



## 基于注意力机制和双分支网络的胸部疾病分类

宋子岩, 罗川, 李天瑞, 陈红梅

### 引用本文

宋子岩, 罗川, 李天瑞, 陈红梅. [基于注意力机制和双分支网络的胸部疾病分类](#)[J]. 计算机科学, 2024, 51(11A): 230900116-6.

SONG Ziyan, LUO Chuan, LI Tianrui, CHEN Hongmei. [Classification of Thoracic Diseases Based on Attention Mechanisms and Two-branch Networks](#) [J]. Computer Science, 2024, 51(11A): 230900116-6.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

#### Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于多模态融合的动态恶意软件检测方法](#)

Multimodal Fusion Based Dynamic Malware Detection

计算机科学, 2024, 51(11A): 240200098-7. <https://doi.org/10.11896/jsjcx.240200098>

#### [基于开放集的入侵检测方法研究](#)

Study on Open Set Based Intrusion Detection Method

计算机科学, 2024, 51(11A): 231000033-6. <https://doi.org/10.11896/jsjcx.231000033>

#### [基于加权特征融合的物联网设备识别方法](#)

IoT Devices Identification Method Based on Weighted Feature Fusion

计算机科学, 2024, 51(11A): 240100137-9. <https://doi.org/10.11896/jsjcx.240100137>

#### [基于CNN结合BiGRU的恶意流量分类算法研究](#)

Study on Malicious Traffic Classification Algorithm Based on CNN Combined with BiGRU

计算机科学, 2024, 51(11A): 231100106-9. <https://doi.org/10.11896/jsjcx.231100106>

#### [基于深度学习智能反射面辅助通信系统的联合波束成形](#)

Deep Learning Based Joint Beamforming in Intelligent Reflecting Surface Enhanced Wireless Communication Systems

计算机科学, 2024, 51(11A): 231200125-5. <https://doi.org/10.11896/jsjcx.231200125>

# 基于注意力机制和双分支网络的胸部疾病分类

宋子岩<sup>1</sup> 罗川<sup>1</sup> 李天瑞<sup>2</sup> 陈红梅<sup>2</sup>

1 四川大学计算机学院 成都 610065

2 西南交通大学计算机与人工智能学院 成都 611756

(zy\_song2150@163.com)

**摘要** 基于胸部 X 光片的胸部疾病分类对于提高诊断准确率、减轻医疗卫生系统压力具有重要意义。不同胸部疾病的发病区域尺寸存在巨大差异,是基于胸部 X 光片的胸部疾病分类任务面临的主要挑战。在对发病区域较小的疾病进行分类时,图像中大部分区域均为噪声区域,传统方法难以有效应对疾病间巨大的尺寸差异。针对这一问题,提出了一种结合多尺度特征的掩码构造方法,以 DenseNet-121 为特征提取器,构建了一个双分支网络,使用全局网络进行总体分类,并将微小病变区域送入局部分支以减轻噪声区域的干扰,最终利用基于注意力机制分支特征融合模块自适应地融合两个分支的分类特征信息。在 ChestX-ray14 数据集上进行了对比实验、消融实验和参数敏感性分析,结果表明,所提方法对 14 种胸部疾病分类的平均 AUC 高于现有方法,具有有效性且对参数不敏感。

**关键词**:胸部疾病分类;深度学习;胸部 X 光片;注意力机制;特征融合

中图法分类号 TP391

## Classification of Thoracic Diseases Based on Attention Mechanisms and Two-branch Networks

SONG Ziyang<sup>1</sup>, LUO Chuan<sup>1</sup>, LI Tianrui<sup>2</sup> and CHEN Hongmei<sup>2</sup>

1 College of Computer Science, Sichuan University, Chengdu 610065, China

2 School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China

**Abstract** Thoracic disease classification based on chest radiographs is important to improve diagnostic accuracy and reduce the pressure on the healthcare system. The huge variation in the size of the regions of different thoracic diseases is the main challenge in the chest radiograph-based classification of thoracic diseases. When classifying diseases with small onset regions, most of the regions in the image are noisy regions, and it is difficult for traditional methods to cope with the huge size differences among diseases effectively. To address this problem, a mask construction method combining multi-scale features is proposed, using DenseNet-121 as the feature extractor, a two-branch network is constructed, in which the global network is used for the overall classification, and tiny lesion regions are fed into the local branches to mitigate the interference of noisy regions. The branch feature fusion module based on the attentional mechanism is used to fuse the classification features from the two branches' information adaptively. Comparison experiments, ablation experiments, and parameter sensitivity analyses are performed on the ChestX-ray14 dataset. The experimental results show that the average AUC of the proposed method for classifying 14 thoracic diseases is higher than that of the existing methods, which is effective and parameter insensitive.

**Keywords** Thoracic disease classification, Deep learning, Chest radiographs, Attention mechanism, Feature fusion

### 1 引言

胸部疾病是全球最常见的健康问题,给每个国家医疗卫生系统都带来了沉重的负担。调查显示,2019 年仅间质性肺病和肺结节病的发病数和死亡数分别有 2420 万例和 169 833 例<sup>[1]</sup>。胸部 X 光片(CXR)、磁共振成像(MRI)和计算机断层扫描(CT)都是常见的医学成像方式。其中,CXR 具有价格低廉和辐射量低的优点,是胸部疾病诊断中最常见和最普遍的放射学检查类型之一。在临床诊断中,通常由临床医生对 CXR 图像进行人工分析。由于经验、信心、工作环境等的

不同,临床医生对于 CXR 图像的诊断结果也有所不同<sup>[2]</sup>。相较于新手医生,放射科专家往往只通过 CXR 图像中少数关键区域来获得最宝贵的信息。除此之外,由于胸部疾病病理变化复杂,即使是专业知识丰富、有充足临床经验的医生,在诊断时也难免出现错误。所以,利用深度学习技术提出一种自动化的胸部疾病分类方法对于帮助临床医生更加高效准确地做出诊断具有重要意义。

不同胸部疾病的特征往往出现在不同的位置,例如胸腔积液的特征通常出现在胸膜腔内,但是肺炎的感染特征只会在肺部出现<sup>[3]</sup>。这要求网络能够依据不同疾病关注不同的

基金项目:国家自然科学基金(62076171,62376230),四川省自然科学基金(2022NSFSC0898)

This work was supported by the National Natural Science Foundation of China(62076171,62376230) and Natural Science Foundation of Sichuan Province, China(2022NSFSC0898).

通信作者:罗川(c\_luo@scu.edu.cn)

区域。此外, CXR 图像包含复杂的病理信息, 但是每张 CXR 图像上都存在相同的组织和器官。这种极高的视觉相似性可能误导网络, 使得网络难以分辨哪些维度上的信息更重要<sup>[4]</sup>。因此, 网络需要具备区分不同维度信息重要性的能力。此外, 不同的疾病往往还存在明显的尺寸差异, 如结节的发病区域很小, 但是肺炎的发病区域却比结节大得多。对于发病区域小的疾病而言, 全局图像中包含相当程度的噪声<sup>[5]</sup>。因此, 网络难以适应大的尺寸变化, 对于小的区域难以聚焦。

目前已经有许多工作致力于使用深度学习方法解决上述问题<sup>[6]</sup>。Wang 等<sup>[7]</sup>发布了一个大型的多类别 CXR 数据集, 即 ChestX-ray14, 该数据集涵盖了肺炎、肺结核、肺结节等常见胸部疾病, 并且评估了多个传统的 CNN 结构在此数据集上进行多标签分类的有效性, 极大地推动了深度学习在 CXR 图像分类上的应用。Rajpurkar 等<sup>[8]</sup>使用密集连接<sup>[9]</sup>和批量标准化<sup>[10]</sup>方法优化了一个 121 层的卷积神经网络, 并且在肺炎诊断方面超过了放射科医生的平均水平。Baltruschat 等<sup>[11]</sup>探究了训练策略、网络结构和非图像特征等对于疾病分类的影响。这些早期方法并未将医学领域的先验知识与深度学习相结合。所以, 后续出现了许多引入医学领域的先验知识的方法。其中, 一些方法考虑了标签间潜在的依赖关系。Guan 等<sup>[12]</sup>提出考虑标签之间的相关性, 通过调整不同类别的权重来突出标签间的依赖关系, 并且抑制不相关标签引入的噪声。Ge 等<sup>[13]</sup>则是考虑了存在和不存在类别之间的标签关系, 提出了 Multi-label Softmax Loss (MSML) 和 Correlation Loss (CorLoss) 来解决多标签和图像间视觉相似度高的问题。另一些工作使用注意力机制来引导网络关注特定信息。Cai 等<sup>[14]</sup>提出了一种注意力挖掘策略来提高网络对于疾病的敏感性, 让网络更加关注疾病区域。Li 等<sup>[15]</sup>将 SENet<sup>[16]</sup>与 ResNet 相结合, 让网络能够关注与肺炎相关的通道, 来解决 CXR 图像中大量病变之间的高度视觉相似性的难题。Saednia 等<sup>[17]</sup>使用 LSTM 构建了循环注意力机制, 引导网络关注图像中更重要的具有异常信息的区域。Wang 等<sup>[18]</sup>利用 Grad-CAM 构建了注意力分支, 使得分类网络在不需要边界框监督的情况下也能够具备定位能力。Guan 等<sup>[5]</sup>提出构建注意力热力图, 使用局部分支引导网络关注小尺寸病变区域。还有一些工作将分割与分类相结合, 分割裁剪出特定区域, 通过减少噪声的干扰来提高分类性能。Kamal 等<sup>[19]</sup>和 Chen 等<sup>[20]</sup>将解剖学知识引入深度学习中, 将分割出的解剖区域与分类结合, 大大抑制了无关区域的影响。

在已有工作中, 使用全局和局部两个分支构建网络, 依据深度特征计算热力图推断疾病相关区域掩码, 构造局部分支的输入的方法取得了较好效果。但是, 考虑到浅层特征中丰富的空间信息能为确定疾病区域提供空间信息, 本文提出了一种结合浅层和深层特征的掩码推断方法。此前许多使用双分支网络的工作中, 只使用简单的拼接操作融合不同分支信息。通过这一简单方式进行特征融合难以充分利用两个分支的信息, 甚至一个分支可能对另一分支产生干扰, 导致最终分类效果变差。对此, 本文提出了一个基于注意力的分支特征融合模块, 引入了通道注意力和空间注意力, 引导网络自适应地融合两个分支的信息。

本文的主要贡献如下: 1) 提出了一种结合全局分支中不同尺度的信息来推断疾病位置的方法, 此方法产生的位置

掩码有利于网络聚焦于病变区域, 减少了噪声的干扰; 2) 提出了一个基于注意力的分支特征融合模块, 结合了通道注意力和空间注意力, 能够引导网络更好地利用全局和局部信息。

## 2 基于注意力机制的双分支网络框架

### 2.1 网络结构

图 1 为本文方法的网络结构图, 网络主要由全局分支、局部分支、掩码推断模块和基于注意力的分支特征融合模块构成。下面将介绍模型的总体流程。输入图像  $I_e$  首先被送入预训练的全局分支进行微调。然后选择全局分支的 3 个中间特征送入掩码推断模块, 综合其中的病变位置信息, 得到掩码  $M$ 。根据  $M$  从原始图像中裁剪对应区域  $I_l$ , 将尺寸调整为与全局分支的输入尺寸一致后, 送入预训练的局部分支进行微调。此后, 全局和局部分支作为特征提取器, 分别从全局尺度和病变区域尺度获取疾病特征信息。得到的特征信息将被送入基于注意力的分支特征融合模块充分融合并完成分类任务。

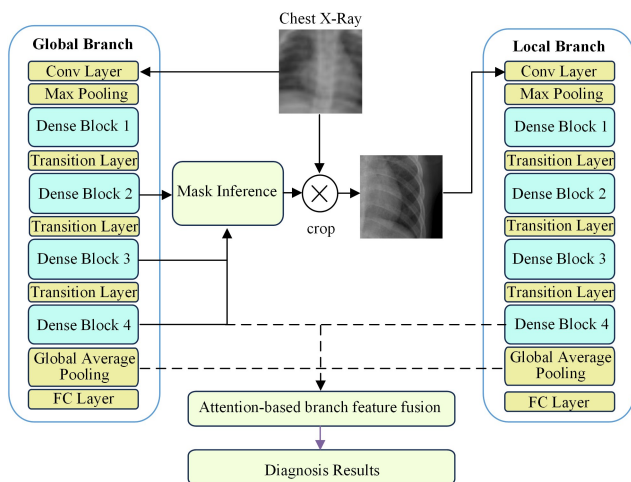


图 1 网络结构图

Fig. 1 Network architecture

### 2.2 双分支网络

为了应对胸部疾病存在的尺寸差异和位置差异, 本文提出的网络中使用了两个平行的分支, 即全局分支和局部分支。这两个分支作为特征提取器, 分别提取全局图像和重点病变区域的特征。在 CNN 网络的各种变体中, DenseNet-121 被认为更容易训练并且具有更强的捕捉特征的能力, 所以本文使用了在 ImageNet<sup>[21]</sup>上进行预训练的 DenseNet-121 作为骨干网络。DenseNet-121 的结构由一个具有 64 个大小为  $7 \times 7$  的核的卷积层开始, 然后是一个  $3 \times 3$  的最大池化层。随后, 有 4 个连续的密集块, 每个密集块分别由 12, 24, 48 和 32 个密集连接的卷积层组成。前 3 个密集块的每个块后面都有一个过渡层。在密集块之后, 使用全局平均池化层, 最后是包含 1000 个神经元并采用 Softmax 激活函数的全连接层。为了让 DenseNet-121 更适合胸部疾病分类问题, 本文对其结构进行了调整, 在最后一个全连接层中只保留了 14 个神经元, 对应了要分类的 14 种胸部疾病, 并且将激活函数替换为了 Sigmoid 函数。

### 2.3 多尺度掩码推断

本文提出了一种构建二进制掩码的方法, 该掩码结合了不同深度特征图中的信息, 能够定位全局图像中与疾病分类

高度相关的区域。

通常情况下,在浅层网络提取特征时,感受野较小,则特征图中每个像素点对应的感受野重叠区域很小,此时的特征为颜色、边缘和纹理等低级特征,特征图分辨率高,称为浅层特征或低级特征。低级特征中富含空间信息和小目标信息。与之不同的,深度网络提取特征时,随着下采样或卷积次数的增加,感受野逐渐变大,特征图中每个像素点代表一个区域的信息。此时的特征图分辨率低,具有更丰富的语义信息,可以捕捉较大目标的信息,称为深层特征或高级特征。高级特征更加鲁棒和抽象,对于理解图像内容更有帮助。在定位全局图像中与疾病分类高度相关的区域时,既需要高分辨率信息来确定准确的位置,又需要丰富的语义信息来确定该区域与病变相关,因此结合低级和高级特征是很有必要的。在骨干网络 DenseNet-121 中共有 4 个密集块,在兼顾准确性和高效性的前提下,使用最后 3 个 block 的组合更有利于疾病分类<sup>[4]</sup>。下面将详细介绍掩码推断的过程。

首先,  $f_{b1}$ ,  $f_{b2}$  和  $f_{b3}$  代表最后 3 个密集块输出的特征图,尺寸分别为  $28 \times 28$ ,  $14 \times 14$  和  $7 \times 7$ 。然后,通过下采样,将  $f_{b1}$  和  $f_{b2}$  的尺寸调整到与  $f_{b3}$  一致后拼接。拼接后的结果使用  $1 \times 1$  的卷积将通道数调整为 1024,得到融合特征  $f$ 。  $f_k(x, y)$  表示第  $k$  通道空间位置  $(x, y)$  处的激活值,其中  $k \in \{1, \dots, K\}$ ,  $K=1024$ 。对  $f_k(x, y)$  取绝对值并且统计各通道的最大值,得到热力图  $H$ 。

$$H(x, y) = \max_k (|f_k(x, y)|), k \in \{1, \dots, K\} \quad (1)$$

其中,  $H$  中每个位置上的值代表该位置对于疾病分类的重要度。使用阈值构建掩码是一种简单有效的方法<sup>[5]</sup>,在热力图

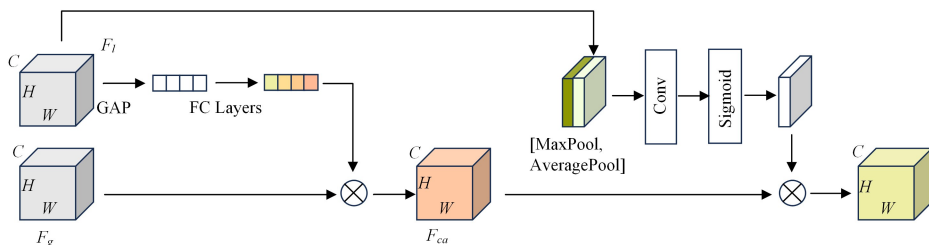


图 2 基于注意力的分支特征融合模块结构图

Fig. 2 Diagram of attention-based branch feature fusion module

在激发阶段,使用了一种 Sigmoid 激发的简单门控机制来充分捕捉通道间的依赖关系,即强调重要特征又抑制不重要的特征。具体的操作是将  $z$  输入两个连续的 FC 层。为了限制模型的复杂性并且有助于优化,第一个 FC 层的衰减率为  $r$ ,使用 ReLU 激活的降维层,  $r$  通常为  $1/16$ ;然后是一个返回原始通道数,使用 Sigmoid 激活的升维层。这一部分的输出  $M_c \in R^{(1,1,C)}$  可以表示为:

$$M_c = \sigma(W_2 \delta(W_1)) \quad (4)$$

其中,  $W_1$  和  $W_2$  是与两个 FC 层相关的权重矩阵,  $\delta(\cdot)$  是 ReLU 函数,  $\sigma(\cdot)$  是 Sigmoid 函数。然后,对  $M_c$  和全局特征图  $F_g$  进行哈达玛乘积,得到  $F_{ca}$ 。

空间注意力部分参考了 CBAM<sup>[22]</sup>,具体结构如图 2 右侧所示。其具体流程如下:输入特征  $F_l \in R^{(H,W,C)}$  平行地沿通道进行平均池化和最大池化,得到两个特征图  $F_s^{avg} \in R^{(H,W,1)}$  和  $F_s^{max} \in R^{(H,W,1)}$ ;然后,通过标准卷积层对这两个特征图进行串联和卷积,使用 Sigmoid 函数激活,生成二维空间注意力图。空间注意力  $M_s \in R^{(H,W)}$  的计算式如下:

中如果某位置的值大于阈值  $\tau$ ,就将掩码中该位置赋值为 1。

$$M(x, y) = \begin{cases} 1, & H(x, y) > \tau \\ 0, & \text{else} \end{cases} \quad (2)$$

得到掩码后,选择掩码中的最大连通区域,使用横轴和纵轴上的最小和最大坐标  $[x_{min}, y_{min}, x_{max}, y_{max}]$  表示。最后,从全局图像中裁剪出连通区域对应的部分,并通过双线性插值法将其大小调整为与全局图像相同。

## 2.4 基于注意力的分支特征融合模块

在此前的许多工作中,不同分支的特征通过拼接、取最大值等简单直接的方式结合在一起。这类简单操作无法充分利用不同分支的信息,因为不同的分支在类别、尺寸等不同方面有不同的侧重。此外,不加选择地结合两个分支的特征可能会放大某一分支中的噪声影响,导致另一分支难以发挥作用。所以本文提出使用注意力机制引导特征进行有选择的融合。

基于注意力的分支特征融合模块包括通道注意力和空间注意力两部分。对于通道注意力,使用结构简单的 SE 模块<sup>[16]</sup>来选出疾病相关的特征通道。如图 2 所示,左侧 SE 模块由全局平均池化层和两个 FC 层组成。SE 模块首先对卷积得到的特征图进行挤压操作,得到通道级的全局特征。然后对全局特征进行激发操作,学习各个通道间的关系,得到不同通道的权重。输入特征图  $F_l \in R^{(H,W,C)}$ ,其中  $H, W, C$  分别对应特征图的高度、宽度和通道。在挤压阶段,通过对输入特征执行全局平均池化操作,生成通道相关统计量,称作  $z$ 。  $z$  的第  $c$  个元素的计算式为:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W f_l(i, j) \quad (3)$$

$$M_c = \sigma(f^{7 \times 7}([AvgPool(F_l); MaxPool(F_l)])) \\ = \sigma(f^{7 \times 7}(F_s^{avg}; F_s^{max})) \quad (5)$$

其中,  $\sigma$  表示 Sigmoid 函数,  $f^{7 \times 7}$  表示滤波器大小为  $7 \times 7$  的卷积运算。最后,将  $F_{ca}$  与  $M_c$  进行哈达玛乘积,得到融合两个分支信息的特征,并使用以 Sigmoid 函数为激活函数的 FC 层来获得最终的疾病分类。

## 3 实验

### 3.1 数据集

实验使用 ChestX-ray14 数据集对提出的方法的分类性能进行定量评估。该数据集包含了来自 32717 名不同病人的 112120 张正面胸部 X 光图片。数据集中的所有图像均标注了包括 14 种疾病和 no finding 在内的 15 种标签,每张图片的标签被设为一个 14 维的向量  $L = [l_1, l_2, \dots, l_c, \dots, l_C]$ ,其中  $l_c \in \{0, 1\}$ ,  $C=14$ ,每一位对应一种疾病。此外,实验将数据集分为 3 个子集,70% 为训练集,10% 为验证集,20% 为测试集,数据的具体分布如图 3 所示。

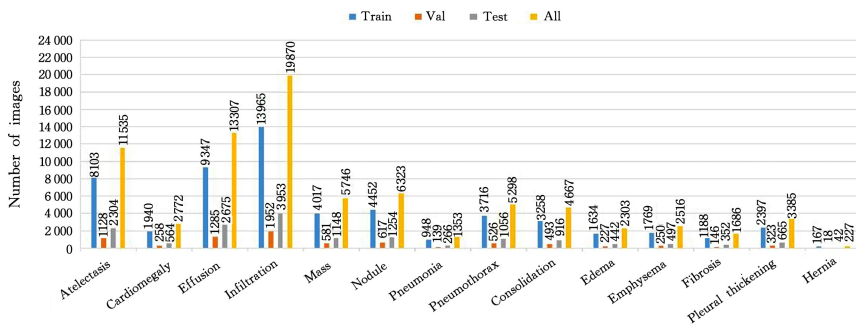


图3 ChestX-ray14中训练集、验证集、测试集的数据及总体数据的分布

Fig. 3 Distribution of data of training set, validation set and testing set and overall data in ChestX-ray14

## 3.2 实验设置

在训练时,大小为  $1024 \times 1024$  的原始图像被缩放至  $256 \times 256$ ,随后随机裁剪为  $224 \times 224$ ,并使用随机水平翻转进行数据增强。所提出的网络训练可以通过以下 5 个步骤来实现:1)使用全局图像微调在 ImageNet 数据集上预训练的全局分支;2)冻结全局分支的全部参数,根据全局分支中的特征推断出掩码并从全局图像中裁剪出局部分支的输入,微调在 ImageNet 数据集上预训练的局部分支;3)随机初始化基于注意力机制的分支特征融合模块的权重;4)冻结全局分支和局部分支的全部参数,训练基于注意力机制的分支特征融合模块;5)端到端地对网络整体进行微调。用于优化网络各个部分的损失函数为二元交叉熵损失(BCE Loss)。训练网络的各个部分时使用的 mini-batch 分别为 64,32,32。学习率的初始值为 0.001,在 15 个 epoch 后缩小 10 倍。训练时的权重衰减和动量分别为 0.0001 和 0.9。在测试过程中,原始图像被缩放至  $256 \times 256$ ,然后通过中心裁剪得到大小为  $224 \times 224$  的输入图像。经实验验证后,掩码推断模块的阈值  $\tau$  被设为 0.4。

## 3.3 实验结果与分析

### 3.3.1 对比实验

为了验证本文方法的有效性,将其与 6 种已有的在 ChestX-ray14 数据集上的胸部疾病分类方法进行了对比。这 6 种方法分别为:DCNN<sup>[7]</sup>、LSTM-Net<sup>[23]</sup>、Li 等提出的方法<sup>[24]</sup>、CRAL<sup>[12]</sup>、TSCN<sup>[20]</sup> 和 AG-CNN<sup>[5]</sup>。DCNN 采用弱监督方法对胸部疾病进行多标签分类。LSTM-Net 利用 LSTM 挖掘标签之间的依赖关系来预测 14 种胸部疾病。Li 等提出的方法利用类别信息和有限的位置信息进行分类和定位,实现了高精度的疾病预测。CRAL 提出为无关疾病类别分配较小权重,同时为相关疾病类别分配较大权重。TSCN 利用分割网络 U-Net 对肺部区域进行分割,然后将原始 CXR 图像和肺部区域图像分别输入并行特征提取器,最后利用 Hilbert-Schmidt 独立性准则获得两个分支的权重,将两个分支的结果进行融合。AG-CNN 利用全局分支中最深层的特征构建局部分支的输入,最终拼接两个分支的特征完成了对胸部疾病的分类,对比实验结果如表 1 所列。

表 1 与 6 种已有方法的对比实验结果

Table 1 Comparative experimental results between the proposed method and six existing methods

Method	Atel	Card	Effu	Infi	Mass	Nodu	Pneu1	Pneu2	Cons1	Edem	Emph	Fibr	P_T	Hean	Mean
DCNN	0.700	0.810	0.759	0.661	0.693	0.669	0.658	0.799	0.703	0.805	0.833	0.786	0.684	0.872	0.745
LSTM-Net	0.733	0.856	0.806	0.673	0.718	0.777	0.684	0.805	0.711	0.806	0.842	0.743	0.724	0.775	0.761
Li et al.	0.80	0.87	0.87	0.70	0.83	0.75	0.67	0.87	0.80	0.88	0.91	0.78	0.79	0.77	0.81
CRAL	0.781	0.880	0.829	0.702	0.834	0.773	0.729	0.857	0.754	0.850	0.908	0.830	0.778	0.917	0.816
TSCN	0.785	0.887	0.831	<b>0.703</b>	0.833	<b>0.798</b>	0.731	<b>0.881</b>	0.754	0.849	<b>0.930</b>	<b>0.833</b>	0.782	0.921	0.823
AG-CNN	<b>0.802</b>	0.905	0.874	0.696	<b>0.839</b>	0.742	0.732	0.873	<b>0.807</b>	<b>0.896</b>	0.884	0.779	<b>0.783</b>	0.930	0.825
Ours	0.801	<b>0.907</b>	<b>0.875</b>	0.699	0.837	0.742	<b>0.734</b>	0.876	<b>0.807</b>	<b>0.896</b>	0.888	0.788	0.782	<b>0.938</b>	<b>0.826</b>

从结果中可以得出,所提出的方法在所有方法中取得了最高的平均 AUC 值,为 0.826。本文方法在心脏肿大(Card)、胸腔积液(Effu)、肺炎(Pneu1)、肺实变(Cons)、水肿(Edem)和肺疝(Hean)中准确率较其他方法更高,AUC 值分别为 0.907,0.875,0.734,0.807,0.896 和 0.938。综上所述,本文提出的方法对胸部疾病的分类精度较目前已有方法总体上具有最优的性能,并在多个类别的分类中具有较大优势。

### 3.3.2 消融实验

为了验证本文提出的多尺度掩码推理方法和基于注意力的分支特征融合方法的有效性,对两种方法进行消融实验验证。

如表 2 所列,首先,仅使用 DenseNet-121 对 ChestX-ray14 数据集进行端到端分类(见表中 G)。然后,引入多尺度

掩码推断方法。很明显,使用该方法引入局部分支后(见表中 G+L),分类性能得到了提高。与表 3 中下采样策略下局部分支分类性能及网络总体分类性能结合起来看,局部分支只接受并处理特定小发病区域信息,所以单一分支的整体分类性能较差,但与全局分支组合后的分类效果高于二者中任意一个,提升了网络对 11 种疾病的分类准确率,说明小发病区域信息有利于提升网络整体分类性能。但此时也有一些疾病类别的分类性能有所下降,这表明分支中的某些信息可能干扰了模型,在这种情况下模型的鲁棒性较差。之后,引入基于注意力的分支特征融合模块(见表中 G+L+F),与前一阶段相比,分类性能得到了进一步提高。具体来说,基于注意力的分支融合模块提高了大多数疾病的分类准确性。如果没有该模块,局部分支可能会导致一些疾病的分类准确率下降,如结节、水肿和纤维化。但引入该模块后,模型特征融合的稳定性

得到了提高,局部分支产生的干扰也被大大抑制,因此与全局分支相比,所有结果都有非负的提高。

表2 消融实验结果

Table 2 Results of ablation experiment

Method	Atel	Card	Effu	Infi	Mass	Nodu	Pneu1	Pneu2	Cons	Edem	Emph	Fibr	P_T	Hean	Mean
G	0.800	0.898	0.873	0.694	0.833	0.742	0.724	0.873	0.805	0.895	0.884	0.783	0.778	0.926	0.822
G+L	<b>0.801</b>	0.904	0.874	0.698	<b>0.837</b>	0.739	0.731	0.871	<b>0.807</b>	0.893	0.883	<b>0.778</b>	0.780	0.936	0.824
G+L+F	<b>0.801</b>	<b>0.907</b>	<b>0.875</b>	<b>0.699</b>	<b>0.837</b>	<b>0.742</b>	<b>0.734</b>	<b>0.876</b>	<b>0.807</b>	<b>0.896</b>	<b>0.888</b>	<b>0.788</b>	<b>0.782</b>	<b>0.938</b>	<b>0.826</b>

表3 特征图采样策略分析的结果

Table 3 Results of feature map sampling strategy analysis

Method	Atel	Card	Effu	Infi	Mass	Nodu	Pneu1	Pneu2	Cons	Edem	Emph	Fibr	P_T	Hean	Mean
Up_L	0.767	0.890	0.850	0.682	0.799	0.683	0.715	0.837	0.793	0.864	0.826	0.725	0.740	0.881	0.789
Up	0.800	0.905	<b>0.875</b>	0.694	<b>0.839</b>	0.741	0.732	0.875	<b>0.808</b>	0.893	0.885	0.781	<b>0.782</b>	0.932	0.825
Mid_L	0.771	0.889	0.849	0.684	0.801	0.687	0.714	0.834	0.793	0.873	0.830	0.731	0.740	0.906	0.793
Mid	<b>0.801</b>	0.904	<b>0.875</b>	0.696	<b>0.839</b>	<b>0.744</b>	0.732	0.874	<b>0.808</b>	0.894	0.884	0.782	0.781	0.930	0.825
Down_L	0.764	0.893	0.847	0.683	0.780	0.678	0.716	0.833	0.787	0.871	0.815	0.736	0.741	0.895	0.788
Down	<b>0.801</b>	<b>0.907</b>	<b>0.875</b>	<b>0.699</b>	0.837	0.742	<b>0.734</b>	<b>0.876</b>	0.807	<b>0.896</b>	<b>0.888</b>	<b>0.788</b>	<b>0.782</b>	<b>0.938</b>	<b>0.826</b>

### 3.3.3 参数敏感性分析

本文对多尺度掩码推断方法中不同尺寸特征图采样策略和阈值  $\tau$  进行了敏感性分析。

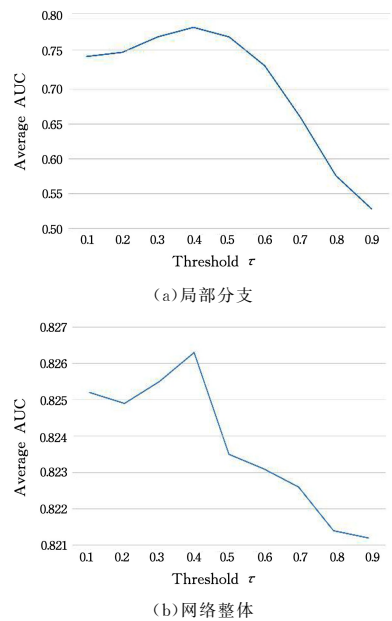
#### 1) 特征图采样策略分析

该部分分析了网络对于多尺度掩码推断方法中不同尺寸特征图采样策略的敏感性。通常,特征图越大,掩码推理的候选区域就越大,空间信息就越丰富,但噪声干扰就越大。特征图越小,掩膜推理的候选区域越小,但语义信息越丰富。本文使用的采样策略为将第二个密集块和第三个密集块的特征图尺寸向下采样为  $7 \times 7$ 。该部分对另外两种采样策略进行了实验,包括将第三和第四密集块的特征图尺寸向上采样到  $28 \times 28$ ,以及将第二和第四密集块的特征图尺寸向中间采样到  $14 \times 14$ 。这3种策略下局部分支和网络整体的具体实验数据如表3所列,其中 Up\_L 表示向上采样策略下的局部分支,Up 代表向上采样策略的总体网络,Mid 与 Down 部分同理。就局部分支而言,采用向上、中间策略时,局部分支的平均分类精度较高。然而,在采用向下策略时,网络整体的平均分类准确率达到最高值。在使用向上和中间策略时,由于特征图较大,局部分支中学习到的特征表征可能更接近全局分支。结合消融实验部分得出的结果,推断此时学习到的特征表征更接近全局分支,从而使得局部分支的平均分类准确率更高。此外,两个相似分支的特征融合可能并不能起到有效的信息补充作用,因此达不到最优结果。采用向下策略时,特征图的规模较小,抽象语义信息较多,仅靠局部分支很难取得良好的分类效果。但局部分支可能从这些抽象信息中学到了与全局分支互补的特征表征,从而在分支信息融合后达到最优结果。总体而言,在使用不同采样策略时,网络分类性能在较小范围内波动,网络对于采样策略不敏感。

#### 2) 阈值 $\tau$ 敏感性分析

该部分分析了多尺度掩码推断方法中网络对参数  $\tau$  的敏感性。具体实验结果如图4所示。 $\tau$  趋于0和1的极限情况很容易预见。当  $\tau$  接近0时,大部分区域被保留,裁剪后的图像对应于原始图像的大部分区域。当  $\tau$  接近1时,大部分区域被丢弃,裁剪后剩余区域非常小,难以用于模型训练。从图4(a)中可以得出结论,当  $\tau$  在0.1到0.6之间时,局部分支分类性能变化不大;当  $\tau$  在0.7到0.9之间时,该分支自身的分类性能会逐渐下降到很低的水平。这与之前的推断一致。从

图4(b)中可以得出,网络整体的分类性能随  $\tau$  的变化在很小的范围内波动,当  $\tau=0.4$  时达到最佳。因此,网络对参数  $\tau$  不敏感,可以保持稳定的性能。

图4 网络在不同阈值  $\tau$  下的实验结果Fig. 4 Experimental results of the network with different thresholds  $\tau$ 

**结束语** 本文提出了一种新颖的胸部疾病分类方法,采用双分支结构来构建网络,提出了一种结合不同尺寸特征来推断疾病相关区域掩码的多尺度掩码推断方法和基于注意力的分支特征融合模块。多尺度掩码推断方法构造的局部分支能够学习小尺寸区域中的疾病特征,得到的特征经过分支特征融合模块与全局分支融合后,实现了对胸部X光片中信息的充分利用。在ChestX-ray14数据集上对14种胸部疾病分类的对比实验、消融实验和参数分析验证了本文方法的有效性和鲁棒性。在今后的研究中,我们将探索使用合理的重采样方法解决数据集中存在的数据倾斜和标签分布不均匀的问题,减轻数据集中数据分布对网络性能的限制。此外,我们还将探索在掩码推断方法中通过计算L1距离、L2距离等方式替换最大值操作,推断出更准确的疾病相关区域。

## 参 考 文 献

- [1] MA X Q, ZHU L L, KURCHE J S, et al. Global and regional burden of interstitial lung disease and pulmonary sarcoidosis from 1990 to 2019: results from the global burden of disease study 2019[J]. *Thorax*, 2022, 77(6): 596-605.
- [2] MCLAUGHLIN L, BOND R, HUGHES C, et al. Computing eye gaze metrics for the automatic assessment of radiographer performance during x-ray image interpretation [J]. *International Journal of Medical Informatics*, 2017, 105: 11-21.
- [3] HOSSAIN M I, HOSSAIN S M, ZUNAED M, et al. A novel attention mechanism using anatomical prior probability maps for thoracic disease classification from x-ray images [J]. *arXiv: 2210.02998*, 2022.
- [4] 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.
- [5] GUAN Q J, HUANG Y P, ZHONG Z, et al. Thorax disease classification with attention guided convolutional neural network [J]. *Pattern Recognition Letters*, 2020, 131: 38-45.
- [6] ÇALLI E, SOGANCIOGLU E, VAN G B, et al. Deep learning for chest x-ray analysis: a survey [J]. *Medical Image Analysis*, 2021, 72: 102125.
- [7] 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]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 2097-2106.
- [8] RAJPURKAR P, IRVIN J, ZHU K, et al. Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning [J]. *arXiv: 1711.05225*, 2017.
- [9] 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.
- [10] IOFFEG S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// *International Conference on Machine Learning*. 2015: 448-456.
- [11] BALTRUSCHAT I M, NICKISCH H, GRASS M, et al. Comparison of deep learning approaches for multi-label chest x-ray classification[J]. *Scientific Reports*, 2019, 9(1): 6381.
- [12] 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.
- [13] GE Z Y, MAHAPATRA D, CHANG X J, et al. Improving multi-label chest x-ray disease diagnosis by exploiting disease and health labels dependencies [J]. *Multimedia Tools and Applications*, 2020, 79: 14889-14902.
- [14] CAI J Z, LU L, HARRISON A P, et al. Iterative attention mining for weakly supervised thoracic disease pattern localization in chest x-rays[C]// *Medical Image Computing and Computer Assisted Intervention*. 2018: 589-598.
- [15] LI B C, KANG G X, CHENG K, et al. Attention-guided convolutional neural network for detecting pneumonia on chest x-rays [C]// *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2019: 4851-4854.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 7132-7141.
- [17] SAEDNIA K, JALALIFAR A, EBRAHIMI S, et al. An attention-guided deep neural network for annotating abnormalities in chest x-ray images: visualization of network decision basis [C]// *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society*. 2020: 1258-1261.
- [18] WANG H Y, JIA H Z, LU L, et al. Thorax-net: an attention regularized deep neural network for classification of thoracic diseases on chest radiography [J]. *IEEE Journal of Biomedical and Health Informatics*, 2019, 24(2): 475-485.
- [19] KAMAL U, ZUNAED M, NIZAM N B, et al. Anatomy-xnet: an anatomy aware convolutional neural network for thoracic disease classification in chest x-rays [J]. *IEEE Journal of Biomedical and Health Informatics*, 2022, 26(11): 5518-5528.
- [20] 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.
- [21] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge [J]. *International Journal of Computer Vision*, 2015, 115: 211-252.
- [22] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module [C]// *Proceedings of the European Conference on Computer Vision*. 2018: 3-19.
- [23] YAO L, POBLENZE, DAGUNTS D, et al. Learning to diagnose from scratch by exploiting dependencies among labels [J]. *arXiv: 1710.10501*, 2017.
- [24] LI Z, WANG C, HAN M, et al. Thoracic disease identification and localization with limited supervision [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 8290-8299.



**SONG Ziyan**, born in 1999, postgraduate. His main research interest is medical image processing and analysis.



**LUO Chuan**, born in 1987, Ph.D, associate professor, Ph.D supervisor, is a member of CCF (No. 45031M). His main research interests include data mining and machine learning.