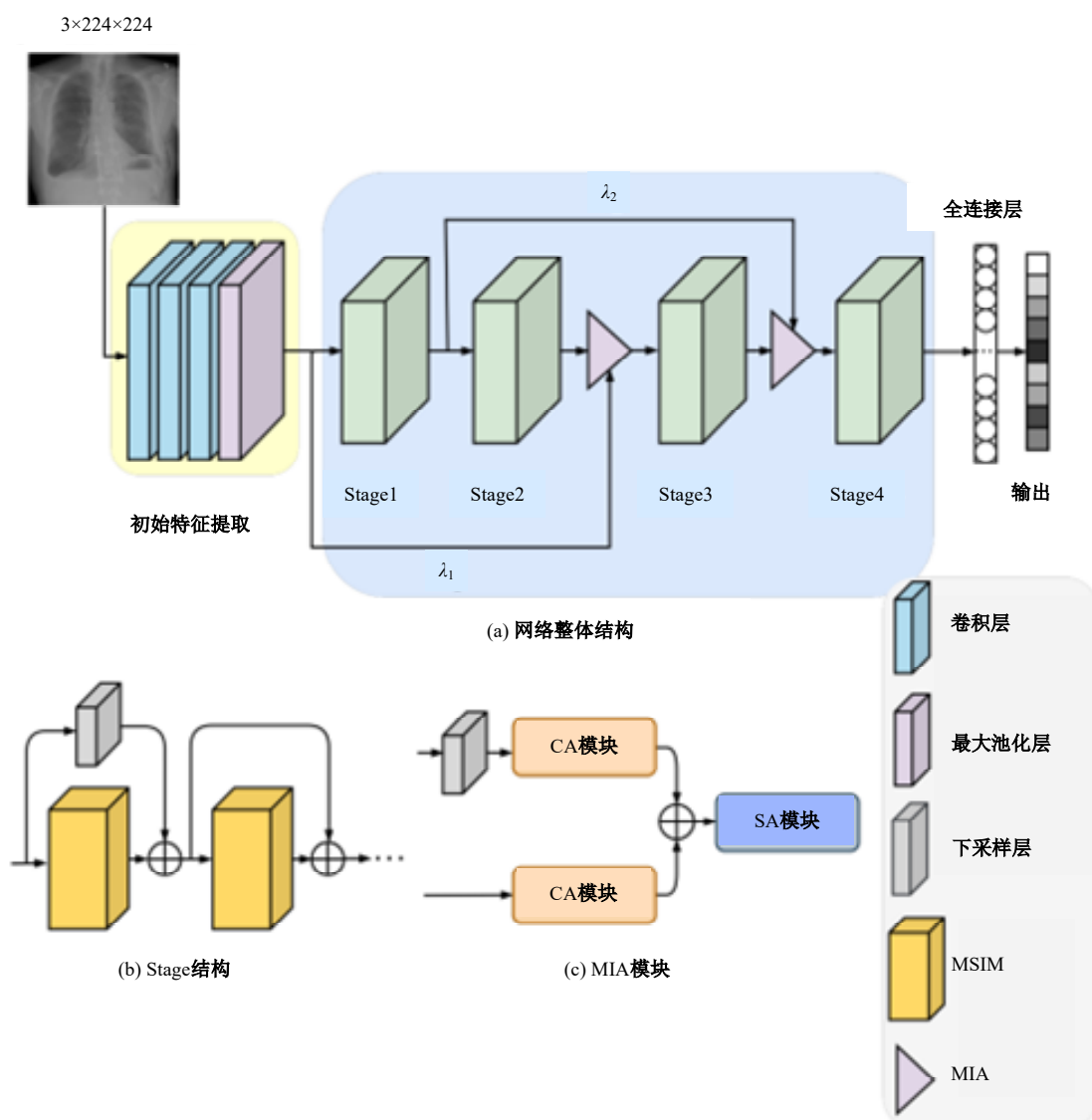


图 1 MIM-Net 模型架构

Fig. 1 MIM-Net model architecture

取后的特征图进入 Stage1(S_1)~Stage4(S_4)4个阶段,各阶段包含的 MSIM 模块数量分别为 3、3、4、3。如图 1(b)所示,MSIM 模块在 4 个阶段中采用残差连接方式,在一定程度上缓解了数据集样本数量类不平衡导致的网络过拟合问题。此外,将初始特征提取后的特征图与经 S_2 处理后的特征图同时输入 MIA 模块,其中包含更低级语义信息的特征图乘以超参数 λ_1 后作为 MIA 模块的输入之一,经 MIA 模块处理后的图像作为 S_3 的输入。同样地,将 S_1 和 S_3 阶段处理后的特征图输入 MIA 模块,经 MIA 模块处理后的图像作为 S_4 阶段的输入。超参数 λ 反映了网络训练过程中,较浅层次的语义信息对网络整体分类能力的影响。通过上述设计,网络深层信息和浅层信息得到充分融合,

使网络专注于病灶区域,抑制不相关背景区域干扰。下文将分别介绍 MSIM 模块和 MIA 模块。

1.1 多路空间信息复用模块

1.1.1 超像素和亚像素操作

空间信息复用的关键思想是将多个尺度的空间信息映射到通道中。如图 2(b)所示,给定尺寸为 $C \times H \times W$ 的特征图(其中, C 为通道数, H 和 W 分别为特征图的高度和宽度),定义一个大小为 $r \times r$ 的窗口,将各通道特定位置的特征值映射到输出特征图唯一的 $r \times r$ 窗口中,该操作即超像素(Super-pixel)操作。它将尺寸为 $C \times H \times W$ 的特征图映射成尺寸为 $C/r^2 \times rH \times rW$ 的特征图,实现了特征图通道信息与空间信息的复用,并通过卷积操作进行特征学习。

亚像素(Sub-pixel)操作为超像素操作的逆操作,如图 2(c)所示。定义一个大小和步长均为 $r \times r$ 的窗口,将窗口中的特征映射到各个通道中,每个窗口的特征值对应通道中唯一的特征位置。总的来说,亚像素操作可将尺寸为 $C \times H \times W$ 的特征图映射为 $r^2 C \times H/r \times W/r$ 的特征图,使下采样的空间信息能够与通道信息复用,增大了后续卷积操作的感受野。

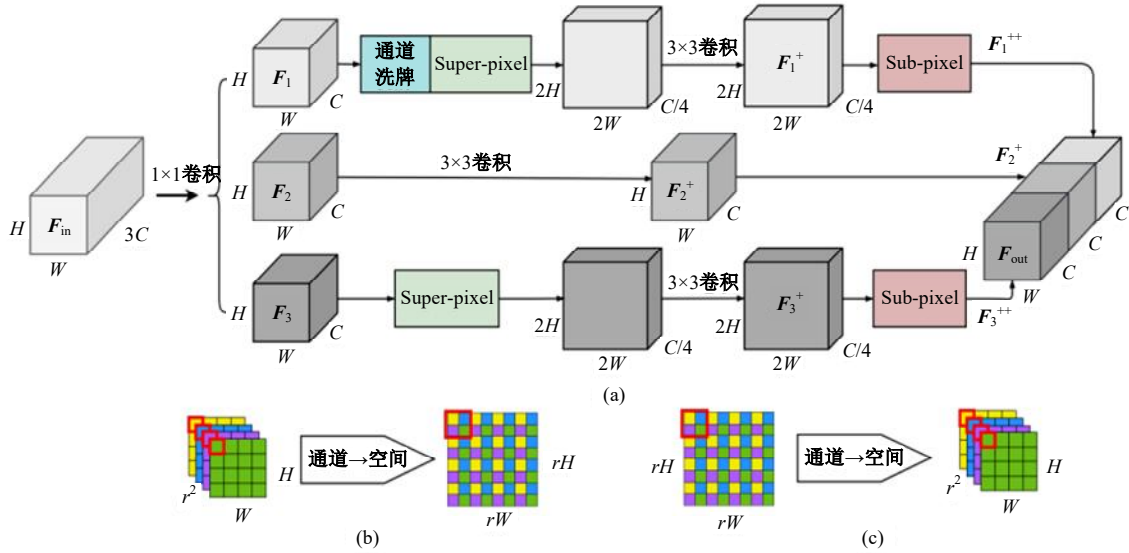


图 2 MSIM 模块

Fig. 2 Multiple spatial information multiplexing block

输入特征图 F_{in} 首先通过 1 个 1×1 卷积,在提取特征信息的同时将通道数降为原来的 $1/3$,将输出的特征图复制 3 份,分别记为 F_1 、 F_2 和 F_3 。 F_1 先通过通道洗牌操作打乱通道顺序,再经过 $r=2$ 的超像素操作和 3×3 卷积实现空间信息复用,得到 F_1^+ 。 F_3 先进行 $r=2$ 的超像素操作,再通过 3×3 卷积进行特征学习,得到 F_3^+ 。 F_1^+ 和 F_3^+ 最终分别通过 $r=2$ 的亚像素操作还原为原始的特征图尺寸,得到 F_1^{++} 和 F_3^{++} 。 F_1^{++} 、 F_3^{++} 与 F_2 经过 3×3 卷积后的输出特征图 F_2^+ 进行逐通道拼接,得到最终的输出特征图 F_{out} ,其表达式为:

$$F_{out} = [F_1^{++}, F_2^+, F_3^{++}] \quad (1)$$

现有的多尺度处理通常基于跨不同层实现,而 MSIM 模块通过超像素操作和亚像素操作,增强了对同一层内多尺度信息的利用,并结合通道洗牌操作进一步增强了跨通道的信息流交互,在不引入额外参数的情况下,实现了空间和通道上的多路信息流复用。MSIM 模块帮助网络捕捉具有代表性的特征,用于胸部疾病的细粒度分类,提

1.1.2 多路空间信息复用模块

不同疾病的病变在 CXR 图像上的大小和位置存在显著差异,这就要求网络具有可靠的多尺度特征提取能力,以充分结合特征图的多尺度特征进行融合学习。本文提出的 MSIM 模块不仅实现了特征图通道信息与空间信息的复用,还通过通道洗牌和超像素操作的结合,建立了更多通道间的信息交流。MSIM 的具体实现方式如图 2(a) 所示。

升了网络的表现能力和预测性能。

1.2 MIA 模块

网络训练阶段应更加关注病变相关的特征信息。通道注意力机制^[15]可以对通道间的依赖关系进行建模,增加病变相关通道的权重,并对无关通道进行抑制。此外,CXR 图像上通常有多个病变区域,且大小、形状均不同,因此网络在训练过程中,不仅要关注特征图的全局信息,还要关注病灶在特征图中的空间位置信息。因此,本文提出的多尺度融合注意力(MIA)模块将通道注意力和空间注意力相结合,使网络更多地关注病灶关键位置,减少对无关部分的关注。MIA 模块的结构如图 3 所示。

MIA 模块以中间层特征图 X_1 和包含更高级语义信息的特征图 X_2 作为输入。其中, X_1 在执行后续操作前先进行下采样,得到 X_1' , X_1' 和 X_2 的尺寸相同。首先, X_1' 和 X_2 分别经过通道注意力模块,具体做法为:对输入特征图分别进行全局最大池化和全局平均池化,对特征映射进行两个维度

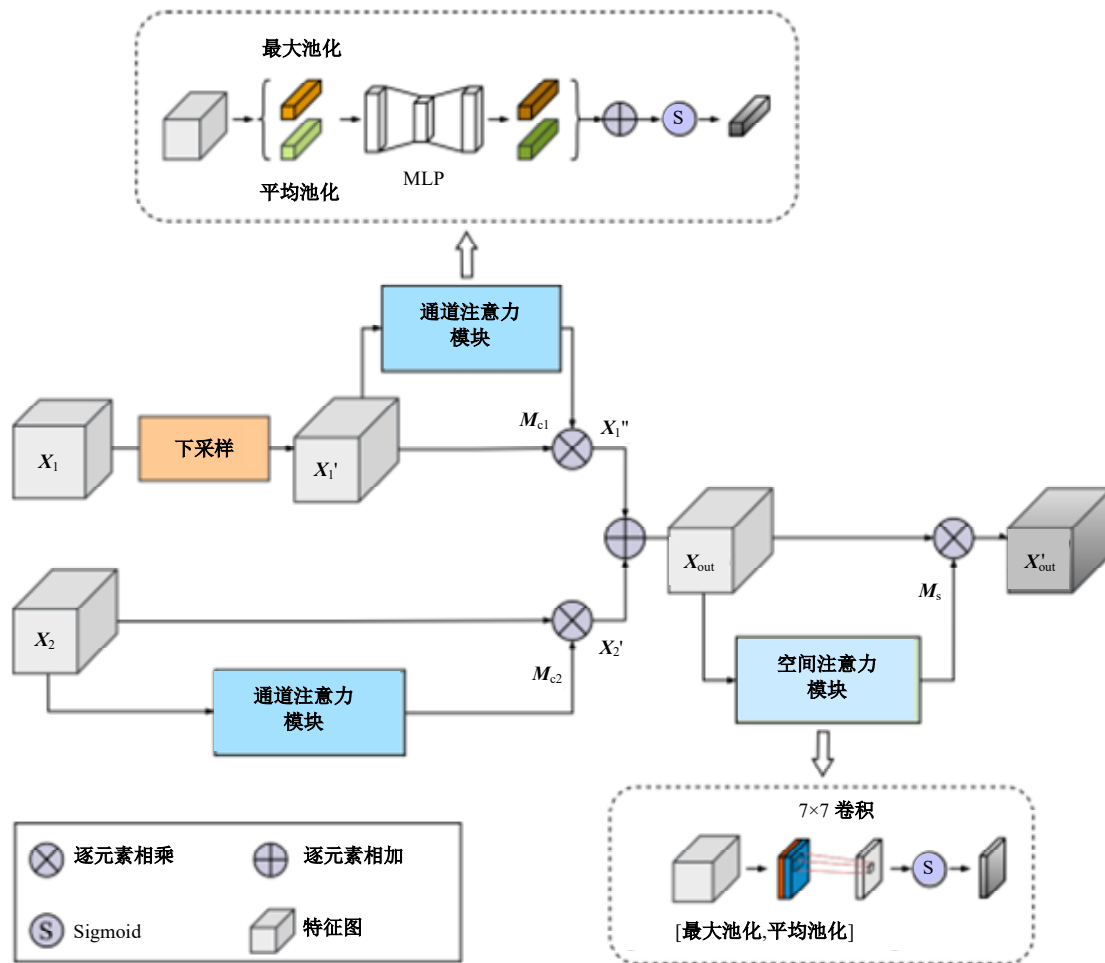


图 3 MIA 模块

Fig. 3 Multi-scale integration attention block

的压缩,获得两种不同维度的特征表达。池化后的特征图各自通过一个多层感知器(Multilayer perceptron, MLP),即先经过一个全连接层将通道数降至原来的 $1/4$,再通过另一个全连接层恢复通道数。两个多层感知器网络的输出在通道维度堆叠后,经 Sigmoid 激活函数将特征图各通道的权重归一化,得到通道注意力的权重矩阵 M_{c1} 、 M_{c2} 。同时,将输入特征图 X_1' 、 X_2 分别与各自的权重矩阵 M_{c1} 、 M_{c2} 相乘,得到 X_1'' 和 X_2' 。

其次,将 X_1'' 和 X_2' 进行逐元素相加,实现低层语义信息和高层语义信息的特征融合,得到 X_{out} 。对 X_{out} 执行空间注意力操作,具体做法为:在通道维度上对 X_{out} 分别进行最大池化和平均池化,将池化后的特征图在通道维度堆叠,通过 1 个 7×7 卷积核融合通道信息,生成通道数为 1 的权重矩阵,卷积后的权重矩阵经 Sigmoid 函数实现空间权重归一化,将归一化后的权重矩阵 M_s 与 X_{out} 相乘得到 MIA 模块的最终输出 X'_{out} 。MIA 模块将

网络训练过程中的深层信息和浅层信息相融合,能够提取病灶的多尺度信息,挖掘疾病间的潜在语义关系,自动捕捉病灶位置变化,从而提高疾病分类的准确性。

1.3 非对称移位焦点损失

样本分布不均会影响多标签分类任务准确率的提高。在 ChestX-ray14 数据集中,疝气、肺炎等病理正样本(疾病样本)数量较少,简单易分的负样本(无疾病样本)数量过多,使得网络优化方向倾向于无疾病样本的无效学习,只能分辨出有无疾病,无法分辨具体的疾病种类,且疾病间存在特征差异,学习难易程度也不同。为解决这一问题,本文采用优化后的焦点损失函数,将其称为非对称移位焦点损失函数。受 Ridnik 等^[16]的启发,当使用焦点损失函数进行多标签训练时,若设置较高的 γ 以降低简单易分负样本的贡献权重,则会消除罕见正样本的贡献权重,因此损失函数首先对正样本和负样本进行解耦,即:

$$L = \begin{cases} \frac{1}{K} \sum_{k=1}^K (1 - p_k)^{\gamma^+} \log p_k, y_k = 1 \\ \frac{1}{K} \sum_{k=1}^K (p_k)^{\gamma^-} \log (1 - p_k), y_k = 0 \end{cases} \quad (2)$$

式中: K 为类别数量; y_k 为图片的疾病类别; p_k 为网络预测图片类别的正确概率。定义 γ^+ 和 γ^- 替代焦点损失函数中的 γ , 设置 $\gamma^+ < \gamma^-$, 以更好地控制正样本和负样本对损失函数的贡献, 从而在通过 γ^- 降低易分负样本权重的同时, 不会过多消除罕见正样本的贡献权重。本文设置 $\gamma^+ = 0$, 使正样本产生简单的交叉熵损失; 设置 $\gamma^- = 2$, 控制非对称聚焦水平。

此外, 焦点损失函数通过设置参数 γ 调整网络对难分类样本的关注度, 即 γ 值越大, 焦点损失函数对易分负样本的权重衰减越大。然而, 由于 ChestX-ray14 数据集中样本的不平衡程度很高, 其简单易分的无关负样本过多, 普通焦点损失函数只能通过设置高 γ 对其权重进行一定程度的衰减。因此, 本文采用概率移位法设置概率偏移因子 s , 对非常易分的负样本执行硬阈值分割, 即当它们的概率非常低时完全丢弃, 使网络更加关注负样本中较难分类的部分, 即:

$$L = \begin{cases} \frac{1}{K} \sum_{k=1}^K (1 - p_k)^{\gamma^+} \log p_k, y_k = 1 \\ \frac{1}{K} \sum_{k=1}^K (p_k^s)^{\gamma^-} \log (1 - p_k^s), y_k = 0 \end{cases} \quad (3)$$

式中: p_k^s 为经概率偏移因子 s 处理后的网络预测图片类别的正确概率。其表达式为:

$$p_k^s = \max(p_k - s, 0) \quad (4)$$

当负样本的预测得分在 $0 \sim s$ 时, 认为该负样本的分类非常容易, 经概率移位运算后阈值为 0, 本文设置 $s = 0.2$ 。非对称移位焦点损失函数可使网络模型根据胸片数据集中的样本分布, 更加关注正负样本中难以分类的部分, 解决了焦点损失易削弱罕见疾病样本权重的问题, 提高了网络的整体分类能力。

2 实验结果与分析

2.1 数据集与数据预处理

本文采用两个公开胸部 X 光片数据集作为评估基准, 分别为美国国立卫生研究院发布的 ChestX-ray14 和斯坦福大学研究人员发布的 CheXpert。

ChestX-ray14 数据集的样本分布信息如图 4 所示。该数据集包含 30 805 名患者的 112 120 张正面视图 X 射线图像, 每张图像均标记有 14 种常见胸腔疾病的一种或多种类型, 具体为肺不张、心脏肿大、积液、渗透、肿块、肺结节、肺炎、气胸、肺实变、水肿、肺气肿、纤维化、胸膜增厚和疝气。由图 4 可以看出, 不同疾病的样本分布存在显著的不均衡性, 这种情况会增加分类网络的训练难度。

CheXpert 数据集包含 65 240 名患者的 224 316 次 X 射线扫描图像, 该数据集中每种疾病类别有阳性、阴性和不确定 3 种标记, 其中不确定表示医生仅通过 X 光片尚无法判断是否患病。由于该数据集不完全开源, 公开的验证集仅包含肺不张、心脏肿大、肺实变、水肿和胸腔积液 5 种疾病的 234 张图像, 因此本文仅对这 5 种疾病的预测精度进行评估, 每种疾病的样本分布如表 1 所示。

在数据预处理阶段, 本文对 CXR 图像做了一系列增强操作。首先, 将图像尺寸缩放至 256×256 , 并随机裁剪为 224×224 ; 其次, 以 0.5 的概率

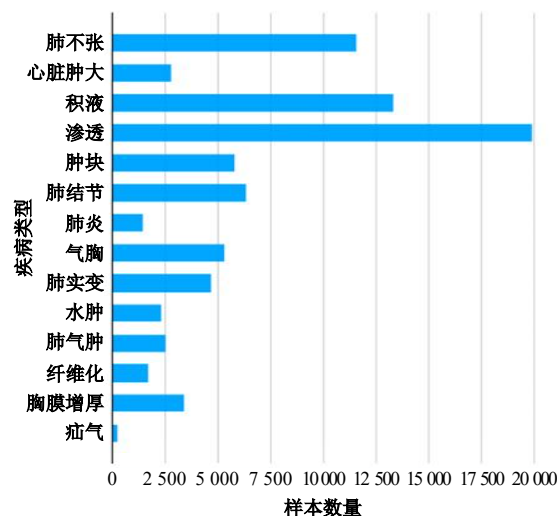


图 4 ChestX-ray14 数据集中各类型图像数量

Fig. 4 Number of images of each type in ChestX-ray14 dataset

表 1 CheXpert 数据集中各类型图像数量

Table 1 Number of images of each type in CheXpert dataset

疾病种类	阳性	不确定	阴性
肺不张	29 333	29 377	165 606
心脏肿大	23 002	6 597	194 717
肺实变	12 730	23 976	187 610
水肿	48 905	11 571	163 840
胸腔积液	75 696	9 419	139 201

水平翻转图像,并在 $[-5^\circ, 5^\circ]$ 范围内随机旋转图像;最后,将图像转换为向量格式,并对RGB 3个通道进行归一化,使其更易被神经网络处理。

2.2 实现细节与评价指标

本文所有实验均在NVIDIA RTX A5000 (24 GB) GPU平台上进行,采用PyTorch开源深度学习框架,初始学习率为0.000 1,每两个轮次学习率衰减为原来的0.9倍。为提高网络的收敛速度和学习能力,实验中的骨干网络先在ImageNet数据集上进行预训练,当验证集的损失不再减少或开始增加时训练停止。

本文采用受试者工作特征(Receiver operating characteristic, ROC)曲线与ROC曲线下面积(Area under ROC curve, AUC)值作为网络识别每种病理能力的评价指标,其中ROC曲线下面积可以反映预测精度,面积越大准确率越高。在ROC曲线中,横坐标假阳率(False positive rate, FPR)表示所有未患病样本中被错误地判断为患病的样本数百分比,纵坐标真阳率(True positive rate)表示所有患病样本中被正确地判断为患病的样本数百分比。FPR和TPR的精确计算公式为:

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (5)$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

式中:TP、FP、TN和FN分别为样本中的真阳(患病样本被正确判断为患病)、假阳(未患病样本被错误判断为患病)、真阴(未患病样本被正确判断为未患病)和假阴(患病样本被错误判断为未患病)。

AUC值越高,表明对应疾病的分类准确度越高,即网络的分类能力越强。

2.3 超参数实验分析

本节将分析MIM-Net中 λ_1 和 λ_2 的取值对网络分类精度的影响。在MIM-Net网络中,超参数 λ 控制了MIA模块中包含更浅层信息的特征图所占权重。 λ_1 为经初始特征提取后的特征图在进入第一个MIA模块前所乘的权重参数, λ_2 为 S_1 阶段处理后的特征图在进入第二个MIA模块前所乘的权重参数。在保证其他设置不变的情况下,设置多组 λ_1 和 λ_2 的不同组合训练网络,实验结果如图5所示。

实验选择0、0.3、0.6、1作为 λ 的候选值,散点颜色对应右边色柱上AUC值大小,反映不同超参

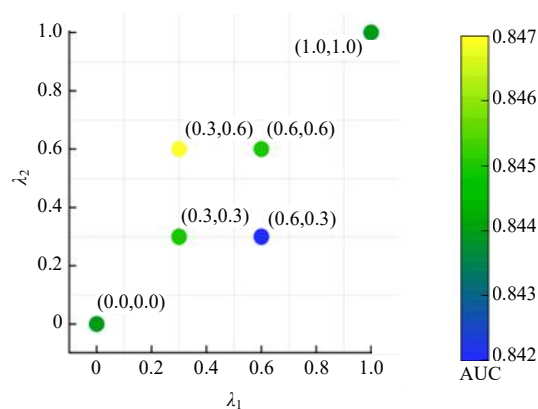


图5 不同 λ 组合的AUC散点图

Fig. 5 AUC scatter plots of different λ combinations

数组组合对分类精度的影响。从图5可以看出, λ 的不同取值对分类结果影响较大,当 $\lambda_1=0.3$ 、 $\lambda_2=0.6$ 时,网络分类效果最佳。 λ 值越大,输出的特征图中浅层语义信息的影响越大,这些浅层特征包含大量病灶空间位置信息,多尺度信息的融合学习有助于提高神经网络对不同疾病病灶的识别性能。

2.4 实验结果与分析

本文将MIM-Net与ChestX-ray14数据集上的先进网络进行比较,通过计算每种疾病类别的AUC值及所有类别的平均AUC值进行评估。从图6可以看出,每种疾病的ROC曲线均位于图中左上方,说明MIM-Net的整体分类性能较好,对所有疾病的预测均有良好表现。

表2为MIM-Net与其他网络的比较结果,对比网络包括Wang等^[4]提出的DCNN、Guan等^[17]

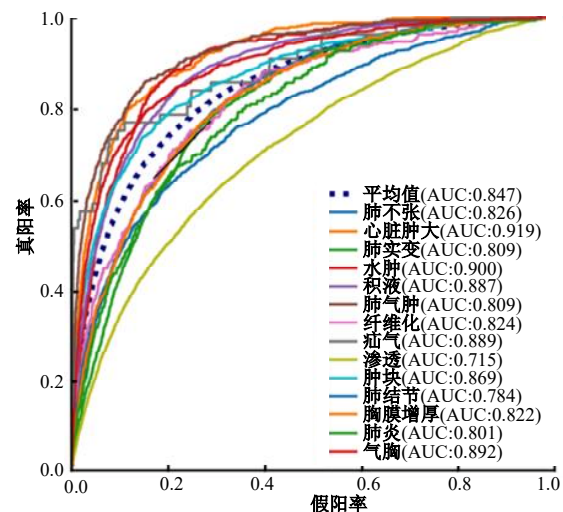


图6 ChestX-ray14数据集中14种疾病的ROC曲线

Fig. 6 ROC curves of 14 diseases on the ChestX-ray14 dataset

表 2 各网络在 ChestX-ray14 数据集上的分类性能对比

Table 2 Comparison of the classification accuracy of each network using the ChestX-ray14 dataset

疾病种类	DCNN	ConsultNet	Deformab-CDAM-D	A ³ Net	PCSA _{Net}	CheXGAT	PCAN	SSGE	MIM-Net
肺不张	0.700	0.785	0.820	0.779	0.807	0.787	0.785	0.792	0.826
心脏肿大	0.810	0.899	0.912	0.895	0.910	0.879	0.897	0.892	0.919
积液	0.759	0.835	0.890	0.836	0.879	0.837	0.837	0.840	0.887
渗透	0.661	0.699	0.714	0.710	0.698	0.699	0.706	0.714	0.715
肿块	0.693	0.838	0.865	0.834	0.824	0.839	0.834	0.848	0.869
肺结节	0.669	0.775	0.772	0.777	0.750	0.793	0.786	0.812	0.784
肺炎	0.658	0.738	0.762	0.737	0.750	0.741	0.730	0.733	0.801
气胸	0.799	0.871	0.903	0.878	0.850	0.879	0.871	0.885	0.892
肺实变	0.703	0.763	0.810	0.759	0.802	0.755	0.763	0.753	0.809
水肿	0.805	0.850	0.896	0.855	0.888	0.851	0.854	0.848	0.900
肺气肿	0.833	0.924	0.914	0.933	0.890	0.945	0.921	0.948	0.924
纤维化	0.786	0.831	0.808	0.838	0.812	0.842	0.817	0.827	0.824
胸膜增厚	0.684	0.776	0.815	0.791	0.768	0.794	0.791	0.795	0.822
疝气	0.872	0.922	0.876	0.938	0.915	0.931	0.943	0.932	0.889
平均 AUC	0.745	0.822	0.840	0.826	0.825	0.827	0.824	0.830	0.847

提出的 ConsultNet、胡锦涛等^[10]提出的 Deformab-CDAM-D、Wang 等^[11]提出的 A³Net、Chen 等^[13]提出的 PCSANet、Lee 等^[18]提出的 CheX-GAT、Zhu 等^[12]提出的 PCAN 及 Chen 等^[19]提出的 SSGE。表中加粗数据表示该方法得到的指标最佳。实验结果表明,本文提出的 MIM-Net 分类能力较对比网络有显著提高,且在心脏肿大、水肿、肺气肿 3 种疾病的检测中效果突出,AUC 值均 > 0.9。与所有对比网络相比,MIM-Net 在 7 种疾病的识别上达到最优水平,对积液和肺实变的识别效果也接近最佳水平。值得一提的是,本文方法对肺炎的识别准确率对比比网络的最佳结果提高了 3.9%。此外,与大多数方法一样,MIM-Net 识别肺结节和渗透的能力仍需改进。渗透在影像学上边缘模糊,其诊断依赖于细微的纹理变化,肺结节属于小病变,易受无关特征的影响,这两种疾病的识别相对困难。与上述网络相比,MIM-Net 充分利用不同采样阶段的残差特征,增强了疾病在空间位置上的病理特征表达,因此在大多数疾病识别上取得了较好的效果。

为验证 MIM-Net 的鲁棒性和泛化能力,本文在相同的实验平台和实验环境下,于 CheXpert 数据集上评估其性能。表 3 展示了 MIM-Net 在 CheXpert 上与其他网络的对比结果,对比网络包括 Irvin 等^[7]提出的多 CNN 组合模型(表 3 中的

Ensemble)、Guan 等^[17]提出的 ConsultNet、Zhu 等^[12]提出的 PCAN 及 Pham 等^[20]提出的 DCNN。其中“U-Zeros”表示将不确定性标签数据视为未患病样本,“U-Ones”表示将不确定性标签数据视为患病样本。

从表 3 的实验结果来看,MIM-Net 在两种标签处理方案下均取得了最优结果,平均 AUC 值分别为 0.901 和 0.893,与在 ChestX-ray14 数据集上的结果具有一致性,显示出 MIM-Net 在分类任务中的优越性。当使用“U-Ones”策略时,MIM-Net 对肺不张和心脏肿大两种疾病的分类精度均高于其他网络;当使用“U-Zero”策略时,MIM-Net 对肺不张的 AUC 值比其他网络的最佳结果显著提高 3.1%。CheXpert 数据集上的实验表明,即便对于临床特征差异较大的疾病,MIM-Net 仍然具有很高的识别精度,验证了其在不同 CXR 图像数据集中具有较强的泛化能力。

2.5 消融实验

为验证 MIM-Net 中各模块的有效性,本文在 ChestX-ray14 数据集上开展消融实验,研究 MIA 模块和 MSIM 模块的有效性。实验以 MIM-Net 为基准,通过删除相应的模块探索其对分类精度的影响。本节通过移除 MIA 模块并使用 ResNet 的瓶颈结构替换 MSIM 模块,以及单独替换 MSIM 模块和单独移除 MIA 模块的方式设计了 3

表 3 各网络在 CheXpert 数据集上的分类精度对比

Table 3 Comparison of the classification accuracy of each network on the CheXpert dataset

疾病类型	Ensemble (U-Ones)	ConsultNet (U-Ones)	PCAN (U-Ones)	MIM-Net (U-Ones)	Ensemble (U-Zeros)	ConsultNet (U-Zeros)	DCNN (U-Zeros)	MIM-Net (U-Zeros)
肺不张	0.858	0.847	0.848	0.859	0.811	0.804	0.745	0.842
心脏肿大	0.832	0.868	0.865	0.885	0.840	0.874	0.813	0.873
肺实变	0.899	0.923	0.908	0.904	0.932	0.940	0.882	0.901
水肿	0.941	0.924	0.912	0.935	0.929	0.894	0.921	0.928
积液	0.934	0.926	0.940	0.924	0.931	0.923	0.930	0.923
平均值	0.893	0.898	0.895	0.901	0.889	0.889	0.858	0.893

组消融实验。将消融实验测得的每种疾病 AUC 值和平均 AUC 值与本文网络的结果进行比较,消融实验结果如表 4 所示。

首先,单独移除 MIA 模块会使平均 AUC 值降低 0.3%;其次,单独替换 MSIM 模块会使平均 AUC 值降低 0.8%。实验结果显示,本文提出的两个模块将平均 AUC 值提升了 0.9%(从 0.838 增加到 0.847),且对每种疾病的识别准确率几乎均有提升,说明本文提出的网络对提高胸部疾病分类精度是有效的。此外,同时使用 MSIM 模块和 MIA 模块有助于网络提取多尺度特征,进一步增强了通道间和空间特定位置的病理特征表达。

同时,本文在 ChestX-ray14 和 CheXpert 数据集上进行对比实验,探讨不同损失函数对网络分类能力的影响。其中,CheXpert 采用“U-Ones”策

表 4 MSIM 和 MIA 模块消融实验对比结果

Table 4 Comparison results of the ablation experiment on MSIM and MIA module

MSIM	×	×	✓	✓
MIA	×	✓	×	✓
肺不张	0.822	0.811	0.824	0.826
心脏肿大	0.911	0.907	0.918	0.919
积液	0.877	0.892	0.885	0.887
渗透	0.714	0.718	0.715	0.714
肿块	0.865	0.858	0.866	0.869
肺结节	0.779	0.791	0.786	0.784
肺炎	0.765	0.771	0.799	0.801
气胸	0.852	0.879	0.876	0.892
肺实变	0.817	0.801	0.805	0.809
水肿	0.895	0.891	0.897	0.900
肺气肿	0.917	0.915	0.926	0.924
纤维化	0.803	0.817	0.814	0.824
胸膜增厚	0.819	0.808	0.823	0.822
疝气	0.892	0.885	0.881	0.889
平均 AUC	0.838	0.839	0.844	0.847

略。实验分别用二元交叉熵损失函数、普通焦点损失函数代替非对称移位焦点损失函数,实验结果如表 5 所示。

实验结果表明,MIM-Net 在两个数据集上使用焦点损失函数获得的平均 AUC 值优于使用二元交叉熵损失函数的结果。焦点损失函数通过焦点因子增强了网络对难分类样本的关注,而非对称移位焦点损失函数在普通焦点损失函数的基础上,更加关注正负样本中难以分类的部分,进一步提升了网络的分类能力。

表 5 损失函数对结果的影响

Table 5 Influence of loss function on the results

损失函数	平均 AUC	
	ChestX-ray14	CheXpert
交叉熵损失函数	0.838	0.889
焦点损失函数	0.843	0.893
非对称移位焦点损失函数	0.847	0.901

2.6 可视化分析

本节通过梯度加权类激活映射算法^[21]反向传播模型,并计算特征映射的梯度信息,生成病灶区域热力图以显示 CXR 图像中网络分类的关键区域,直观感受网络的分类能力。为验证网络识别的准确性,将生成的热力图与专业医生在 ChestX-ray14 数据集中提供的病变标注图进行比较,实验结果如图 7 所示。

图 7 中右侧热力图中突出显示的部分代表模型中最受关注的部分,也是诊断的主要依据。由图 7 可以看出,CXR 图像中激活的区域与左侧专业医生标注的区域一致,说明本文网络的预测结果和医生给出的人工标注重合度很高。此外,突出显示的区域非常集中,肉眼可以快速定位病变发生的区域;对于不同大小的病变,网络也可以实现准确识别,进一步验证了网络的分类效果。

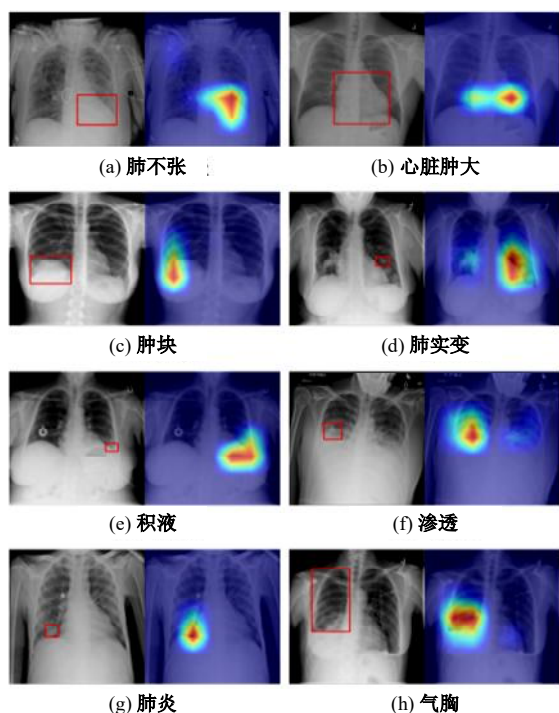


图 7 病灶区域标注图与可视化热力图

Fig. 7 Lesion area annotation map and visual heat map

3 结束语

本文提出一种面向多标签胸部X光片分类任务的多尺度注意力信息复用网络MIM-Net。该网络中的MSIM模块具有很强的空间信息复用和跨通道信息交流的能力,可以快速获取不同大小病变的感受野信息;采用MIA模块实现对病灶多尺度信息的提取,减少网络特征提取过程中图像噪声及病灶无关区域的影响;使用非对称移位焦点损失函数提高网络的整体分类准确性。实验结果表明,MIM-Net在ChestX-ray14和CheXpert数据集上均取得了很好的效果。此外,本文通过病灶区域可视化,将网络关注区域与人工标注进行对比,为人工诊断提供了直观的依据,进一步说明了该网络的应用意义。

参考文献:

- [1] Wei X L, Li W, Zhang M M, et al. Medical hyperspectral image classification based on end-to-end fusion deep neural network[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 68(11): 4481-4492.
- [2] 刘桂霞, 田郁欣, 王涛, 等. 基于双输入3D卷积神经网络的胰腺分割算法[J]. 吉林大学学报: 工学版, 2023, 53(12): 3565-3572.
- [3] 王雪, 李占山, 吕颖达. 基于多尺度感知和语义适配的医学图像分割算法[J]. 吉林大学学报: 工学版, 2022, 52(3): 640-647.
- [4] 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]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, USA, 2017: 3462-3471.
- [5] Krizhevsky A, Sutskever I, Hinton E G. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [6] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, USA, 2016: 770-778.
- [7] 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 Washington, DC: AAAI Press, 2019: 590-597.
- [8] Jiang X B, Zhu Y, Cai G, et al. MXT: A new variant of pyramid vision transformer for multi-label chest x-ray image classification[J]. Cognitive Computation, 2022, 14(4): 1362-1377.
- [9] Wang W H, Xie E Z, Li X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2021: 548-558.
- [10] 胡锦涛, 聂为之, 宋丹, 等. 可形变Transformer辅助的胸部X光影像疾病诊断模型[J]. 浙江大学学报: 工学版, 2023, 57(10): 1923-1932.
- Hu Jin-bo, Nie Wei-zhi, Song Dan, et al. Chest X-ray imaging disease diagnosis model assisted by de-

- formable transformer[J]. Journal of Zhejiang University (Engineering Science), 2023, 57(10): 1923–1932.
- [11] 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.
- [12] Zhu X F, Pang S M, Zhang X 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.
- [13] Chen K, Wang X Q, Zhang S W. Thorax disease classification based on pyramidal convolution shuffle attention neural network[J]. IEEE Access, 2022, 10: 85571–85581.
- [14] Lu Z C, Deb K, Boddeti V N. MUXConv: Information multiplexing in convolutional neural networks[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, 6: 12041–12050.
- [15] Woo S H, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]//Computer Vision—ECCV 2018. Cham, Switzerland: Springer Nature Switzerland AG, 2018: 3–19.
- [16] Ridnik T, Ben B E, Zamir N, et al. Asymmetric loss for multi-label classification[C]//IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2021: 82–91.
- [17] 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.
- [18] Lee Y W, Huang S K, Chang R F. CheXGAT: A disease correlation-aware network for thorax disease diagnosis from chest X-ray images[J]. Artificial Intelligence in Medicine, 2022, 132: 102382.
- [19] Chen B Z, Zhang Z, Li Y J, et al. Multi-label chest X-ray image classification via semantic similarity graph embedding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(4): 2455–2468.
- [20] Pham H, Le T, Tran D, et al. Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels[J]. Neurocomputing, 2021, 437: 186–194.
- [21] Selvaraju R, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]//IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 618–626.