

doi:10.12068/j.issn.1005-3026.2024.12.001

# 融合多尺度注意力机制的冠状病毒肺炎 CT诊断方法

单鹏, 张林, 肖洪明, 赵玉良  
(东北大学秦皇岛分校 控制工程学院, 河北 秦皇岛 066004)

**摘要:** 人工智能诊断是肺部感染的重要辅助诊断方法之一。然而, 现有的方法大多基于深度学习, 具有模型稳定性不足、复杂度高、准确率低的问题。提出融合多尺度注意力机制的浅层模型, 实现了准确率高并且结构简单的新型冠状病毒肺炎CT诊断。首先, 将收集到的两个新型冠状病毒肺炎数据集融合成一个数据集, 解决了因数据集不足导致训练的模型不稳定。其次, 通过在浅层网络 ResNet18 后 3 层中引入多尺度注意力, 弥补了模型特征提取能力的不足。最后, 搭建了一个具有 3 层全连接层的分类器, 改进模型的分类能力, 进而提高了肺部 CT 的分类准确率。结果表明, 所提模型准确率达到 95.41%, 性能超过 ResNet50, ResNet101, VGG16, DenseNet169 等网络, 并且模型参数数量仅有  $12.24 \times 10^6$ , 比 ResNet50 和 VGG16 等网络低 50% 左右。

**关键词:** 肺炎; 深度学习; 多尺度注意力; CT; 分类器

中图分类号: TP 391.4

文献标志码: A

文章编号: 1005-3026(2024)12-1673-07

## CT Diagnosis Method for Coronavirus Pneumonia with Integrated Multi-scale Attention Mechanism

SHAN Peng, ZHANG Lin, XIAO Hong-ming, ZHAO Yu-liang

(School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China.  
Corresponding author: ZHANG Lin, E-mail: 2272326@stu.neu.edu.cn)

**Abstract:** Artificial intelligence (AI)-based diagnosis has become an important auxiliary method for detecting lung infections. However, most existing approaches rely on deep learning, which are often plagued by issues such as insufficient model stability, high complexity, and low accuracy. This paper presents a shallow model which incorporates a multi-scale attention mechanism to achieve both high accuracy and a simple structure for diagnosing COVID-19 from CT scans. Firstly, two datasets of COVID-19 CT images are combined into a single dataset to address the issue of model instability caused by insufficient data. Secondly, by introducing multi-scale attention (MA) in the final three layers of the shallow ResNet18 network, the model's feature extraction capability is enhanced. Finally, classifier with three fully connected layers (CTFCL) is constructed to improve the classification performance of the model, thereby increasing the accuracy of lung CT classification. Experimental results show that the proposed model achieves an accuracy of 95.41%, outperforming networks such as ResNet50, ResNet101, VGG16, and DenseNet169. Furthermore, the model has only  $12.24 \times 10^6$  parameters, which is approximately 50% fewer than networks like ResNet50 and VGG16.

**Key words:** pneumonia; deep learning; multi-scale attention; CT; classifier

肺炎是一种由多种病原体引起的呼吸道感染, 严重时可导致呼吸衰竭和死亡。尤其是新型冠状病毒造成的肺炎, 其主要通过飞沫传播, 给

收稿日期: 2023-07-03

基金项目: 中央高校基本科研业务费专项资金资助项目(N2023021)。

作者简介: 单鹏(1985-), 男, 河南平舆人, 东北大学副教授。

公共健康带来巨大挑战.为有效控制该肺炎的扩散,及时诊断并且隔离感染者是关键措施.最初检测的唯一有效技术是实时逆转录聚合酶链式反应(reverse transcription-polymerase chain reaction, RT-PCR)<sup>[1]</sup>,但 RT-PCR 既耗时又容易造成误诊.在计算机辅助诊断技术的发展下,利用深度学习技术对新型冠状病毒肺炎进行检测得到了广泛的应用<sup>[2]</sup>.该技术主要利用胸部 X 光和肺部 CT 两种影像学检查来检测感染者.大多数医院都配备了 X 光机,但是胸部 X 光不能准确地鉴别软组织,肺部 CT 扫描则更加有效.针对肺部 CT 图像,研究者设计了有关新型冠状病毒肺炎的多种深度学习诊断模型,主要有两种类型.

一种是以单个网络作为骨干分类网络的模型. CoroNet<sup>[3]</sup>网络在 Xception<sup>[4]</sup>的基础上添加了全连接层和 Dropout 层,还采用了迁移学习的方法以避免过拟合现象.实验结果显示,CoroNet 在小数据集上实现了较高的准确率,同时还具备较低的计算量.然而,由于数据集数量少且患者信息不足,实验的可靠性存在一定限制. Wang 等<sup>[5]</sup>提出了一种全自动深度学习系统,主要用于 COVID-19 的识别分类和预后分析.它包括 3 个部分:肺部区域的自动分割、非肺区域的抑制以及分类和预后分析.作者将 DenseNet121<sup>[6]</sup>和 FPN<sup>[7]</sup>相结合并使用了两次迁移学习,最后通过可视化算法对网络中学习到的特征进行了双重分析.实验结果显示, DenseNet121+FPN 模型的准确率仅达到 85%.相比之下,以 ResNet50<sup>[8]</sup>作为骨干的 3D 网络模型框架 COVNet 性能更出色,该网络优势在于特征提取中包括二维局部和三维全局两部分,为了增强模型的可解释性,研究者使用了 Grad-CAM<sup>[9]</sup>方法对深度学习模型的敏感区域进行了可视化.实验结果表明,COVNet<sup>[10]</sup>具备高准确率和优秀的特征提取能力,然而,其复杂的模型结构和大的计算量在实际应用中存在一定限制.

另一种是以多网络融合的分类模型. Tiwari 等<sup>[11]</sup>提出了一种基于多网络融合的分类模型,该模型采用了 CapsNet 和 VGG16 的结合,展现了一定的准确率,但仅适用于二分类任务. DenseCapsNet<sup>[12]</sup>是一种结合了密集卷积网络和胶囊神经网络的模型,充分利用了 2 种网络的优势,使其更加适用于小规模数据集.该方法在肺炎 CT 图像性能检测中取得了 90.7% 的准确率和

90.9% 的 F1 得分,还实现了 96% 的检测灵敏度. COFE-Net<sup>[13]</sup>是一种基于模糊逻辑原理的分类模型,它将 InceptionV3<sup>[14]</sup>, Inception ResNet-V2<sup>[15]</sup>和 DenseNet201 三种网络相结合,并用于计算决策值以进行肺炎 CT 图像的检测分类.采用 Sugeno 模糊测量明显减少了对病灶区域搜索范围,但确定最佳测量仍需要耗费大量的时间和计算成本.在多个数据集上的实验取得了令人满意的效果.多融合的网络模型将两种或多种基础模型的优势结合起来,明显提升了检测效果.然而,这也导致了网络的深度增加,进而带来计算量和参数数量的增加.

受上述研究启发,结合全局与局部信息提高网络对肺炎 CT 图的特征提取对模型性能具有巨大提升.同时为了兼顾模型的实际应用性,减少模型参数数量和计算量具有重要研究意义.本文首先将收集到的两个新型冠状病毒肺炎数据集融合成一个数据集,解决了因数据集不足导致训练的网络不稳定的问题.其次,通过在浅层网络 ResNet18 后 3 层中引入多尺度注意力(multi-scale attention, MA)模块,弥补了因网络结构简单而特征提取能力低的不足.最后,搭建了一个三层分类器 CTFCL,改进网络模型分类能力,进而提高了肺部 CT 影像的分类准确率.实现了准确率高并且模型结构简单的新型冠状病毒肺炎 CT 图分类,并且易于部署在大多应用场景.

## 1 实验方法

图 1 展示了 RMCNet 模型的网络结构,其主要由 3 部分组成: ResNet18, MA 模块<sup>[16]</sup>以及 CTFCL. CT 图片在进入 RMCNet 网络之前先进行预处理操作,将输入图像的大小统一调整为 224×224,并将训练图像进行随机裁剪,裁剪区域为原图像的 50%,同时应用水平翻转来进行数据增强. ResNet18 作为主干网络,主要负责从原始图像中提取 3 种不同尺度的特征图  $f_1$ ,  $f_2$  和  $f_3$ . MA 模块通过提取  $f_1$ ,  $f_2$ ,  $f_3$  上的多尺度特征,提高网络对不同尺度上的病理特征提取能力;同时,通过注意力池化将特征图  $f_3$  和注意力图  $A$  融合来提高关键特征的表达能力,使网络更能关注到图像中具有辨识性的病理特征区域. 3 层分类器则对融合后的多尺度特征进行病情类别诊断.

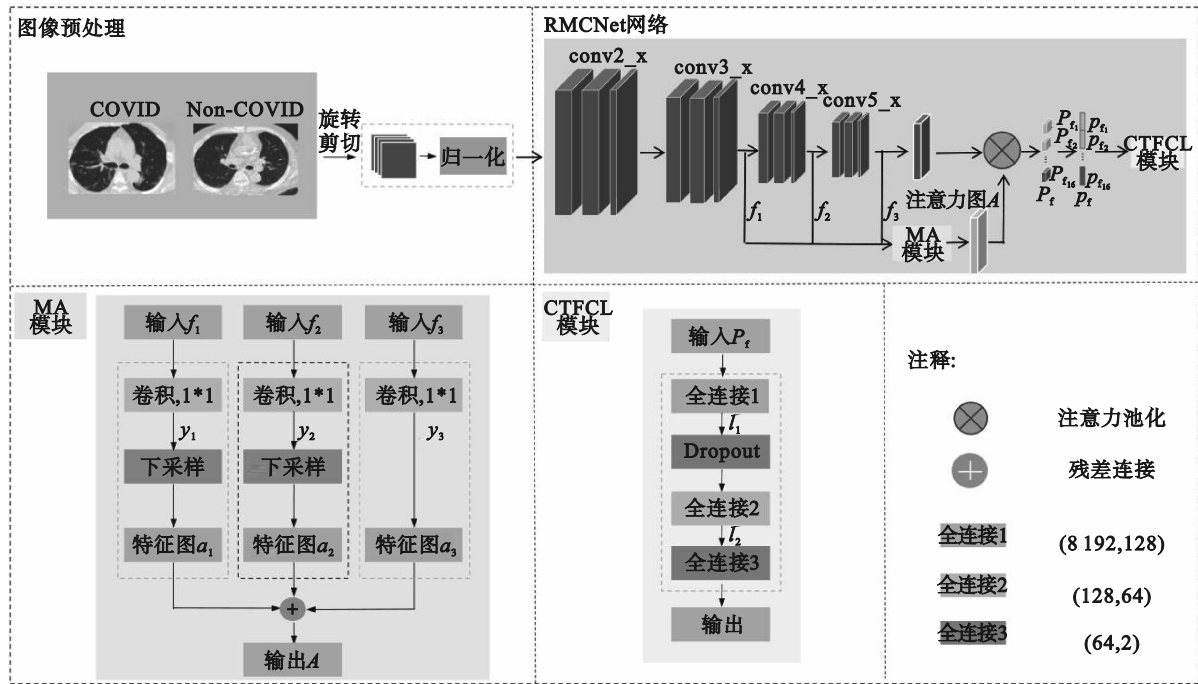


图 1 RMCNet模型结构图

Fig. 1 RMCNet model structure diagram

ResNet18 网络: ResNet18 作为主干网络, 具有 18 层深度, 主要由卷积层、残差块和全连接层构成. 除了输入和输出组件外, ResNet18 的其余层可分为 4 块, 分别为 conv2\_x, conv3\_x, conv4\_x 和 conv5\_x. 每块均由 4 个 3×3 卷积层组成, 输出通道数分别为 64, 128, 256 和 512, 输出特征图大小分别为 56×56, 28×28, 14×14 和 7×7.

MA 模块: 将 ResNet18 中的 conv3\_x, conv4\_x, conv5\_x 块的输出作为 MA 模块在 3 个尺度上的输入  $f_1, f_2, f_3$ , 3 个尺度的特征图的大小分别为 128×28×28, 256×14×14, 512×7×7. 分别利用 3 个 1×1 的卷积对输入特征图的通道数进行调整.

$$y_{i,j,k} = \sum_{c=0}^{C-1} W_{1,1,c,k} \times f_{i,j,c}. \quad (1)$$

式中:  $W$  为卷积核;  $f$  为输入特征图;  $y$  为输出特征图;  $C$  为输入特征图的通道数;  $k$  为输出特征图的通道数;  $i$  和  $j$  分别为输出特征图上的空间位置. 由于卷积核大小为 1×1, 每个输出特征图上的像素点只与输入特征图上对应的像素点以及通道上的权重有关系, 与周围像素点无关. 然后将  $f_1$  和  $f_2$  下采样到 7×7 大小.

$$a_{i,j,k} = \frac{1}{m \times n} \sum_{p=0}^{m-1} \sum_{q=0}^{n-1} X_{i \times s + p, j \times s + q, k}. \quad (2)$$

式中:  $m, n$  为池化区域大小;  $s$  为步幅大小;  $X$  为输入特征图;  $a$  为输出特征图;  $p$  和  $q$  是池化中池化区域的坐标, 其取值范围分别是  $[0, m-1]$  和

$[0, n-1]$ . 每个池化区域内的像素值取平均值作为输出, 以减小特征图的尺寸, 同时可以提高模型的鲁棒性, 减少过拟合; 然后通过残差连接生成注意力图  $A$  (16×7×7).

在注意力池化 (attention pooling, AP)<sup>[17]</sup> 中将注意力图  $A$  和主干网络特征映射  $f_3$  按照式 (3) 进行融合得到特征  $P_{\bar{t}}$  (2 048×7×7).

$$P_{\bar{t}} = A_t \odot f_3 (t = 1, 2, \dots, N). \quad (3)$$

式中: 符号  $\odot$  表示两个张量之间的元素相乘;  $N$  表示注意力图的数量, 本文设为 16. 在 AP 之后, 通过全局平均池化 (global average pooling, GAP) 操作来获得  $p_{\bar{t}}$  (2 048×1×1):

$$p_{\bar{t}} = \frac{1}{H \times L} \times \sum \sum X(u, v) (t = 1, 2, \dots, N). \quad (4)$$

其中:  $X(u, v)$  表示输入特征图的第  $u$  行, 第  $v$  列的元素;  $H$  和  $L$  分别代表输入特征图的高度和宽度. 这一步将特征图压缩成一个向量, 以便于进行分类等任务. 最后将生成的 16 个  $p_{\bar{t}}$  通过式 (5) 拼接得到  $p_t$ .

$$p_t = \text{Cat}([p_{\bar{t}1}, p_{\bar{t}2}, \dots, p_{\bar{t}16}]). \quad (5)$$

分类器中的 3 层全连接层参数分别设置为 (8 192, 128), (128, 64) 和 (64, 2); 在每层的神经元内不使用激活函数, 而使用线性函数.

$$l = M \times p_t + b. \quad (6)$$

其中:  $M$  为权重矩阵;  $p_t$  代表输入向量;  $b$  为偏置

向量;  $I$  是输出向量. 同时为了提高模型的泛化能力, 在第一层全连接后面增加了 Dropout<sup>[18]</sup> 层, 并将其参数概率值设为 0.02.

## 2 实验结果分析

### 2.1 数据集

本文数据集由公共数据集 SARS-CoV-2 CT 和 COVID-CT 合并而成. SARS-CoV-2 CT 数据集<sup>[19]</sup>(见 www.kaggle.com) 由巴西圣保罗医院的 1 252 例 SARS-CoV-2 阳性 (COVID-19) 感染者和 1 230 例未感染者的肺部 CT 图片组成. 另一个数据集<sup>[20]</sup> 包括 349 张 (216 名 COVID-19 患者) 和 463 张 (非 COVID-19 患者) 肺部 CT 图片. 按照 14:3:3 的比例将 3 294 张肺部扫描图片随机划分为训练集、验证集和测试集.

### 2.2 评价指标

为了综合评估不同模型的性能, 采用召回率 (recall)、精确率 (precision)、F1 得分、AUC 值和准确率 (accuracy) 作为评价指标. 真阳性样本数目为 TP, 真阴性样本数目为 TN, 假阳性样本数目为 FP, 假阴性样本数目为 FN. 评价指标的计算公式如下:

1) 准确率 ( $A_c$ ): 分类模型正确预测的样本数量与总样本数量的比例.

$$A_c = \frac{TP + TN}{TP + TN + FP + FN}. \quad (7)$$

2) 精确率 ( $P$ ): 指模型在预测为阳性的样本中, 实际为阳性的比例.

$$P = \frac{TP}{TP + FP}. \quad (8)$$

3) 召回率 ( $R$ ): 模型正确预测为阳性的样本数与实际阳性样本数的比例.

$$R = \frac{TP}{TP + FN}. \quad (9)$$

4) F1 得分: 模型综合指标.

$$F1 = \frac{2 \times P \times R}{P + R}. \quad (10)$$

### 2.3 模型参数实验

#### 2.3.1 超参数设置

RMCNet 在 PyTorch 框架下实现, 在两块 GTX 1080TI GPU 上进行训练. CT 图像大小统一调整为 224×224. 优化器为 ADAM, 损失函数为二元交叉熵<sup>[21]</sup>, 批次大小为 32.

#### 2.3.2 学习率选择

在深度学习模型的训练过程中, 学习率是一个关键的超参数, 它确定了每次迭代中模型参数的更新程度. 适当的学习率可以加快模型的收敛速度, 而过大或过小的学习率则会导致训练不稳定和性能下降等问题.

为了确定最佳的学习率, 本文设计了不同的学习率组合, 例如 0.000 05, 0.000 1, 0.000 2, 采用余弦退火策略, 以 ResNet18+SE+CTFCL (RSCNet) 模型为例, 图 2 给出了不同学习率下 RSCNet 的性能评价指标. 当学习率为 0.000 1 时, 模型在测试集上的准确率、精确率以及召回率值均达到最优; 相比之下, 当学习率为 0.000 2 时, 模型的收敛速度会更快, 但分类性能低; 当学习率为 0.000 05 时, 模型的收敛速度较慢, 其召回率以及精确率比学习率为 0.000 1 的低. 综合考虑, 0.000 1 应作为最优学习率设置, 以达到最佳训练效果.

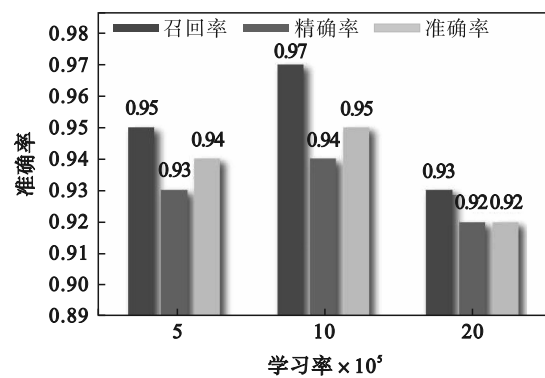


图 2 RSCNet 模型学习率对比

Fig. 2 RSCNet model learning rate comparison

### 2.4 主流算法对比实验

由表 1 可知, 在与主流算法模型对比中 RMCNet 网络在测试集上实现了 95.41% 的准确率, 96.25% 的召回率以及 95.45% 的 F1 指标. 同时随着模型的深度增加, 其分类性能也越好. RMCNet 则通过浅层的网络结构达到了最优的效果, 展现了对大规模数据集依赖性低和模型训练简单的优势. 在图 3a 的雷达图中, 可以直观看出 RMCNet 模型在准确率、召回率和 F1 得分 3 个指标上达到最好, 在 AUC 指标上仅次于 DenseNet169 和 DenseNet201, 并且该模型的多边形面积占比最大, 表明其综合性能最好. 同时表 1 中 RMCNet 模型参数数量仅为  $12.24 \times 10^6$ , 这比 ResNet50 和 ResNet101 模型参数数量少 50% 以上, 比具有网络结构优势的 DenseNet169 和 DenseNet201 少  $9 \times 10^6 \sim 24 \times 10^6$ , 是 VGG16 模型参数数量的十分之一.

此外, RMCNet 模型的每秒浮点操作数 (floating-point operations per second, FLOPS)<sup>[22]</sup> 仅为  $1.82 \times 10^9$ , 比其他模型低 50% 以上. 从参数数

量和 FLOPS 的对比结果来看, RMCNet 模型具有更少的模型参数和更小的计算量, 证明了该模型在保证性能的同时更具有实用性.

表 1 RMCNet 网络与主流算法的对比  
Table 1 Comparison between RMCNet network and mainstream algorithms

模型	召回率	精确率	F1 得分	AUC 值	准确率	参数量	FLOPS
VGG16	0.895 8	0.947 1	0.920 7	0.968 3	0.922 9	$138.36 \times 10^6$	$15.5 \times 10^9$
ResNet50	0.954 1	0.894 5	0.923 3	0.978 2	0.920 8	$25.56 \times 10^6$	$4.12 \times 10^9$
ResNet101	0.912 5	0.956 3	0.933 9	0.968 3	0.935 4	$44.55 \times 10^6$	$7.48 \times 10^9$
DenseNet169	0.945 8	0.953 7	0.949 7	0.985 3	0.950 0	$14.15 \times 10^6$	$3.42 \times 10^9$
DenseNet201	0.945 8	0.945 8	0.945 7	0.982 7	0.945 6	$18.09 \times 10^6$	$4.37 \times 10^9$
RMCNet	0.962 5	0.946 7	0.954 5	0.981 4	0.954 1	$12.24 \times 10^6$	$1.82 \times 10^9$

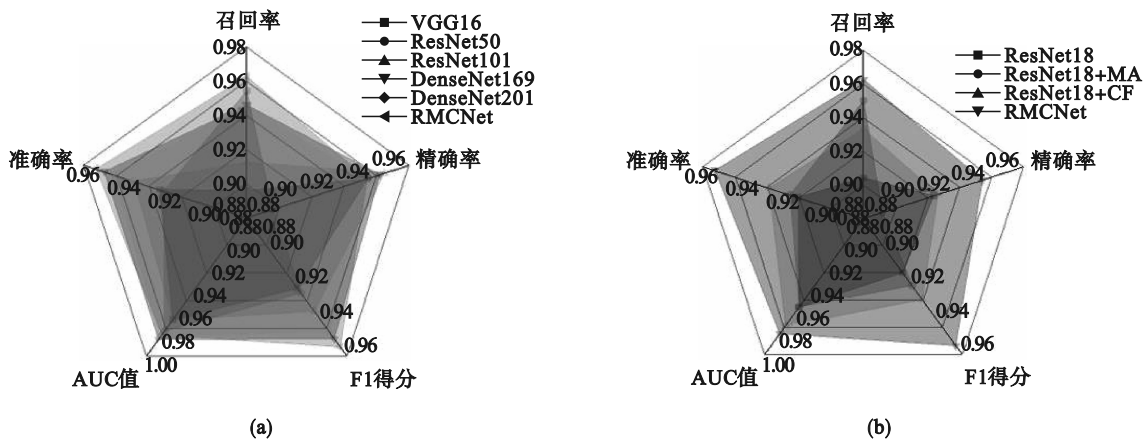


图 3 对比实验雷达图

Fig. 3 Comparison experiment radar map

(a)—RMCNet 与主流算法对比; (b)—RMCNet 消融实验对比.

图 4 绘制了各模型分类结果的混淆矩阵, 行表示实际标签, 列表示模型的预测结果. 结果显示: 在阳性感染者中, RMCNet 模型预测准确率达到 96%, 比其他主流模型高 1%~2%. 这表明该模型在减少假阴性上具有较大优势, 能够实现尽可能多的患者得到正确的诊断和治疗, 防止病毒进一步扩散和治疗延误. 在阴性感染者中, RMCNet 模型准确率也能达到 95%, 比除 DenseNet169 以外的模型都要高, 展现了该模型在减少假阳性上具有较好的表现.

### 2.5 注意力对比实验

本文对比了 3 种注意力机制模块: 压缩和激励 (squeeze and excitation, SE) 模块<sup>[23]</sup>、卷积块注意力模块 (convolutional block attention module, CBAM)<sup>[24]</sup> 以及 MA 模块. 将这些模块分别集成到 ResNet18+CTFCL (RCNet) 模型, 并选取最优学习率 0.000 1. 实验结果如图 5 所示, 使用 MA 模块

的网络在测试集上达到了最高的准确率, 其次是 SE 模块, CBAM 模块性能最低. 这证明同时关注不同尺度的特征是 MA 模块优于其他模块的主要原因.

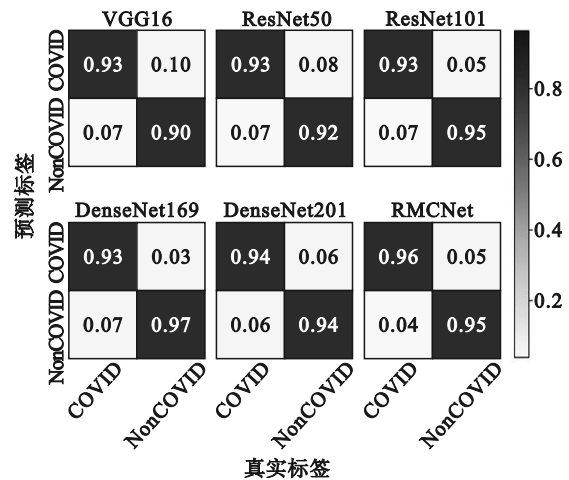


图 4 测试集混淆矩阵

Fig. 4 Test set confusion matrix

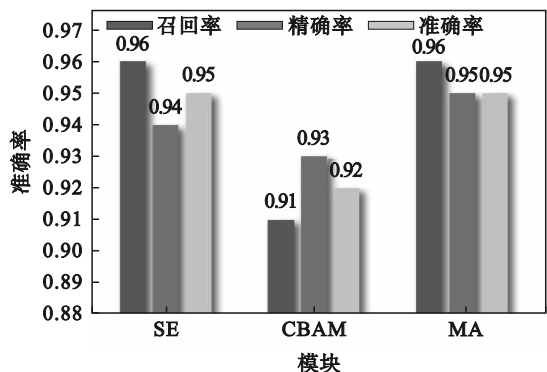


图5 注意力模块对比

Fig. 5 Comparison of attention module

### 2.6 消融实验

为了验证 RMCNet 中多尺度注意力机制和分类器的有效性,本文以 ResNet18 为基础模型设计了消融实验.如表 2 所示,模型 ResNet18+MA 在 ResNet18 网络中引入 MA,相较于 ResNet18 网络,准确率提高了 0.31%,召回率提高了 4.59%,但精确率下降了 2.88%. ResNet18+CTFCL 在 ResNet18 网络中引入 CTFCL,准确率、精确率和召回率都有所提高,证明了分类器的有效性.同时,消融实验中模型的雷达图显示(如图 3b 所示) RMCNet 模型的每个指标都达到了最优,这表明在 ResNet18 的主干网络中,MA 和 CTFCL 在特征提取和分类方面表现出色,两者相互配合提高了胸部 CT 图像分类性能.

表 2 RMCNet 消融实验对比

模型	召回率	精确率	F1 得分	AUC 值	准确率
ResNet18	0.904 1	0.919 4	0.911 7	0.958 1	0.912 5
ResNet18+MA	0.950 0	0.890 6	0.919 3	0.957 7	0.916 6
ResNet18+CTFCL	0.933 3	0.921 8	0.927 5	0.974 5	0.927 1
RMCNet	0.962 5	0.946 7	0.954 5	0.981 4	0.954 1

### 2.7 RMCNet 网络可视化

为了揭示模型在分类决策过程中的依据以及模型关注的敏感性区域,使用 Grad-CAM 技术对肺部 CT 图像分类模型进行可视化分析,如图 6 所示.在数据集中随机选取 5 张阳性 CT,第 1~6 列分别为 VGG 16, ResNet50, ResNet101, DenseNet169, DenseNet201 以及 RMCNet 的可视化结果.通过观察不同的敏感性区域,结果显示 RMCNet 网络对比于 VGG 16, ResNet50, ResNet101 等网络能够更好地关注 COVID-19 感染区域,同时忽略无关的背景区域,从而提高了

预测结果的准确性.与其他算法相比,该模型展现出更好的定位能力和分类性能.这表明,MA 模块在特征提取中起到了关键作用.

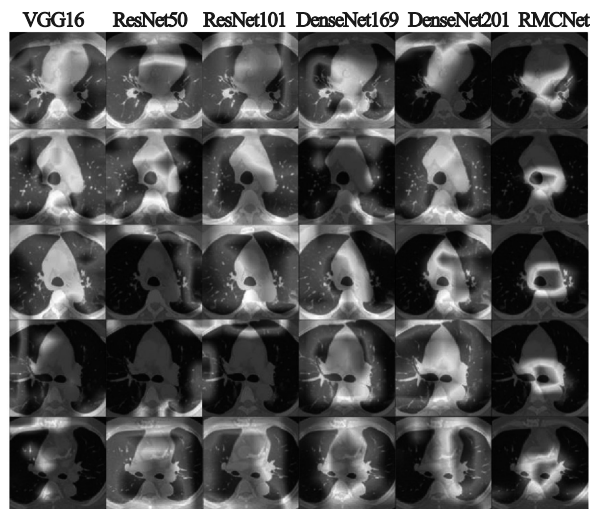


图 6 Grad-CAM 对网络特征可视化

Fig. 6 Grad-CAM visualizes network features

## 3 结 语

在肺炎 CT 图像分类中,现有网络模型存在准确率低、模型复杂和计算量大的问题.为了解决这些问题,本文构建了多尺度注意机制的浅层卷积神经网络对肺炎 CT 图像进行分类的新方法.相比于现有的方法,通过引入多尺度注意力机制,能够更好地捕捉图像中的重要特征,进而提高分类的准确性.其次,采用了 2 个 COVID-19 数据集的组合,扩大训练样本的规模,缓解由于数据集规模较小而引起的模型过拟合和不稳定性问题.最后,引入 3 层全连接分类器进一步提高了模型的性能.在实验中对比其他主流深度学习模型,并证明了该方法在 COVID-19 CT 图像分类任务上的优越性能.实验结果显示, RMCNet 实现了 95.41% 的准确率,并且具有模型结构简单和计算量小的优势,为肺炎 CT 图像的正确诊断和临床决策提供了新的解决方案,具有广泛的应用潜力.

### 参考文献:

[1] Freeman W M, Walker S J, Vrana K E. Quantitative RT-PCR: pitfalls and potential [J]. *Biotechniques*, 1999, 26(1):112-125.

[2] 叶端胜,华云松.基于深度学习的肺炎影像分割研究[J]. *软件工程与应用*, 2022(6):1490-1499. (Ye Duan-sheng, Hua Yun-song. Research on pneumonia image segmentation based on deep learning [J]. *Software*

- Engineering and Applications*, 2022(6):1490–1499.)
- [ 3 ] Khan A I, Shah J L, Bhat M M. CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest X-ray images[J]. *Computer Methods and Programs in Biomedicine*, 2020, 196:105581.
- [ 4 ] Chollet F. Xception: deep learning with depthwise separable convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 1251–1258.
- [ 5 ] Wang S, Zha Y F, Li W M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis [J]. *European Respiratory Journal*, 2020, 56 (2) : 2000775.
- [ 6 ] 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. Honolulu, 2017:4700–4708.
- [ 7 ] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017:2117–2125.
- [ 8 ] Wu Z F, Shen C H, Van Den Hengel A. Wider or deeper: revisiting the ResNet model for visual recognition [J]. *Pattern Recognition*, 2019, 90: 119–133.
- [ 9 ] Li L, Qin L X, Xu Z G, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy [J]. *Radiology*, 2020, 296(2): 65–71.
- [10] Lucas M, Lerma M, Furst J, et al. RSI-Grad-CAM: visual explanations from deep networks via Riemann-Stieltjes integrated gradient-based localization [C]// International Symposium on Visual Computing. Cham: Springer, 2022: 262–274.
- [11] Tiwari S, Jain A. Convolutional capsule network for COVID-19 detection using radiography images [J]. *International Journal of Imaging Systems and Technology*, 2021, 31(2): 525–539.
- [12] Quan H, Xu X S, Zheng T T, et al. DenseCapsNet: detection of COVID-19 from X-ray images using a capsule neural network [J]. *Computers in Biology and Medicine*, 2021, 133: 104399.
- [13] Banerjee A, Bhattacharya R, Bhateja V, et al. COFE-Net: an ensemble strategy for computer-aided detection for COVID-19 [J]. *Measurement*, 2022, 187: 110289.
- [14] Xia X L, Xu C, Nan B. Inception-v3 for flower classification [C]//2017 2nd International Conference on Image, Vision and Computing (ICIVC). Chengdu, 2017: 783–787.
- [15] Ferreira C A, Melo T, Sousa P, et al. Classification of breast cancer histology images through transfer learning using a pre-trained inception ResNet V2 [C]// International Conference on Image Analysis and Recognition. Cham: Springer, 2018: 763–770.
- [16] Li J X, Wang Y Q, Wang S, et al. Multiscale attention guided network for COVID-19 diagnosis using chest X-ray images [J]. *IEEE Journal of Biomedical and Health Informatics*, 2021, 25(5): 1336–1346.
- [17] Fu J L, Zheng H L, Mei T. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017:4438–4446.
- [18] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. *The Journal of Machine Learning Research*, 2014, 15 (1) : 1929–1958.
- [19] Jaiswal A, Gianchandani N, Singh D, et al. Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning [J]. *Journal of Biomolecular Structure and Dynamics*, 2021, 39(15): 5682–5689.
- [20] Polsinelli M, Cinque L, Placidi G. A light CNN for detecting COVID-19 from CT scans of the chest [J]. *Pattern Recognition Letters*, 2020, 140: 95–100.
- [21] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017:2980–2988.
- [22] Li Y S, Chen Y P, Dai X Y, et al. Micronet: improving image recognition with extremely low flops [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, 2021: 468–477.
- [23] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 7132–7141.
- [24] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 3–19.