

文章编号 1004-924X(2024)05-0714-13

DRT Net: 面向特征增强的双残差 Res-Transformer 肺炎识别模型

周 涛^{1,2}, 彭彩月^{1,2*}, 杜玉虎^{1,2}, 党 培^{1,2}, 刘凤珍^{1,2}, 陆惠玲³

- (1. 北方民族大学 计算机科学与工程学院, 宁夏 银川 750021;
2. 北方民族大学 图像图形智能处理国家民委重点实验室, 宁夏 银川 750021;
3. 宁夏医科大学 医学信息与工程学院, 宁夏 银川 750004)

摘要: 针对肺部 X 射线图像的病灶区域较小、形状复杂, 与正常组织间的边界模糊, 使得肺炎图像中的病灶特征提取不充分的问题, 提出了一个面向特征增强的双残差 Res-Transformer 肺炎识别模型, 设计 3 种不同的特征增强策略对模型特征提取能力进行增强。设计了组注意力双残差模块(GADRM), 采用双残差结构进行高效的特征融合, 将双残差结构与通道混洗、通道注意力、空间注意力结合, 增强模型对于病灶区域特征的提取能力; 在网络的高层采用全局局部特征提取模块(GLFEM), 结合 CNN 和 Transformer 的优势使网络充分提取图像的全局和局部特征, 获得高层语义信息的全局特征, 进一步增强网络的语义特征提取能力; 设计了跨层双注意力特征融合模块(CDAFFM), 融合浅层网络的空间信息以及深层网络的通道信息, 对网络提取到的跨层特征进行增强。为了验证本文模型的有效性, 分别在 COVID-19 CHEST X-RAY 数据集上进行消融实验和对比实验。实验结果表明, 本文所提出网络的准确率、精确率、召回率、F1 值和 AUC 值分别为 98.41%, 94.42%, 94.20%, 94.26% 和 99.65%。DRT Net 能够帮助放射科医生使用胸部 X 光片对肺炎进行诊断, 具有重要的临床作用。

关键词: 肺炎识别; X 射线图像; 特征增强; 双残差结构; Transformer

中图分类号: TP399 文献标识码: A doi: 10.37188/OPE.20243205.0714

DRT Net: dual Res-Transformer pneumonia recognition model oriented to feature enhancement

ZHOU Tao^{1,2}, PENG Caiyue^{1,2*}, DU Yuhu^{1,2}, DANG Pei^{1,2}, LIU Fengzhen^{1,2}, LU Huiling³

- (1. College of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China;
2. Key Laboratory of Image and Graphics Intelligent Processing of State Ethnic Affairs Commission,
North Minzu University, Yinchuan 750021, China;
3. School of Medical Information & Engineering, Ningxia Medical University, Yinchuan 750004, China)
* Corresponding author, E-mail: peng_caiyue@163.com

Abstract: Deep learning for lung X-ray image recognition has emerged as a prominent research area. The challenge lies in the small, complexly shaped lesion areas within lung X-rays, where the boundary between the lesion and normal tissue is often unclear, complicating feature extraction in pneumonia images.

收稿日期: 2023-05-11; 修订日期: 2023-07-05.

基金项目: 国家自然科学基金资助项目(No. 62062003); 宁夏自然科学基金资助项目(No. 2022AAC03149)

This paper introduces a Dual Res-Transformer pneumonia recognition model focused on feature enhancement. It incorporates three feature enhancement strategies to augment the model's feature extraction capabilities. The model's key components include: the Group Attention Dual Residual Module (GADRM), which leverages a dual-residual structure for effective feature fusion and enhances local feature extraction through channel shuffle, channel attention, and spatial attention; the Global-Local Feature Extraction Module (GLFEM), which applies at the network's higher levels, merging CNN and Transformer benefits to extract comprehensive global and local image features, thereby boosting the network's semantic feature extraction; and the Cross-layer Dual Attention Feature Fusion Module (CDAFFM), designed to merge shallow network spatial information with deep network channel information, enhancing the network's cross-layer feature extraction. The model's efficacy was validated through ablation and comparative experiments on the COVID-19 CHEST X-RAY dataset. Results demonstrate the network's high performance, with accuracy, precision, recall, F1 score, and AUC values of 98.41%, 94.42%, 94.20%, 94.26%, and 99.65%, respectively. This model offers significant assistance to radiologists in diagnosing various pneumonia cases using chest X-rays, marking a crucial advancement in computer-aided pneumonia diagnosis.

Key words: pneumonia recognition; X-ray image; feature enhancement; dual residual model; Transformer

1 引言

肺炎作为一种常见的胸部疾病,是由微生物和其他环境因素引起的呼吸道感染。它导致肺部积液和呼吸困难,对人体的呼吸系统造成了严重的危害。2019年下半年爆发的新型冠状病毒感染(Corona Virus Disease 2019, COVID-19),已影响全球人口的健康^[1]。传统的肺炎诊断手段主要依赖于医生多年的临床经验,可能存在主观误差,且难以大规模实施。胸部影像学在肺部感染性病变的诊断中具有重要的价值,对患者进行疗效评价可利用高分辨率胸部CT^[2]及X线检查^[3]。利用深度学习技术对患者肺部影像特征进行分析,从而进行病情评估,能够降低复杂的医学数据给医务人员带来的负担,提高诊断效率,辅助医生尽快形成针对性诊疗方案^[4]。

残差神经网络(Residual Neural Network, ResNet)^[5]广泛应用在医学图像领域,目前在肺部疾病的临床辅助诊断方面取得了良好的效果。Zhou等^[6]提出一种基于CT图像的改进的注意力ResNet新冠肺炎识别模型,以适应新冠肺炎病灶区域的特性,实现对于新冠肺炎的识别。Chen等^[7]提出一种基于Inception-ResNet的COVID-19

诊断三分类模型,在Inception-Resnet中使用自注意力机制对肺部病变进行分类,进一步提升了卷积神经网络的分类性能。Huang等^[8]提出了一种名为非局部通道注意力ResNet的深度学习神经网络,将ResNet与非局部模块和信道注意力机制相结合,用于监测COVID-19 PN患者的肺水肿程度,帮助临床医生为患者制定适当的治疗方法。Rajpal等^[9]提出了一种注意力瓶颈残差网络(AB-ResNet)对COVID-19的正常和异常病例进行分类,并通过提出的基于边缘的图切割分割(E-GCS)来定位疾病感染区域,在精度更高的同时实现了有效的网络性能。Chen等^[10]提出双非对称特征学习网络DualCheXNet,用于多标签胸部疾病分类,通过结合ResNet和DenseNet,从胸部图像中捕获更多的判别特征,提高胸部疾病分类性能。吴宣言等^[11]提出一种深层聚合残差密集网络(DLA-RDNet),用于超声图像左心室分割,在下采样部分,结合ResNet与DenseNet的优势提出残差密集网络(RDNet),充分利用所有卷积层的层次信息,实现了较高的增长率。李镛等^[12]提出一种结合三重注意力机制的双路径卷积神经网络(TADPN),将ResNet和DenseNet结合的双路径网络(Double Path Network, DPN)作为骨干网

络,并利用3种不同形式的注意力机制改进DPN,在维持参数量稳定的同时提高网络复杂度,进而提升对胸片疾病的分类精度。

肺部X射线图像病变区域多样化,存在分布广泛、形状复杂、大小不一等特点,且图像中病灶区域存在与周围组织对比度有限、边界不清晰的问题,肺炎感染症状的特异性会导致模型不能很好地关注图像中的病变区域,难以充分提取病变区域的有效特征进行分类。针对上述问题,本文设计了组注意力双残差模块(Group Attention Dual Residual Module, GADRM),采用通道混洗、通道注意力与空间注意力进行高效的特征提取,融合不同通道之间的特征信息,使用两个不同操作的残差连接同时对特征进行重复挖掘利用,增强单个模块对病变区域特征的提取能力;设计了全局局部特征提取模块(Global Local Feature Extraction Module, GLFEM),在特征提取网络末端结合CNN和Transformer的优势,使得网络同时关注全局以及局部的高语义特征信息,进一步增强网络的语义特征提取能力;构造了跨层双注意力特征融合模块(Cross-layer Dual Attention Feature Fusion Module, CDAFFM),利

用空间注意力增强浅层网络的纹理、形状等低语义信息,对深层网络的高语义信息进行通道增强,将二者融合获得更丰富的上下文信息,对网络提取到的跨层特征进行增强。

2 整体网络结构

ResNet的出现有效缓解了由于网络深度增加带来的梯度消失和网络退化问题,能够加快神经网络的训练速度,并且大幅提升深度网络的泛化能力和稳定性,常用于医学图像的肺部分类。然而,肺部X射线图像的病灶区域较小、形状复杂,与正常组织间的边界模糊,模型常常无法提取图像的全局特征和局部特征,且难以聚焦于病灶区域。此外,原始残差网络采用卷积操作进行特征提取,无法获得病灶的全局信息。

为了有效利用X光图像中的全局与局部病变区域特征,提高模型对于不同类型肺炎的识别能力,本文提出面向特征增强的双残差 Res-Transformer肺炎识别模型DRT Net。该模型的整体结构如图1所示,包含GADRM(A)、

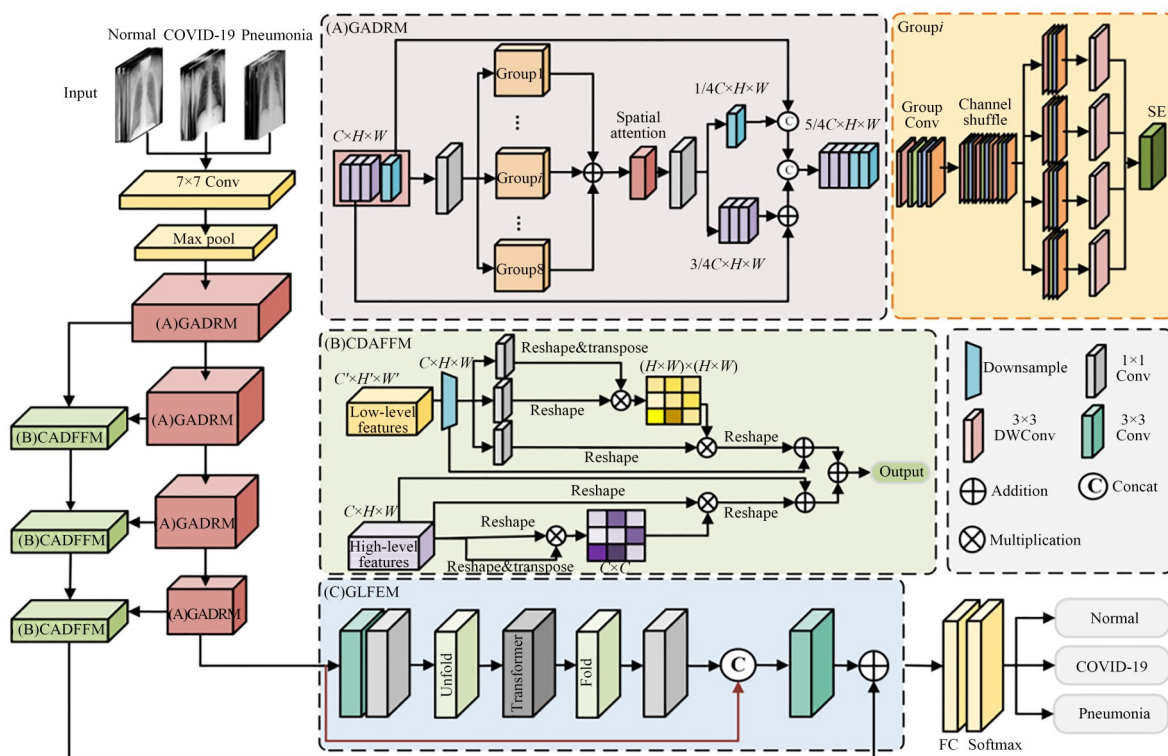


图1 DRT Net整体框架

Fig. 1 Overall framework of DRT Net

步骤一:首先将输入特征 X_r 进行通道混洗,其流程如图 3 所示。首先通过 Reshape 操作将输入通道从一维变成二维,其中一维表示卷积组数,另一维是每个卷积组包含的通道数;然后,进行 Transpose 操作将扩展出的二维进行置换;最后进行 Flatten 展平操作,将置换后的通道展平,完成最后的通道混洗。计算过程如下:

$X'_r = \text{Flatten}(\text{Transpose}(\text{Reshape}(X_r)))$, (1)
其中 $r=1,2,3,4$ 。

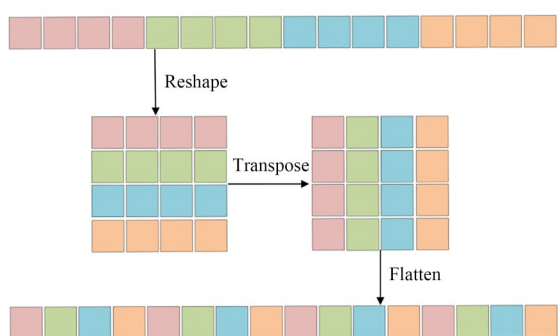


图 3 通道混洗操作

Fig. 3 Channel shuffle operation

步骤二:使用 $\text{DWConv}()$ 表示 GADRM 中 3×3 深度可分离卷积(图 2 中 $3 \times 3 \text{DWConv}$),并使用 Y_r 表示 $\text{DWConv}()$ 的输出。计算过程如下:

$$Y_r = \text{DWConv}(X'_r), \quad (2)$$

式中 $r=1,2,3,4$ 。

步骤三:利用通道映射之间的相互依赖性可以改进特定语义的特征表示。将第 r 组的特征图 Y_r 输入到 SE 模块^[14]。首先,通过自适应全局平均池化(Global Average Pooling, GAP)将特征层的长宽进行压缩,只留下通道维度的信息。然后,使用两个全连接层 $\text{FC}_{\text{Sigmoid}}$ 和 FC_{Relu} 对通道信息进行加权,得到各分组通道的权重:

$$W_r = \text{FC}_{\text{Sigmoid}}(\text{FC}_{\text{Relu}}(\text{GAP}(Y_r))), \quad (3)$$

式中 $r=1,2,3,4$ 。

最后将各通道的权重与各分组特征图相乘,得到通道加权特征图:

$$G_i = W_r \times Y_r, \quad (4)$$

式中: $r=1,2,3,4, i=1,2, \dots, 8$ 。

阶段三:利用空间注意模块来聚合空间关系,使网络更加关注图像的病灶区域,如图 4 所

示。首先将所有组得到的特征图 G_i 进行求和得到:

$$G = \sum_{i=1}^8 G_i, \quad (5)$$

式中 $i=1,2, \dots, 8$ 。

接着对特征图 G 使用全局平均池化(Global Average Pooling, GAP)和全局最大池化(Global Max Pooling, GMP)来获取两种不同的上下文信息;将两个特征图 $\text{GAP}(G) \in \mathbb{R}^{1 \times H \times W}$ 和 $\text{GMP}(G) \in \mathbb{R}^{1 \times H \times W}$ 拼接起来,再通过一个 3×3 的卷积操作获得特征图 G 空间维度的权重:

$$W = \text{Conv}(\text{Concat}[\text{GAP}(G), \text{GMP}(G)]), \quad (6)$$

式中,Concat 表示特征图拼接操作,Conv 表示一个标准的 3×3 卷积操作。

最后在权重 W 和输入特征图 G 之间进行逐元素乘法运算,得到最终的空间注意力图:

$$N = W \odot G. \quad (7)$$

阶段四:最后特征图 N 经过一个 1×1 卷积层操作后,按通道数将其划分成两部分,四分之三个通道的特征图 N_1 被添加到下方跳跃连接路径中,与初始特征图 M_1 相加;另外四分之一通道的特征图与上方跳跃连接路径的初始模块做拼接操作;最终将通道合并作为输出。其数学表示如下:

$$\text{Output} = \text{Concat}[\text{Concat}[M_2, N_2], M_1 + N_1], \quad (8)$$

其中:Output 表示该模块的最终输出,Concat 表示特征图沿通道方向的拼接操作。

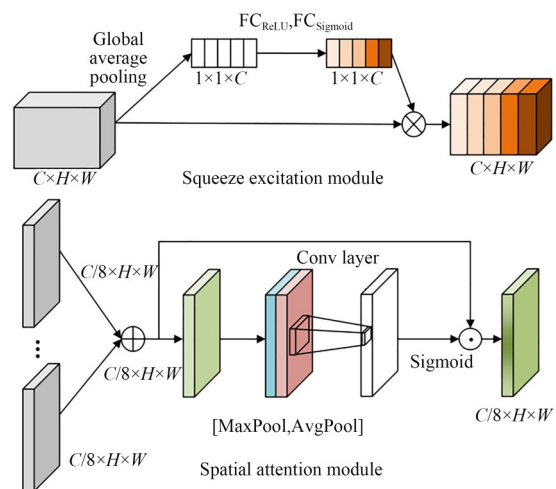


图 4 挤压激励模块和空间注意力模块

Fig. 4 Squeeze excitation module and spatial attention module

2.2 全局局部特征提取模块

CNN具有平移不变性和局部敏感性等归纳偏置,可以很好地捕捉图像细粒度特征和局部信息,是计算机视觉领域的主流模型^[15]。但是CNN感受野有限,现有的CNN方法通常只使用从网络深层提取的高级语义信息进行分类,不具备获取全局信息的能力,这会导致有用信息的丢失,使模型的分类效果不佳。Transformer编码器解码器结构使并行计算得以实现,提高了模型的训练效率,并且利用自注意力机制能够捕获数据之间的长距离信息和依赖关系。

本文在网络深层设计了一个GLFEM同时提取局部信息和全局信息。该模块对高层语义信息进行全局建模,同时融合CNN和Transformer的优势,提取更丰富的特征信息,能够使网络充分提取病变区域的有效特征,从而提升网络对于肺炎疾病的分类性能。GLFEM的具体流程如图5所示,该模块由局部信息编码模块、全局信息编码模块和特征融合模块3个子模块组成,它们进行的操作分别为图像局部特征信息提取、图像全局特征信息提取以及特征信息融合。

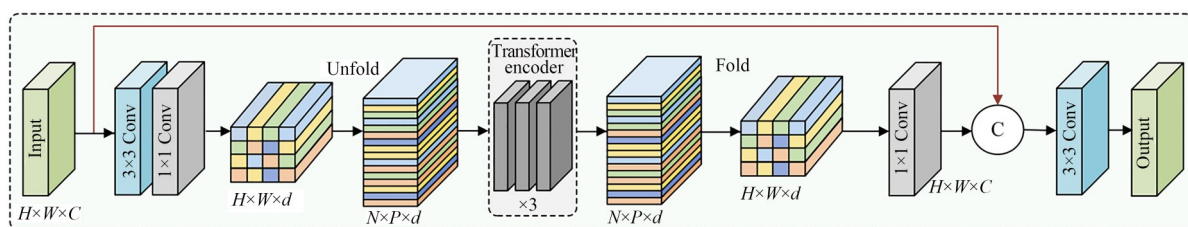


图5 全局局部特征提取模块

Fig. 5 Global local feature extraction module

首先特征图 $X \in \mathbb{R}^{H \times W \times C}$ 进入局部信息编码模块,通过一个卷积核大小为 3×3 的卷积层进行局部信息编码,然后通过一个卷积核大小为 1×1 的卷积层进行通道数的调整,通过学习输入通道的线性组合将张量投影到高维空间,此时 $X \in \mathbb{R}^{H \times W \times d}$ 。

其次,通过 Unfold, Transformer, Fold 结构进行全局的特征建模。为了使网络能够学习具有空间归纳偏置的全局表示,先将输入的特征图划分成 Patch。此时特征图表示为 X_U , 且 $X_U \in \mathbb{R}^{P \times N \times d}$ 。其中 $P=WH$, N 为 Patch 的个数且 $N=HW/P$, H 和 W 分别为 Patch 的高度和宽度。如图5所示,通过 Unfold 操作将相对位置相同的特征图拼接在一块,即图中颜色相同的位置,分别在每个块内进行自注意力计算,相应地减少计算量。然后将拼成的一个序列输入到 Transformer 进行建模。

然后,通过应用 Transformer 来编码 patch 间关系,其数学表示如下:

$$X_G(p) = \text{Transformer}(X_U(p)), 1 \leq p \leq P. \quad (9)$$

通过 Fold 操作将计算完自注意力的特征图重新按照相对位置还原为初始形状。与丢失像素空间顺序的视觉 Transformer 相比,该模块同时保留了 Patch 的顺序与每个 Patch 内像素的空间顺序。

将全局特征建模后的特征块 X_G 通过 1×1 卷积将通道数调整回原始大小,通过一个残差连接与原始输入特征图沿通道方向进行拼接,最后再通过一个 3×3 的卷积层进行特征融合得到输出。

2.3 跨层双注意力特征融合模块

CNN 中的卷积操作在提取特征的同时丢失了底层的纹理细节,使得高层次特征和低层次特征分布在网络两端。高层次特征具有更强的语义信息,但分辨率低且对细节的感知能力较差;浅层特征分辨率高,包含更多位置细节、边缘和纹理等信息,但由于特征提取不充分,其语义性低且噪声多。此外,特征提取网络中不同层关注的信息也是有差异的,利用不同层的特征融合上下文信息能够提升网络的分类性能。然而,简单

的相加易造成信息冗余,并不能充分利用二者的优势,因此本文设计了CDAFFM,用浅层语义信息弥补深层语义信息的缺失,如图6所示。将浅层网络的纹理、形状等低语义信息利用空间注意

力进行增强,将深层网络的高语义信息进行通道增强。筛选后的通道和空间信息相加,使图像浅层上下文信息与深层上下文信息进行高效融合,保留更多有用信息,提升模型的性能。

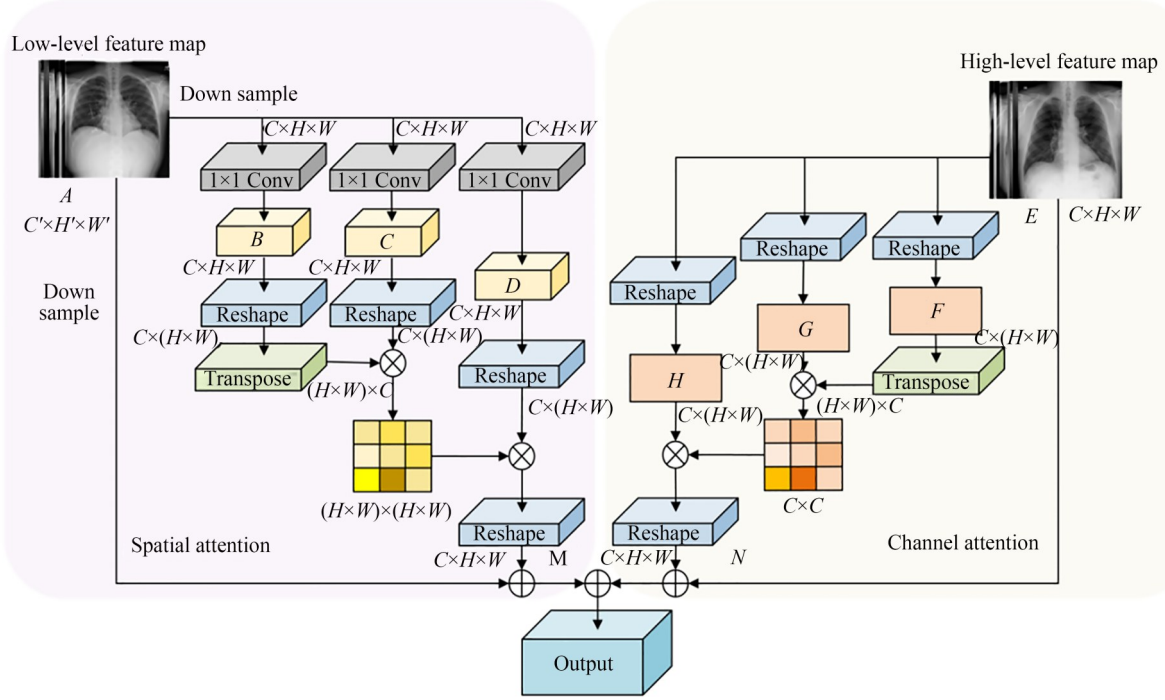


图6 跨层双注意力特征融合模块

Fig. 6 Cross-layer dual attention feature fusion module

对于原始输入的低层特征图A,经过3个相同的 1×1 卷积操作进行特征映射,得到特征图B,C,D。对于特征图B和特征图C,其原始尺寸为 $C \times H \times W$,将二者从三维特征经 Reshape 操作转换到二维特征,尺寸变为 $C \times N (N=H \times W)$,之后将特征图B进行转置,其形状变为 $N \times C$,并与特征图C相乘,获得空间注意力值 W_{spatial} ;特征图D同样经过 Reshape 操作变为 $C \times N$,将特征图D与 W_{spatial} 相乘获得空间维度上筛选后的特征图,最后将特征图 Reshape 为 $C \times H \times W$ 。

空间注意特征图的计算过程如下:

$$S \in \mathbf{R}^{N \times N}; S_{ji} = \frac{\exp(B_i \times C_j)}{\sum_{i=1}^N \exp(B_i \times C_j)}, \quad (10)$$

式中: S_{ji} 度量第 i 个位置对第 j 个位置的影响,即第 i 个位置和第 j 个位置之间的关联程度,其值越大越相似。其输出为:

$$M \in \mathbf{R}^{C \times H \times W}; M_j = \alpha \sum_{i=1}^N (S_{ji} D_i) + A_j, \quad (11)$$

其中: α 表示尺度系数,初始化为0,通过逐渐地学习分配到更大的权重。在每个位置处得到的结果特征M是所有位置上的特征和原始特征的加权和,它具有全局上下文视图,并根据空间注意力图选择性地聚合上下文,使相似的语义特征相互增益,从而提高类内紧凑性和语义一致性。

对于高语义的特征图E,则不经过 1×1 卷积操作进行特征映射,直接进行 Reshape 操作分别获得二维特征图F,G,H。将特征图F进行转置使得其形状变为 $N \times C (N=H \times W)$,转置后的特征图F与G相乘获得通道维度的注意力值 W_{channel} ,将 W_{channel} 与特征图H相乘获得通道维度上筛选后的特征图,最后将特征图 Reshape 为 $C \times H \times W$ 。

通道注意特征图的计算如下:

$$S \in \mathbb{R}^{c \times c}: X_{ji} = \frac{\exp(A_i \times A_j)}{\sum_{i=1}^c \exp(A_i \times A_j)}, \quad (12)$$

式中 X_{ji} 为第 i 个通道对第 j 个通道的影响值。其输出为:

$$N \in \mathbb{R}^{c \times H \times W}: N_j = b \sum_{i=1}^c (X_{ji} A_i) + A_j, \quad (13)$$

其中: β 表示尺度系数,初始化为0,经过逐渐学习分配到更大的权重。每个通道的结果特征为 N ,表示所有通道特征和原始特征的加权和。

最后,将高低层筛选后的特征图相加获得特征结果图。

3 实验及结果分析

3.1 数据集及预处理

本文使用的数据集为公开数据集 COVID-19 CHEST X-RAY DATABASE^[16-17],该数据集来自卡塔尔大学和达卡大学的研究人员以及来自巴基斯坦和马来西亚的合作者与医生合作创建。分类实验选取其中的 COVID-19 阳性病例、正常肺部图像以及病毒性肺炎图像,如图7所示。将数据集经简单筛选后重新分成训练集和验证集。其中,训练集包括2 893张 COVID-19 阳性病例图像、2 400张正常肺部图像以及1 076张病毒性肺炎图像;测试集包括723张 COVID-19 阳性病例图像、600张正常肺部图像以及269

张病毒性肺炎图像。将所有不同尺寸的原始图像缩放至 224×224 像素,然后转换为向量格式并进行像素值归一化处理。

3.2 评价指标

通过对分类模型实验结果的定量对比,能够判断分类模型的优劣。本文主要以准确率(Accuracy, A)、平均精确率(Precision-Macro, P)、平均召回率(Recall-Macro, R)、平均 F1 值(F1 score-Macro, $F1$)和 AUC 值为评价指标,分析了改进的网络模型对肺炎分类效果的影响。

评价指标的数学表达式如下:

$$A = \frac{TN + TP}{FP + TN + TP + FN}, \quad (14)$$

$$P = \frac{TP}{FP + TP}, \quad (15)$$

$$R = \frac{TP}{FN + TP}, \quad (16)$$

$$F1 = \frac{2 \times P \times R}{P + R}, \quad (17)$$

其中: TN , TP , TN 和 FN 分别表示真阳性、假阳性、真阴性和假阴性的数量, TP 是正确标记为阳性的模型预测结果的数量, FP 是错误标记为阳性的模型预测结果的数量, TN 是正确标记为阴性的模型预测结果的数量, FN 是错误标记为阴性的模型预测结果的数量。

本文在三分类模型中单独计算每一类的评价指标,最后求评价指标的算术平均。此外, AUC 被定义为 ROC 曲线下与坐标轴围成的面积,作为数值可以直观地评价分类器的好坏。AUC 越大,分类器效果越好。

3.3 实验环境

实验环境为 Windows Server 2019 Datacenter 的 64 位系统,搭载 Intel Xeon Gold 6154, 3.0GHz x36 CPU 处理器,计算机内存为 256 GB,采用两块并行的 TITAN V 显卡加速图像处理,程序编写采用 Python 语言,基于 GPU 版本的 Pytorch 框架进行网络搭建和训练。使用 Adam 优化器进行优化,学习率衰减设置值为 0.000 1,对于肺部 X 射线数据集的训练周期设为 150,训练批处理大小设置为 8。损失函数使用交叉熵损失。

3.4 消融实验与分析

为了评估模块的有效性,通过不同的网络模

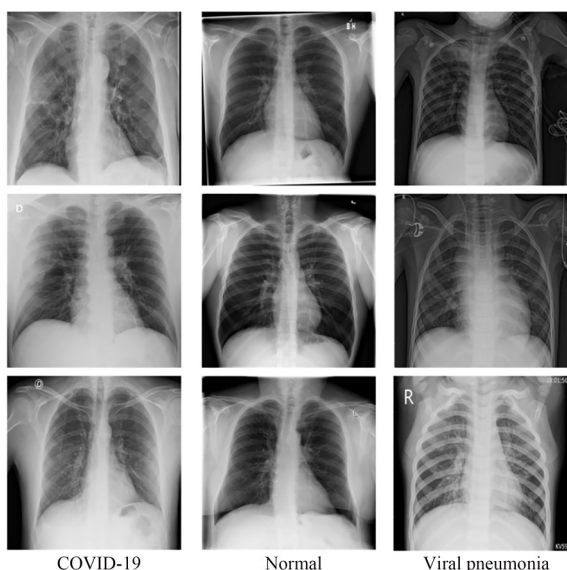


图7 数据集展示

Fig. 7 Dataset display

型来测试各个模块,评估指标包括准确率(A)、宏平均精确率(P)、宏平均召回率(R)、宏平均F1值和AUC值,如表1所示。本文在DPN92网络的基础上依次进行8项实验。Network1:DPN92网络;Network2:添加CDAFFM的DPN92网络;Network3:与GLFEM结合的DPN92网络;Net-

work4:将初始特征提取块改为GADRM的网络;Network5:使用CDAFFM以及GLFEM的DPN92网络;Network6:使用GADRM以及GLFEM的网络;Network7:使用GADRM以及CDAFFM的网络;Network8:本文所提出的DRT Net。

表 1 消融实验结果对比

Tab. 1 Result comparison of ablation experiments

Name	Model	GADRM	GLFEM	CDAFFM	A	P	R	F1	AUC
Network1	DPN92				0.969 2	0.912 8	0.902 6	0.900 3	0.988 4
Network2	DPN92+CDAFFM			✓	0.972 5	0.927 0	0.915 5	0.919 0	0.990 1
Network3	DPN92+GLFEM		✓		0.974 1	0.928 8	0.920 5	0.921 1	0.991 2
Network4	DPN92+GADRM	✓			0.974 7	0.936 3	0.942 7	0.938 6	0.992 8
Network5	DPN92+GLFEM+CDAFFM		✓	✓	0.974 6	0.930 9	0.913 8	0.920 0	0.992 7
Network6	DPN92+GADRM+GLFEM	✓	✓		0.975 5	0.923 5	0.917 3	0.915 3	0.992 0
Network7	DPN92+GADRM+CDAFFM	✓		✓	0.977 9	0.939 1	0.937 6	0.937 4	0.993 0
Network8	DRT Net	✓	✓	✓	0.984 1	0.944 2	0.942 0	0.942 6	0.996 5

与 Network1 相比, Network2 在添加了CDAFFM 模块后的性能参数均有所提升,准确率、精确率、召回率、F1 值与 AUC 值分别提升了 0.33%,1.42%,1.29%,1.87% 与 0.17%,由此证明 CDAFFM 模块能够将不同阶段的特征融合增强,增强了模型对特征的提取能力;Network3 在添加 GLFEM 模块后的准确率、精确率、召回率、F1 值与 AUC 值分别提升了 0.49%,1.60%,1.79%,2.08% 与 0.28%,证明了 GLFEM 模块的有效性;Network4 在添加 GADRM 模块后的准确率提升了 0.55%,精确率提升了 2.35%,召回率提升了 4.01%,F1 值提升了 3.83%,AUC 值提升了 0.44%,由此验证了 GADRM 模块能够使网络具有更好的提取特征。

添加两个模块的 Network5,6,7 的各项评价指标又高于只添加一个模块的 Network2,3,4;添加 GADRM, GLFEM 与 CDAFFM 3 个模块的 DRT Net 性能最好,且与初始的 Network1 模型相比,肺炎分类的准确率从初始的 96.92% 上升到 98.41%,精确率从 91.28% 上升到 94.42%,召回率从 90.26% 上升到 94.20%,F1 值从 90.03% 上升到 94.26%,AUC 值从 98.84% 上升到 99.65%。由此可知,本文所提

出的 DRT Net 性能最优,在肺部 X 射线图像数据集上对于肺炎的分类性能最佳。为了更直观地进行实验对比,绘制消融实验结果雷达图,如图 8 所示。DRT Net 折线位于最外侧,模型性能最优。此外,为了考察不同模型对于三类样本分类预测的标签(Y -Pre)和真实情况(Y -True)之间的差异,本文采用混淆矩阵对消融实验测试结果进行可视化,如图 9 所示。通过混淆矩阵对比可以看出,DRT Net 对三类样本的识别更为均衡且分类效果更优,能够实现肺炎的准确分类。

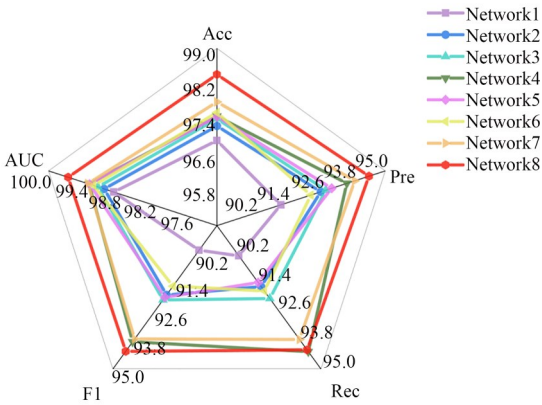


图 8 消融实验结果雷达图

Fig. 8 Radar chart of ablation experiment results

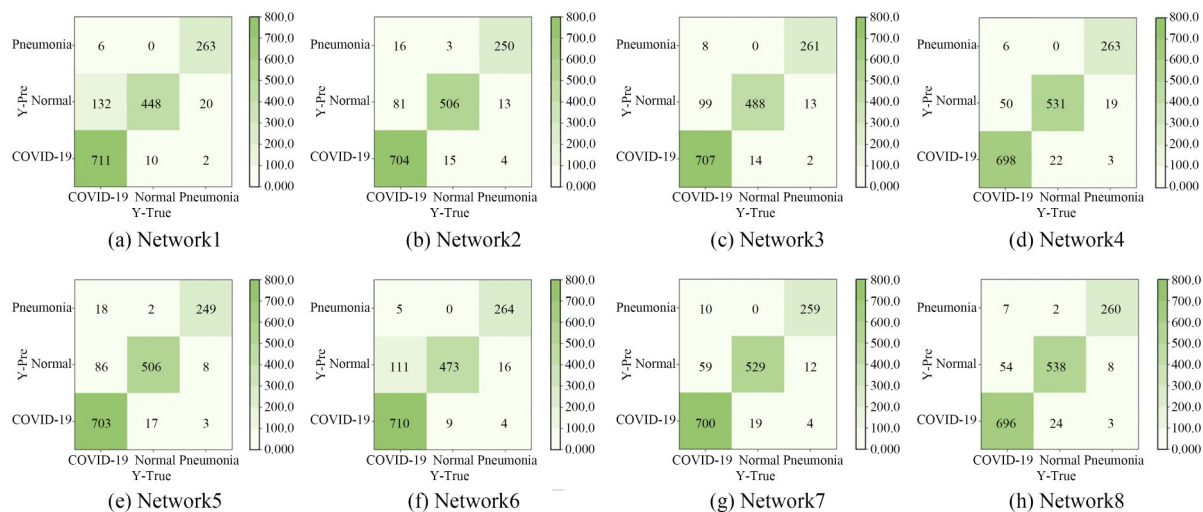


图9 消融实验中各模型的混淆矩阵

Fig. 9 Confusion matrix of each model in ablation experiments

3.5 对比实验与分析

为了验证本文模型对于肺炎的分类能力,在同一数据集^[16-17]上与 ResNet50^[5], ResNet101^[5], Res2Net50^[18], DenseNet121^[19], Resnext101^[20], MobileNetV2^[21]和 DPN92^[22]等网络模型进行对比,实验结果如表2所示。在经过数据预处理的数据集上训练网络,本文模型的准确率为98.41%,精度为94.42%,召回率为94.20%,F1值为94.26%,AUC值为99.65%,性能优于其他网络,具有更好的分类性能。对比实验结果雷达图如图10所示,DRT Net折线位于最外侧,性能最好。本文采用混淆矩阵对各模型测试集的结果进行可视化,结果如图11所示。通过混淆矩阵对比可以看出,DRT Net对于三类样本的识别能力相较其他分类网络更

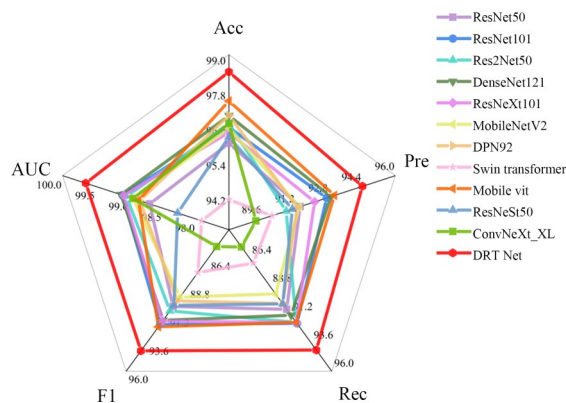


图10 不同模型的肺炎分类结果雷达图

Fig. 10 Radar chart of pneumonia classification results for different models

为均衡且分类效果更优。

表2 各个模型的肺炎X射线图像分类结果

Tab. 2 Classification results of pneumonia X-ray images for each model

Model	A	P	R	F1	AUC
ResNet50 ^[5]	0.959 6	0.914 2	0.907 4	0.905 1	0.986 9
ResNet101 ^[5]	0.965 1	0.927 0	0.919 4	0.920 3	0.991 0
Res2Net50 ^[18]	0.966 4	0.907 5	0.918 4	0.908 8	0.990 4
Densenet121 ^[19]	0.969 1	0.929 0	0.912 5	0.916 7	0.991 2
ResNeXt101 ^[20]	0.963 3	0.921 2	0.918 7	0.917 4	0.990 8
MobileNetV2 ^[21]	0.964 7	0.913 9	0.894 4	0.897 1	0.989 4
DPN92 ^[22]	0.969 2	0.912 8	0.902 6	0.900 3	0.988 4
Swin Transformer ^[23]	0.940 3	0.900 7	0.868 4	0.876 2	0.979 1

续表 2 各个模型的肺炎 X 射线图像分类结果

Tab. 2 Classification results of pneumonia X-ray images for each model

Model	<i>A</i>	<i>P</i>	<i>R</i>	F1	AUC
Mobile vit ^[24]	0.974 1	0.930 7	0.918 4	0.922 3	0.988 6
ReNeSt50 ^[25]	0.961 8	0.910 5	0.902 8	0.904 1	0.982 7
ConvNeXt_XL ^[26]	0.966 4	0.893 0	0.854 6	0.854 3	0.989 4
DRT Net	0.984 1	0.944 2	0.942 0	0.942 6	0.996 5

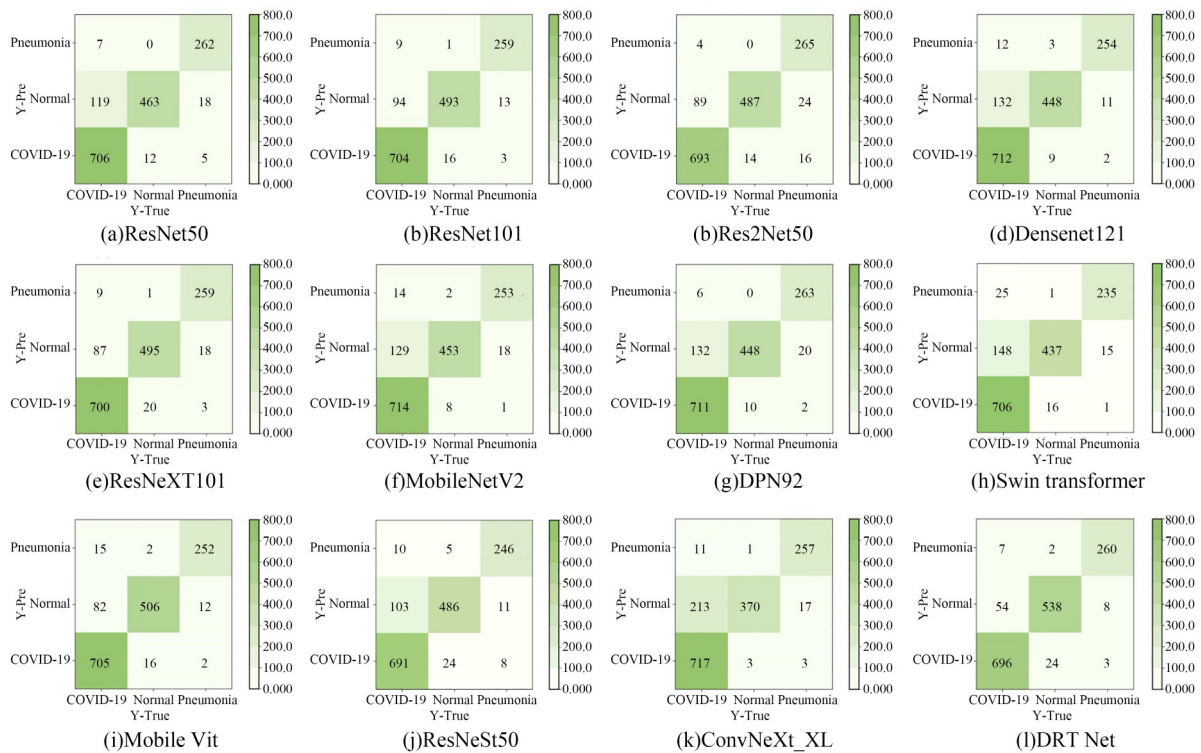


图 11 各模型分类结果的混淆矩阵

Fig. 11 Confusion matrix of classification results for each model

4 结 论

本文提出了一种面向特征增强的双残差 Res-Transformer 肺炎识别模型 DRT Net, 设计了 GADRM, 采用双残差结构进行高效的特征融合, 将通道混洗、通道注意力、空间注意力与双残差结构结合, 提升模型对病灶区域特征的提取能力; 在网络末端采用 GLFEM, 结合 CNN 和 Transformer 的优势使网络充分提取图像的全局特征和局部特征, 对高层语义信息进行全局建

模, 获得高层语义信息的全局特征; 设计了 CDAFFM, 融合浅层网络的纹理、边缘等空间信息以及深层网络的通道信息, 进一步增强网络的特征提取能力。在 COVID-19 CHEST X-RAY 数据集上进行相关实验, 实验结果表明, DRT Net 网络的准确率、精确率、召回率、F1 值和 AUC 值分别为 98.41%, 94.42%, 94.20%, 94.26% 和 99.65%。该模型能够辅助放射科医生使用胸部 X 光影像诊断肺炎病例, 对患者及时展开针对性的治疗。

参考文献:

[1] ZHOU T, LIU F Z, LU H L, *et al.* A review of

deep learning imaging diagnostic methods for COVID-19[J]. *Electronics*, 2023, 12(5): 1167.

- [2] TORRES-RAMIREZ C A, TIMARAN-MONTE-NEGRO D, MATEO-CAMACHO Y S, *et al.* CT-based pathological lung opacities volume as a predictor of critical illness and inflammatory response severity in patients with COVID-19[J]. *Heliyon*, 2022, 8(12): e11908.
- [3] REN K, GU Y, LUO M S, *et al.* Deep-learning-based denoising of X-ray differential phase and dark-field images[J]. *European Journal of Radiology*, 2023, 163: 110835.
- [4] MAMALAKIS M, SWIFT A J, VORSELAARS B, *et al.* DenResCov-19: a deep transfer learning network for robust automatic classification of COVID-19, pneumonia, and tuberculosis from X-rays[J]. *Computerized Medical Imaging and Graphics*, 2021, 94: 102008.
- [5] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016. Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [6] ZHOU T, CHANG X Y, LIU Y C, *et al.* COVID-ResNet: COVID-19 recognition based on improved attention ResNet[J]. *Electronics*, 2023, 12(6): 1413.
- [7] CHEN Y F, LIN Y L, XU X D, *et al.* Classification of lungs infected COVID-19 images based on inception-ResNet [J]. *Computer Methods and Programs in Biomedicine*, 2022, 225: 107053.
- [8] HUANG Q H, LEI Y, XING W Y, *et al.* Evaluation of pulmonary edema using ultrasound imaging in patients with COVID-19 pneumonia based on a non-local channel attention ResNet [J]. *Ultrasound in Medicine & Biology*, 2022, 48(5): 945-953.
- [9] AHILA T, SUBHAJINI A C. E-GCS: detection of COVID-19 through classification by attention bottleneck residual network [J]. *Engineering Applications of Artificial Intelligence*, 2022, 116: 105398.
- [10] CHEN B Z, LI J X, GUO X B, *et al.* Dual-CheXNet: dual asymmetric feature learning for thoracic disease classification in chest X-rays[J]. *Biomedical Signal Processing and Control*, 2019, 53: 101554.
- [11] 吴宣言, 缙新科, 朱子重, 等. 深层聚合残差密集网络的超声图像左心室分割[J]. *中国图象图形学报*, 2020, 25(9): 1930-1942.
- WU X Y, GOU X K, ZHU Z ZH, *et al.* Left ventricular segmentation on ultrasound images using deep layer aggregation for residual dense networks [J]. *Journal of Image and Graphics*, 2020, 25(9): 1930-1942. (in Chinese)
- [12] 李镔, 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[J]. *电子与信息学报*, 2023, 45(4): 1412-1425.
- LI Q, WANG X, GUAN X. A dual-path network chest film disease classification method combined with a triple attention mechanism [J]. *Journal of Electronics & Information Technology*, 2023, 45(4): 1412-1425. (in Chinese)
- [13] 周涛, 刘赞臻, 陆惠玲, 等. ResNet 及其在医学图像处理领域的应用: 研究进展与挑战[J]. *电子与信息学报*, 2022, 44(1): 149-167.
- ZHOU T, LIU Y C, LU H L, *et al.* ResNet and its application to medical image processing: research progress and challenges [J]. *Journal of Electronics & Information Technology*, 2022, 44(1): 149-167. (in Chinese)
- [14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018. Salt Lake City, UT. IEEE, 2018: 7132-7141.
- [15] 范丽丽, 赵宏伟, 赵浩宇, 等. 基于深度卷积神经网络的目标检测研究综述[J]. *光学精密工程*, 2020, 28(5): 1152-1164.
- FAN L L, ZHAO H W, ZHAO H Y, *et al.* Survey of target detection based on deep convolutional neural networks [J]. *Optics and Precision Engineering*, 2020, 28(5): 1152-1164. (in Chinese)
- [16] CHOWDHURY M E H, RAHMAN T, KHAN-DAKAR A, *et al.* Can AI help in screening viral and COVID-19 pneumonia? [J]. *IEEE Access*, 2020, 8: 132665-132676.
- [17] RAHMAN T, KHANDAKAR A, QIBLAWEY Y, *et al.* Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images [J]. *Computers in Biology and Medicine*, 2021, 132: 104319.
- [18] GAO S H, CHENG M M, ZHAO K, *et al.* Res2Net: a new multi-scale backbone architecture [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(2): 652-662.
- [19] HUANG G, LIU Z, VAN DER MAATEN L, *et al.* Densely connected convolutional networks

- [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017. Honolulu, HI. IEEE, 2017: 2261-2269.
- [20] XIE S N, GIRSHICK R, DOLLAR P, *et al.* Aggregated residual transformations for deep neural networks[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017. Honolulu, HI. IEEE, 2017: 1492-1500.
- [21] SANDLER M, HOWARD A, ZHU M L, *et al.* MobileNetV2: inverted residuals and linear bottlenecks[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018. Salt Lake City, UT. IEEE, 2018: 4510-4520.
- [22] CHEN Y P, LI J N, XIAO H X, *et al.* Dual path networks[C]. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. December 4-9, 2017, Long Beach, California, USA. ACM, 2017: 4470-4478.
- [23] LIU Z, LIN Y T, CAO Y, *et al.* Swin Transformer: hierarchical Vision Transformer using Shifted Windows[C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021. Montreal, QC, Canada. IEEE, 2021: 10012-10022.
- [24] ALAM N, KOLAWOLE S, SETHI S, *et al.* Vision transformers for mobile applications: a short survey [J]. *arXiv preprint arXiv: 2305.19365*, 2023
- [25] ZHANG H, WU C R, ZHANG Z Y, *et al.* ResNeSt: split-attention networks [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 19-20, 2022. New Orleans, LA, USA. IEEE, 2022: 2736-2746.
- [26] LIU Z, MAO H, WU C, *et al.* A ConvNet for the 2020s[J]. *arXiv preprint arXiv: 2201.03545*, 2022.

作者简介:



周 涛(1977—),男,宁夏同心人,博士,教授,2010年于西北工业大学获得博士学位,主要从事医学图像分析处理、深度学习、模式识别等方面的研究。E-mail: zhoutaonxmu@126.com

通讯作者:



彭彩月(1999—),女,河南开封人,硕士研究生,主要从事智能医学影像图像处理,计算机辅助诊断方面等方面的研究。E-mail: peng_caiyue@163.com