

# 融合局部与全局特征的肺炎医学影像分类<sup>①</sup>

宋佳航, 刘 静, 王青松, 李 明

(山东中医药大学 智能与信息工程学院, 济南 250355)

通信作者: 李 明, E-mail: 80500285@qq.com



**摘 要:** 针对现有肺炎医学影像识别研究在浅层网络忽略全局特征导致特征提取不全且模型规模较大的问题, 提出了一种基于 CNN 和注意力机制的轻量化模型提高肺炎类型的识别效率. 采用轻量化模型结构减少模型参数量, 通过增大卷积核, 引入高效通道注意力和自注意力机制解决网络重要信息丢失和无法提取底层全局信息的问题, 通过双分支并行提取局部和全局信息并使用多尺度通道注意力提高二者融合质量, 使用 CLAHE 算法优化原始数据. 实验结果表明, 该模型在保证轻量性的同时准确率、灵敏度、特异性较原模型分别提高 2.59%, 3.1%, 1.38%, 并优于当前优秀的其他分类模型, 具有更强的实用性.

**关键词:** 图像处理; 肺炎; 卷积神经网络; 注意力机制; 特征融合

引用格式: 宋佳航, 刘静, 王青松, 李明. 融合局部与全局特征的肺炎医学影像分类. 计算机系统应用, 2023, 32(11): 159-166. <http://www.c-s-a.org.cn/1003-3254/9288.html>

## Pneumonia Medical Image Classification Based on Fusion of Local and Global Features

SONG Jia-Hang, LIU Jing, WANG Qing-Song, LI Ming

(College of Intelligence and Information Engineering, Shandong University of Traditional Chinese Medicine, Jinan 250355, China)

**Abstract:** In view of the problem of incomplete feature extraction and large model size caused by ignoring global features in shallow networks in existing pneumonia medical image recognition research, a lightweight model based on convolutional neural network (CNN) and attention mechanism is proposed to improve the recognition efficiency of pneumonia types. A lightweight model structure is used to reduce the number of model parameters. By increasing the convolution kernel, efficient channel attention and self-attention mechanisms are introduced to solve the problem of loss of important network information and the inability to extract underlying global information. Local and global information is extracted in parallel through dual branches, and multi-scale channel attention is utilized to improve the fusion quality of the two. The CLAHE algorithm is employed to optimize the original data. The experimental results show that the accuracy, sensitivity, and specificity of the model are increased by 2.59%, 3.1%, and 1.38% respectively compared with those of the original model while ensuring lightness, and the proposed model outperforms other current excellent classification models and has stronger practicability.

**Key words:** image processing; pneumonia; convolutional neural network (CNN); attention mechanism; feature fusion

肺炎每年影响着全球上亿人的身体健康, 一般肺炎不具备传染性, 但在 2019 年末, 一种新型冠状病毒导致的肺炎 (COVID-19) 以其极强的传染性吸引了全

人类的眼球. 新冠肺炎会导致患者肺部发炎, 呼吸困难, 如果患者自身基础病多或者免疫力差甚至会导致死亡, 所以能够快速识别肺炎类型, 在短时间内救治患者并

① 基金项目: 国家自然科学基金面上项目 (82174528); 山东中医药大学产教融合协同育人项目 (CJ2021005, CJ2021003)

收稿时间: 2023-04-19; 修改时间: 2023-05-23; 采用时间: 2023-05-29; csa 在线出版时间: 2023-09-19

CNKI 网络首发时间: 2023-09-21

阻断病毒的传播是降低死亡率和遏制疫情蔓延的关键。医学影像用于疾病的辅助诊断是一种非常有效的方法<sup>[1]</sup>，随着医学数据的公开，深度学习技术在疾病辅助诊断任务中应用广泛，针对肺炎识别许多研究者给出了自己的解决方法，Li等<sup>[2]</sup>提出了一种多特征融合网络，可以有效识别肺炎相关病变的内在成像特征及其位置达到提取图像关键特征的目的，并设计相关模块协调不同通道的差异特征来强调重要信息，在肺炎分类任务中取得了较高的准确度。Hussain等<sup>[3]</sup>提出了一种新的CNN模型CoroDet用于多种肺炎的识别，能有效提取图像的重要信息，在三分类实验中取得了94.2%的准确度。面对图像数据不足的情况，有研究者提出使用迁移学习的方法来减少模型对数据的依赖，Ikechukwu等<sup>[4]</sup>在ResNet50和VGG19预训练模型的基础上进行微调，并使用数据增强来减少模型的过拟合，在肺炎类型的识别任务中召回率达到了92%。另外，多项研究表明注意力机制能够帮助网络定位图像中的重要信息，对于解决不同肺炎病灶区别不大，且病灶不明显的问题有重要帮助，李镛等<sup>[5]</sup>在ResNet和DenseNet的基础上加入3种不同注意力机制，在不增加网络参数的情况下对包括肺炎在内的14种胸部疾病进行识别，平均AUC达到了0.8185。谢娟英等<sup>[6]</sup>在ResNeXt模型的基础上加入通道注意力和残差注意力模块，在肺炎数据集中进行实验识别准确度达到96%。当前研究在模型设计改进时普遍规模较大，对硬件资源要求高，并且忽略了CNN网络在低层缺乏全局感受野的问题，致使网络在低层学习到的特征与高层特征具有较大差异<sup>[7]</sup>，而肺炎病灶特征具有多样性，只关注局部特征势必会造成重要信息丢失，降低模型的识别准确率。

针对以上问题，本文采用轻量模型架构，结合CNN和自注意力机制的优点，旨在设计一个能准确识别肺炎类型的轻量化模型，通过对轻量结构进行改进，并引入自注意力机制，让模型兼顾肺炎图像局部与全局特征提取的同时，减少自注意力机制对数据的依赖，提高肺炎类型的识别精度，同时减少计算量和复杂度，保证模型的轻量化。

## 1 本文方法

为了有效提取图像的特征信息并保持模型的轻量化，本文根据肺炎影像的数据特点在轻量结构的基础上做改进设计下采样模块(downsample)和用于提取局部特征信息的卷积模块(conv block)，通过交叉堆叠下

采样模块和卷积模块来搭建基础网络模型，然后引入多头自注意力机制(multi-head self attention, MSA)<sup>[8]</sup>设计全局特征提取模块获取全局特征信息，再设计特征融合模块融合局部与全局特征信息，最后将各模块嵌入到基础网络中完成模型搭建。后文将着重介绍各模块的设计改进和模型的整体结构。

### 1.1 基础网络架构

本文的基础网络包含下采样模块和卷积模块两个部分，卷积模块的设计在倒残差结构(inverted residual block)<sup>[9]</sup>的基础上进行改进，该结构使用深度可分离卷积(depthwise separable convolution)<sup>[10]</sup>用于数据的特征提取，可以减少大量参数，提高模型的轻量化，并加入残差结构(residual block)<sup>[11]</sup>，增强了层间的信息传递，但深度可分离卷积压缩模型时去除了大量的参数，从而损失了一些重要的特征信息，考虑到数据集图像分辨率较大，像素值方差较小，本文将深度卷积的卷积核从3×3调整为5×5来增加模型的有效感受野，提高网络全局信息获取能力，并在深度卷积之后，加入高效通道注意力(efficient channel attention, ECA)模块<sup>[12]</sup>实现局部跨通道的信息交互，突出特征图中有效信息通道。如图1所示，该模块通过全局平均池化(GAP)获得聚合特征，再执行大小为 $k$ 的一维卷积来生成通道权重，其中 $x$ 为输入特征， $k$ 通过信道维度 $C$ 的映射自适应地确定，将权重系数与对应通道特征相乘，实现模型对重要通道的定位，得到输出特征 $\tilde{x}$ 。

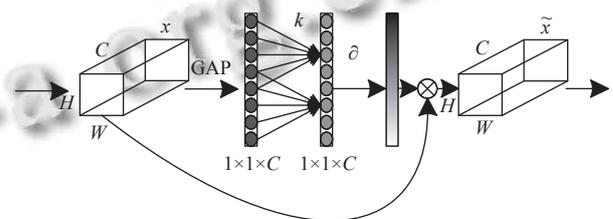


图1 ECA模块结构图

除以上改进外，本文将原模块的全连接层(fully connected, FC)移除，减少模型参数，并在最后的卷积层后添加了BN(batch normalization)<sup>[13]</sup>层和激活函数，避免训练中梯度消失的同时加速模型收敛，增强特征的非线性表达能力。倒残差模块与改进模块结构如图2所示。

另外，本文模型并未通过改变深度卷积的步距来达到下采样的目的，因为增大步距会导致细节信息丢失，影响模型的感知范围，并且下采样时省略残差结构会减弱层间特征信息的传递，所以本文使用单独的下

采样模块用于提高特征维度和去除冗余信息,该模块由卷积层、最大池化层、BN层串联而成,其中卷积层卷积核大小为3×3,主要用于提高特征维度,增加特征多样性,而最大池化层用于特征图的压缩,剔除冗余信息.本文的基础网络由改进的卷积模块和下采样模块交叉堆叠4次得到,卷积模块堆叠个数和基础网络架构如图3所示,在网络最后加入全局平均池化层(GAP)将提取得到的二维特征转化为一维向量并输入到全连接层进行分类得到最终结果.

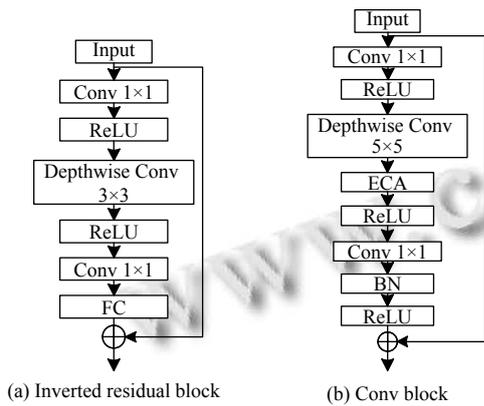


图2 倒残差模块和卷积模块对比图

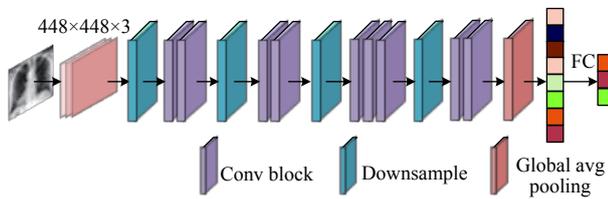


图3 基础网络模型

### 1.2 多头自注意力

对于肺炎医学影像来说,大部分区域为背景、骨

骼、胸廓等无需重点关注的区域,而需要重点关注的区域如胸腔、心脏、肺组织等占比较小,而多头自注意力机制的全局计算方式能帮助模型聚焦病灶区域,提取关键特征信息并建立全局特征依赖关系.多头自注意力机制是在自注意力机制的基础上提出来的,二者结构如图4所示,其中 $Q$ 代表查询(query), $K$ 代表键(key), $V$ 代表值(value),自注意力层通过 $Q$ 与 $K$ 、 $V$ 之间的交互实现特征信息的动态聚合,假设输入序列 $z \in R^{N \times d}$ ,使用3个线性映射矩阵 $U_{qkv} = (W^q, W^k, W^v)$ 将输入序列投影得到 $Q$ 、 $K$ 、 $V$ 这3个向量,然后将 $Q$ 与 $K$ 相乘得到相似度 $A$ ,相似度 $A$ 类似于权重,再将 $A$ 与 $V$ 进行加权得到最终结果.自注意力机制的计算过程如下所示:

$$[Q, K, V] = z \cdot U_{qkv}, U_{qkv} \in R^{d \times 3d_h} \quad (1)$$

$$A = \text{Softmax}(QK^T / \sqrt{d_h}), A \in R^{h \cdot d_h \times d} \quad (2)$$

最后结果可表示为:

$$SA(z) = A \cdot V \quad (3)$$

多头自注意力是对每个 $Q$ 和 $K$ 、 $V$ 做多次自注意力计算,它使用多个自注意力头来并行计算,最后将所有头的输出进行拼接再通过一个可学习的映射矩阵聚合得到最终结果.计算过程如下所示:

$$MSA(z) = \text{Concat}(SA_1(z), SA_2(z), \dots, SA_k(z)) \cdot U_{msa} \quad (4)$$

其中, $N$ 为向量个数, $d$ 表示向量维度, $d_h$ 为每个自注意力头的输出维度, $h$ 表示自注意力头的个数,本文将其数量设为4, $U_{msa} \in R^{h \cdot d_h \times d}$ 是对拼接后的特征进行聚合的映射矩阵,多头能通过 $Q$ 与 $K$ 、 $V$ 的多次交互获得不同的依赖关系,学习到更加多样的特征.

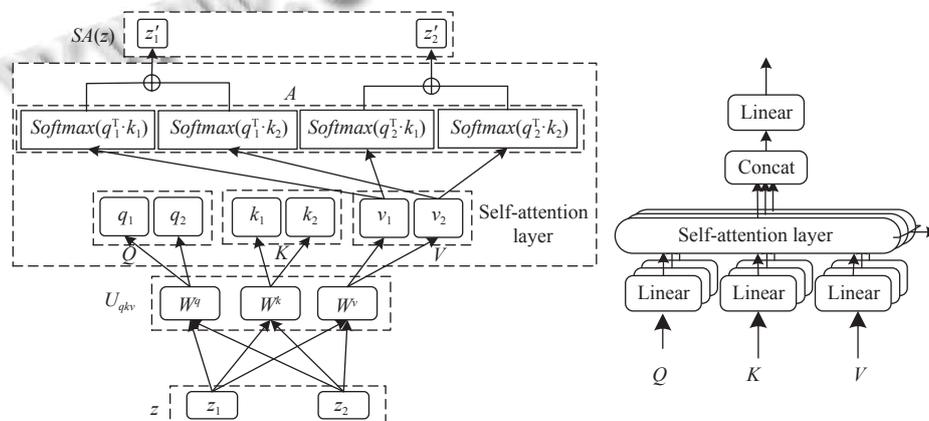


图4 自注意力与多头自注意力

### 1.3 全局特征提取模块

得益于多头自注意力机制具有全局依赖关系学习能力的特点, 本文以其为核心设计了全局特征提取模块 (global block). 由于自注意力的计算量与输入序列的长度成平方关系, 若是以像素为单位作为序列输入, 会产生巨大计算量, 而常用方法如 Vision Transformer (ViT)<sup>[14]</sup> 将图像切分成固定大小的块, 然后使用线性映射将每个块映射到一维向量作为自注意力模块的输入, 需要在向量中加绝对位置偏置, 导致模型参数量大且训练困难, 所以为减少计算量, 本文通过卷积模块提供空间归纳偏置, 摆脱了位置偏置的束缚, 并受 MobileViT<sup>[15]</sup> 的启发构建全局特征提取模块, 能够让多头自注意力学习到具有空间归纳偏置的全局特征, 具体过程如图 5 所示.

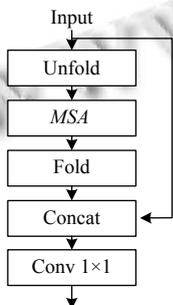


图 5 全局特征提取模块

图 5 中 Unfold 模块是序列化过程, 当输入一个特征图张量  $X \in R^{H \times W \times C}$ , Unfold 模块会将特征图划分为固定大小的块 (patch), 并将每个 patch 对应位置的像素取出, 展开为  $X_U \in R^{P \times N \times C}$  作为自注意力模块的输入序列, 再通过多头自注意力模块 (MSA) 获取全局信息得到输出  $X_G \in R^{P \times N \times C}$ , 计算过程如下所示:

$$X_G(p) = MSA(X_U(p)), 1 \leq p \leq P \quad (5)$$

其中,  $H$ 、 $W$  分别为特征图的高和宽,  $P$  为 patch 的大小,  $N$  是 patch 的数目,  $C$  是维度,  $p \in \{1, \dots, P\}$ , 此过程在全局信息提取过程中不会丢失像素的空间位置信息, 因此获取到全局信息  $X_G$  后通过 Fold 模块执行 Unfold 的逆操作就可得到  $X_F \in R^{H \times W \times C}$ , 使得像素折叠回原特征图, 过程如图 6 所示, 再将带有全局信息的输出与输入数据按通道维度进行拼接, 增强浅层信息的传递, 通过卷积层降维实现通道间的信息交互, 达到提取关键信息的目的.

### 1.4 特征融合模块

由 conv block 和 global block 提取到不同感受野

的特征信息后, 本文设计特征融合模块 (fusion block) 用于局部和全局特征的融合, 如图 7 所示. 先将两个模块提取到的信息从通道维度进行拼接, 考虑到等权重拼接后的特征信息虽然具备不同尺度的感受野, 但不能关注到不同语义特征间的差异, 缺乏对重要通道特征的鉴别能力, 于是在拼接层后引入了多尺度通道注意力模块 (MSCAM)<sup>[16]</sup>, 该模块使用两个不同尺度的分支来提取全局和局部特征的通道注意力, 利用点卷积获取全局和局部通道注意力信息, 将二者相加后通过 Sigmoid 函数生成权重对输入特征做注意力操作得到输出特征. 计算过程如下所示:

$$g(X) = B(P(\delta(B(P(G(X)))))) \quad (6)$$

$$L(X) = B(P(\delta(B(P(X)))) \quad (7)$$

$$X' = X \cdot \sigma(L(X) + g(X)) \quad (8)$$

其中,  $B$  表示 BatchNorm,  $P$  表示点卷积,  $\delta$  表示 ReLU 激活函数,  $G$  表示全局平均池化,  $g(X)$  和  $L(X)$  分别表示全局和局部通道注意力,  $X$  表示输入特征,  $\sigma$  表示 Sigmoid 函数,  $X'$  表示输出特征. MSCAM 模块能进一步对拼接特征进行权重筛选, 凸显重要通道的特征信息, 使得 fusion block 能够高质量的融合不同感受野的特征信息并将其传递到高层网络.

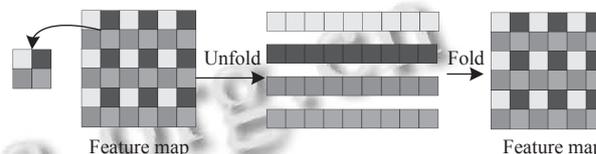


图 6 Unfold 与 Fold 流程图

### 1.5 模型架构

综合以上设计, 本文模型架构如图 8 所示, 最终模型采用并行结构避免因串联在信息传递时造成丢失, 在改进基础模型结构上加入了全局特征提取模块和特征融合模块, 主要由下采样模块、卷积模块、全局特征提取模块和特征融合模块重复堆叠而成, 具体结构如表 1 所示, 全局特征提取模块中多头自注意力的引入能弥补 CNN 在浅层网络感受野小无法获取全局信息的缺陷, 而卷积模块也能帮助其克服获取全局依赖关系时需要大量数据, 导致训练困难的问题. 通过融合各阶段提取到的全局与局部信息, 使用跳跃连接, 能最大程度地减少信息提取不全和信息丢失的问题, 有效提高模型肺炎识别准确率.

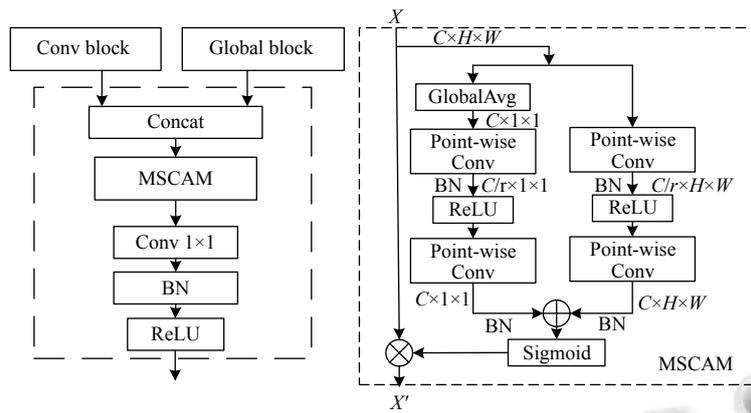


图7 特征融合模块

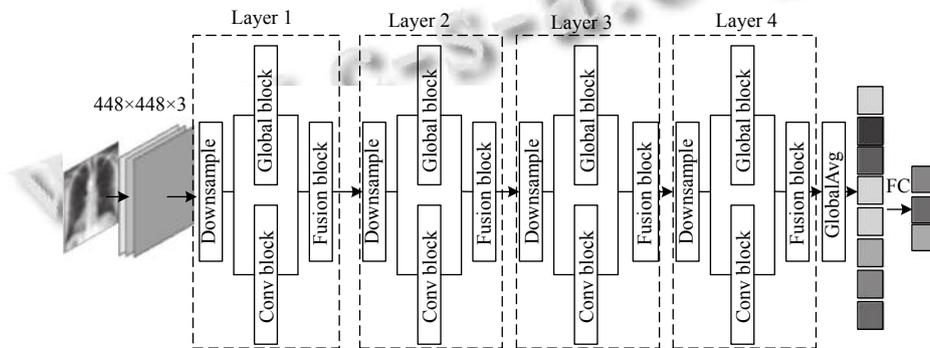


图8 模型框架

表1 模型具体结构

Layer	Layer structure	Output size
Layer 1	{ Downsample, Conv block × 2, Global block, Fusion block }	112 <sup>2</sup> ×48
Layer 2	{ Downsample, Conv block × 2, Global block, Fusion block }	56 <sup>2</sup> ×96
Layer 3	{ Downsample, Conv block × 3, Global block, Fusion block }	28 <sup>2</sup> ×192
Layer 4	{ Downsample, Conv block × 2, Global block, Fusion block }	14 <sup>2</sup> ×384
分类层	Global average pool Full connected	1×3

COVID-19 检测挑战 (<https://cxr-covid19.grand-challenge.org/>) 所使用的数据集, 该数据集包含 3 个数据类别, 分别是正常、普通肺炎、新冠肺炎胸部影像图片, 数据分布情况如表 2 所示, 一共 21 390 张胸部 X 射线图像, 格式为 JPG 和 PNG。

表2 数据分布情况

数据类型	正常	普通肺炎	新冠肺炎	总计
训练集	7151	4273	6534	17958
测试集	1000	1000	1432	3432
总计	8151	5273	7966	21390

## 2 实验及结果分析

### 2.1 实验环境

本文实验在 Windows 10 操作系统下进行, 使用 PyTorch 深度学习框架搭建网络模型, 对应版本为 1.11.0, GPU 使用 GTX 1080 Ti, 内存为 11 GB, 使用 PyCharm 开发工具, Python 版本为 3.8。

### 2.2 数据集及数据预处理

本文使用的数据集是 Grand Challenge 上的胸部

因为图像数据大小不一, 为了适配模型同时减少计算量, 将每张图像缩放至 512×512 像素, 再使用中心裁剪将图像裁剪为 448×448 像素大小并使用随机水平翻转对数据做轻微数据增强, 针对肺炎图像病灶区域对比度差, 病灶模糊的情况, 使用 CLAHE 算法对数据进行优化, 增强细节信息表现, 图像数据处理前后对比如图 9 所示, 最后将图像进行归一化处理完成数据处理过程。



(a) 处理前 (b) 处理后

图9 图像预处理前后对比图

### 2.3 参数设置与训练策略

实验采用可自适应调整学习率的 AdamW 优化器, AdamW 优化器优化了 Adam 权重衰减的缺陷, 对学习率不敏感, 可以减少调参实验代价. 损失函数使用交叉熵损失函数, 学习率采用等间隔调整学习率, 初始学习率为 0.00015, 每训练 8 轮后学习率下降为当前的 0.8 倍, 每一个批次训练 20 张图片, 训练迭代次数设置为 60 轮, 为防止模型过拟合设置了提前停止 (early stopping), 当模型连续 10 轮损失函数未下降便停止训练, 并保留之前表现最好的模型参数.

### 2.4 评价指标

本文使用准确率 (accuracy,  $ACC$ ), 灵敏度 (sensitivity,  $SEN$ ) 和特异性 (specificity,  $SPE$ ) 这 3 种评价指标对模型进行评估. 准确率指的是模型正确识别肺炎类型的样本数占所有样本数的百分比, 灵敏度指的是召回率, 用来衡量被正确辨别的正例个数占所有正类的比例, 特异性指的是被正确辨别的负类个数占所有负类的比例, 它们的公式如式 (9)–式 (11) 所示:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$SEN = \frac{TP}{TP + FN} \quad (10)$$

$$SPE = \frac{TN}{TN + FP} \quad (11)$$

其中,  $TP$  表示真阳性, 为正确划分为正例的数量,  $TN$  表示真阴性, 为正确划分为负例的数量,  $FP$  表示假阳性, 为错误划分为正例的数量,  $FN$  表示假阴性, 为错误划分为负例的数量. 因为本实验数据共 3 个类别, 针对三分类实验, 在计算评价指标的时候采用宏平均 (macro averaging) 的方法来衡量模型, 先计算每个类别的评价指标, 然后将每个类别得到的结果求和算出平均值, 公式如式 (12)–式 (14) 所示, 其中  $ACC_i$ 、 $SEN_i$ 、 $SPE_i$  分别表示第  $i$  个类别的准确率、灵敏度、特异性,  $n$  为类别总数.

$$\overline{ACC} = \frac{1}{n} \sum_{i=1}^n ACC_i \quad (12)$$

$$\overline{SEN} = \frac{1}{n} \sum_{i=1}^n SEN_i \quad (13)$$

$$\overline{SPE} = \frac{1}{n} \sum_{i=1}^n SPE_i \quad (14)$$

### 2.5 实验结果与分析

为了评估模型的有效性, 本文在当前数据集上与其他优秀分类模型和现有研究成果进行三分类对比, 选择的模型包括: ResNet50、VGG16<sup>[17]</sup>、EfficientNet<sup>[18]</sup>、DenseNet<sup>[19]</sup>、MobileViT, 用同样的实验方法对这些模型进行实验, 实验结果比较如表 3 所示, 从表中可以看出, 本文模型在准确率、灵敏度、特异性这 3 个评价指标上均优于以上模型. 此外, 与其他研究人员的研究成果进行比较, 其中 Mohan 等<sup>[20]</sup> 提出了 Siamese Network 模型, 使用迁移学习 (transfer learning) 和一次性学习 (one shot learning) 的方法结合两种损失函数对肺炎进行分类, 因该文并未使用准确率和特异性作为评价指标, 所以本文就灵敏度这一指标与该模型进行比较, 结果优于该模型. Zhan 等<sup>[21]</sup> 提出了一种多模态融合网络, 通过热图来突出图像中的空间和结构特征信息, 该文仅使用准确率作为评价指标, 从表 3 对比可知本文方法准确率明显优于该方法. Azeem 等<sup>[22]</sup> 使用迁移学习的方法在预训练好的模型基础上进行微调对肺炎进行分类, 本文方法在 3 个指标上均优于该方法. 对比实验表明本文方法在肺炎分类任务上有更好的分类效果.

表3 各模型三分类结果比较 (%)

Network	$\overline{ACC}$	$\overline{SEN}$	$\overline{SPE}$
ResNet50	93.47	92.89	96.86
VGG16	92.68	92.27	96.48
EfficientNet	93.96	93.60	97.10
DenseNet	94.84	94.36	97.50
MobileViT	94.72	94.28	97.49
Mohan等 <sup>[20]</sup>	—	95.00	—
Zhan等 <sup>[21]</sup>	95.48	—	—
Azeem等 <sup>[22]</sup>	94.00	94.00	94.20
Ours	96.21	95.79	98.20

另外, 为了验证模型的轻量化, 本文针对参数和计算量与其他模型进行对比实验, 结果如表 4 所示, 从表中可以看出, VGG16 模型的参数量和计算量最大, 而 MobileViT 参数和计算量较少, 但分类效果一般, 本文提出的混合模型与其他模型相比不但参数量和计算量

更少,且分类效果也是最佳,证明了本文模型能在提高肺炎识别准确率的同时保证模型的轻量化。

表4 各模型参数和计算量比较

Network	Params (M)	FLOPs (G)
ResNet50	23.5	16.5
VGG16	442.5	61.8
EfficientNet	10.7	4.0
DenseNet	6.9	11.5
MobileViT	4.9	5.8
Ours	3.6	3.7

最后,采用消融实验来验证改进的基础网络模型、MSCAM、MSA的有效性,并做了5组对比实验,实验结果如表5所示。

表5 消融实验对比 (%)

Number	Network	$\overline{ACC}$	$\overline{SEN}$	$\overline{SPE}$
①	MobileNetV2	93.65	92.73	96.84
②	Basic network	94.63	94.05	97.40
③	Basic network+MSCAM	95.60	95.06	97.89
④	Basic network+MSA	95.68	95.21	97.95
⑤	Basic network+MSCAM+MSA	96.24	95.83	98.22

首先本文的基础网络模型是在 MobileNetV2 的倒残差结构上进行设计改进,所以实验①使用 MobileNetV2 进行分类实验,与实验②改进的基础网络模型进行对比。实验③和实验④分别在基础网络模型中加入特征

融合模块和全局特征提取模块来验证两个模块的有效性,实验⑤则是加入所有模块构成本文的最终模型。从表5中数据可以看出,与原模型相比改进的基础网络3个指标均有提升,在基础网络上分别加入特征融合模块和全局特征提取模块后3个指标也有明显提升,当加入所有模块后准确率,灵敏度,特异性较原模型分别提高2.59%,3.1%,1.38%,此外,本文使用 Grad-CAM 可视化工具对5组实验模型进行特征图可视化生成热力图来进一步探究各模块的有效性,如图10所示,图中颜色暖色越深代表模型对该区域关注度越高,由图10可知,基础网络在先进轻量化模型的基础上进行改进后,对肺炎影像的胸部区域关注度更高,在基础网络上添加 MSCAM 模块后,网络的关注范围变得更广,关注区域也增多,而在基础网络上添加 MSA 模块后,与 MSCAM 模块相比,网络对感兴趣的区域定位更加准确集中。图中最后一列是本文模型所生成的热力图,在改进的基础网络上同时添加了 MSCAM 模块和 MSA 模块后,模型的关注区域不但广,而且定位更加准确,都在肺部的关键区域,MSA 模块保留了 MSCAM 模块关注到的关联程度高的区域,并对冗余关注区域进行了消除,使模型能够更准确地获取重要特征信息,也充分证明了两个模块的有效性。

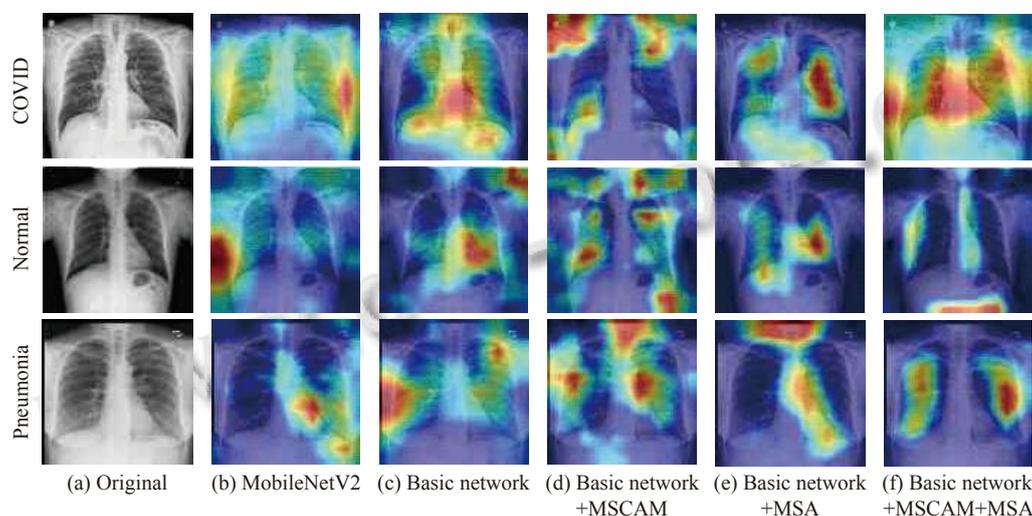


图10 消融实验各模型热力图对比

### 3 结论与展望

本文在卷积神经网络和注意力机制的基础上设计了一种结合二者优点的双分支轻量模型用于肺炎的识别诊断,首先在先进轻量化模型的基础上做改进,通过增大卷积核提高模型的有效感受野,使用深度可分离卷积来保证模型的轻量化,加入 BN 层加速模型收敛,

并引入高效通道注意力模块增强通道之间的信息交互,提高了重要通道的信息表达能力。此外,引入多头自注意力弥补了卷积神经网络在低层缺乏全局感受野的缺点,并设计特征融合模块融合局部与全局信息,使用 CLAHE 算法对数据进行优化,凸显病灶特征。实验结果表明,相比于其他方法,本文方法在肺炎分类任务中

各方面指标均有较大提高,并能在保证模型轻量化的同时取得更好的分类效果,具有较强的实用性,但是目前模型仍有缺点,在实验中发现普通肺炎类型的识别准确率不如其他类,所以后续工作将提高模型对普通肺炎类型的识别精度,并拓展模型对其他肺部疾病的识别诊断,进一步提高模型的泛化能力。

### 参考文献

- 1 唐江平,周晓飞,贺鑫,等.基于深度学习的新型冠状病毒肺炎诊断研究综述.计算机工程,2021,47(5):1-15.[doi:10.19678/j.issn.1000-3428.0060509]
- 2 Li K, Zheng FB, Wu PP, *et al.* Improving pneumonia classification and lesion detection using spatial attention superposition and multilayer feature fusion. Electronics, 2022, 11(19): 3102. [doi: 10.3390/electronics11193102]
- 3 Hussain E, Hasan M, Rahman MA, *et al.* CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images. Chaos, Solitons & Fractals, 2021, 142: 110495. [doi: 10.1016/j.chaos.2020.110495]
- 4 Ikechukwu AV, Murali S, Deepu R, *et al.* ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of pneumonia from chest X-ray images. Global Transitions Proceedings, 2021, 2(2): 375-381. [doi: 10.1016/j.gltip.2021.08.027]
- 5 李镛,王旭,关欣.一种结合三重注意力机制的双路径网络胸片疾病分类方法.电子与信息学报,2023,45(4):1412-1425.[doi:10.11999/JEIT220172]
- 6 谢娟英,夏琴.新冠肺炎 CXR 图像分类新模型 COVID-SERA-NeXt.太原理工大学学报,2022,53(1):52-62.[doi:10.16355/j.cnki.issn1007-9432tyut.2022.01.007]
- 7 Raghu M, Unterthiner T, Kornblith S, *et al.* Do vision transformers see like convolutional neural networks? Proceedings of the 35th Conference on Neural Information Processing Systems. NeurIPS, 2021. 12116-12128.
- 8 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017. 6000-6010. [doi: 10.5555/3295222.3295349]
- 9 Sandler M, Howard A, Zhu ML, *et al.* MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4510-4520.
- 10 Howard AG, Zhu ML, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
- 11 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770-778. [doi: 10.1109/CVPR.2016.90]
- 12 Wang QL, Wu BG, Zhu PF, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 11531-11539.
- 13 Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning. Lille: PMLR, 2015. 448-456. [doi: 10.5555/3045118.3045167]
- 14 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. ICLR, 2021.
- 15 Mehta S, Rastegari M. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer. Proceedings of the 10th International Conference on Learning Representations. ICLR, 2022.
- 16 Dai YM, Gieseke F, Oehmcke S, *et al.* Attentional feature fusion. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021. 3559-3568. [doi: 10.1109/WACV48630.2021.00360]
- 17 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Proceedings of the 3rd International Conference on Learning Representations. San Diego: ICLR, 2014.
- 18 Tan MX, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019. 6105-6114.
- 19 Huang G, Liu Z, van der Maaten L, *et al.* Densely connected convolutional networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 4700-4708. [doi: 10.1109/CVPR.2017.243]
- 20 Mohan V. Detection of COVID-19 from chest X-ray images: A deep learning approach. Proceedings of the 2021 Ethics and Explainability for Responsible Data Science (EE-RDS). Johannesburg: IEEE, 2021. 1-7. [doi: 10.1109/EE-RDS53766.2021.9708594]
- 21 Zhan ZY, Qin YF, Du PY, *et al.* Classification network of COVID-19 based on multi-modality fusion network. Proceedings of the 2021 Ethics and Explainability for Responsible Data Science (EE-RDS). Johannesburg: IEEE, 2021. 1-5. [doi: 10.1109/EE-RDS53766.2021.9708586]
- 22 Azeem MA, Khan MI, Khan SA. COVID-19 detection via image classification using deep learning on chest X-ray. Proceedings of the 2021 Ethics and Explainability for Responsible Data Science (EE-RDS). Johannesburg: IEEE, 2021. 1-4. [doi: 10.1109/EE-RDS53766.2021.9708588]

(校对责编:牛欣悦)