

面向肺炎 CT 图像识别的 DL-CTNet 模型

王威, 黄文迪, 王新*, 王珑润

(长沙理工大学计算机与通信工程学院 长沙 410114)
(wangxin@csust.edu.cn)

摘要: 肺炎常缺乏明显呼吸系症状, 症状多不典型, 易发生漏诊、错诊。利用深度学习技术辅助医务人员安全、高效地检测感染者是一种有效途径。针对 COVID-19 感染者 CT 图像的磨玻璃影、铺路石征、血管扩张等特点, 提出一种可有效地提取 CT 图像中的局部与全局特征的轻量级模型——DL-CTNet。输入预处理的 CT 图像后, 首先采用空洞卷积和动态双路径多尺度特征融合(D-DMFF)模块的 2 个支路提取浅层特征; 然后使用局部与全局特征拼接模块(LGFC)中的 D-DMFF 模块提取局部特征、Swin Transformer 提取全局特征, 并通过拼接获得深层特征; 最后经过全连接层输出分类标签。实验结果表明, 在 2 个 CT 图像数据集上, 验证了 LGFC 模块以及 DL-CTNet 的低复杂度与有效性; DL-CTNet 的分类准确率高达 98.613%, 与其他方法相比, 其能更准确地识别肺炎的 CT 图像。

关键词: 肺炎; 胸部 CT 图像; 卷积神经网络; Transformer

中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2024.19792

DL-CTNet Model for CT Images Recognition of Pneumonia

Wang Wei, Huang Wendi, Wang Xin*, and Wang Longrun

(School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114)

Abstract: Pneumonia often lacks obvious respiratory symptoms, symptoms are more atypical, easy to occur missed diagnosis, misdiagnosis. Deep learning methods can assist medical personnel in diagnosing COVID-19 safely and efficiently. According to the characteristics of ground-glass opacity, crazy paving sign and vasodilatation in CT images of COVID-19 patients, a light weight network DL-CTNet is designed to effectively extract the local and global features in CT images. Firstly, after inputting the pre-processed CT images, the shallow features are extracted by two branches of cavity convolution and dynamic dual-path multi-scale feature fusion(D-DMFF). Then, the D-DMFF module in local and global feature concatenation module (LGFC) is used to extract local features, and Swin Transformer is used to extract global features, and the deep features are obtained by the above two branches in LGFC. Finally, the classification label is output by the fully connected layers. The low complexity and effectiveness of CTNet are verified on two CT image datasets. The experimental results show that the classification accuracy of DL-CTNet is as high as 98.613%, and compared with other methods, DL-CTNet can more accurately detect pneumonia CT images.

Key words: pneumonia; chest CT image; convolutional neural network; Transformer

收稿日期: 2022-06-16; 修回日期: 2022-10-27. 基金项目: 国防科技创新特区项目(2019XXX00701); 湖南省重点研究开发项目(2020SK2134); 湖南省自然科学基金(2022JJ30625). 王威(1974—), 男, 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为计算机视觉、模式识别; 黄文迪(1998—), 男, 硕士研究生, 主要研究方向为计算机视觉、模式识别; 王新(1976—), 女, 硕士, 讲师, 论文通信作者, 主要研究方向为计算机视觉、模式识别; 王珑润(2000—), 男, 硕士研究生, 主要研究方向为计算机视觉、模式识别。

2019年末爆发的新型冠状病毒(2019-nCoV)已遍布全球,其引发的肺炎(coronavirus disease 2019, COVID-19)已严重影响了世界经济的发展以及人民的生活.诊断COVID-19最常用的方法是基于逆转录酶聚合酶链反应(reverse transcription-polymerase chain reaction, RT-PCR)^[1]的检测方法. RT-PCR检测方法具有较高的特异性,但该方法需要消耗许多一次性的试剂资源^[2],且检测所需的时间较长.在肺部影像检测方面,临床广泛使用的医学成像有计算机断层扫描(computed tomography, CT),X射线以及核磁共振成像等. COVID-19包含了一些胸部CT图像可以识别的特征^[3-4].随着疫情防控日益常态化,以传统的人工阅片来检测CT图像中的病灶特征会增大影像科医生的压力.因此,应该寻求一种快速可靠、成本较低的自动检测方法来辅助医务人员诊断COVID-19.

深度学习能够让机器在文字、图像、声音中提取并分析特征,可以解决许多复杂的模式识别难题.卷积神经网络(convolutional neural network, CNN)是深度学习中极具代表性的分支,长期以来,计算机视觉建模一直由CNN主导^[5],其能够出色地完成医学影像识别任务^[6-8].Wang等^[6]采用全局平均池化层改进VGGNet^[9]和ResNet^[10],将改进后的轻量级模型应用到结肠息肉图像识别任务中;Rahimzadeh等^[7]设计一种基于Xception^[11]和ResNet50V2^[12]的模型,该模型可融合2个模型提取的特征,最终在COVID-19的胸部X射线图像分类任务中取得了92.85%的准确率;Wang等^[8]从DenseNet^[13]和MobileNet^[14]中得到启发,提出一种新颖的轻量级模型Dense-MobileNet,在儿童结肠息肉数据集中取得了较好的表现.

除CNN外,计算机视觉领域还流行着另一种深度学习架构——Transformer^[15],该架构最早应用于自然语言处理领域.Dosovitskiy等^[16]将Transformer应用到视觉领域并提出视觉转换器(vision Transformer, ViT),利用其自注意力机制关联起图像中的每个像素,实现全局特征的提取.许多学者将Transformer架构的模型应用于医学影像识别,获得了不错的效果.Gheflati等^[17]首次使用ViT到乳房超声图像分类任务中,并应用不同的数据增强策略训练模型;Chen等^[18]提出一种融合Transformer与CNN这2种架构的模型ViT-V-Net,应用于3D医学影像配准;Jia等^[19]基于CNN与Transformer提出一种多模态特征提取模型BiTr-Unet,将其应用于脑肿瘤分割取得了良好的分割效果.

近年来,许多学者致力于寻找从CT图像检测出COVID-19的深度学习方法.Sohan等^[20]收集了一系列COVID-19的CT图像数据集,使用一种22层的CNN模型对数据集进行大规模评估;Cohen等^[21]详细分析COVID-19在CT图像中的病灶特征,使用ResNet^[10],InceptionV3^[22]和Xception^[11]这3种模型识别COVID-19的CT图像,获得了较好的识别效果;Alshazly等^[23]提出CovidResNet和CovidDenseNet这2种CNN模型,它们支持不同架构之间的迁移学习,有助于提高对COVID-19的CT图像的诊断性能;Fan等^[24]提出一种基于Transformer模块和CNN模块的并行双分支模型Trans-CNNNet,实验结果表明,该模型在COVIDxCT数据集^[25]上的表现明显优于典型的CNN和Transformer模型,如ResNet^[10],DeiT^[26]等.

1 CNN与Swin Transformer

长期以来,CNN模型一直是解决计算机视觉问题的首选架构^[5].随着CNN的应用领域日益宽广,研究人员对CNN架构的思考也日益加深,该架构以更复杂的连接^[10,13,27]和更多样的卷积形式^[14]出现在视觉任务中.为了提高CNN模型的参数利用率并增强模型的特征提取能力,Szegedy等^[27]提出一种具有多支路的Inception模块,并应用到所提出的GoogLeNet中;为了解决CNN的退化问题,He等^[10]在模型设计中运用一种巧妙的跳跃连接,提出易于优化的ResNet模型;Huang等^[13]提出基于密集连接的DenseNet模型,通过特征重用减少参数量与计算量,能有效地缓解梯度消失的问题.

然而,由于CNN的感受野有限,因此不擅长提取特征之间长距离依赖关系.近年来,Transformer架构成功地被应用到计算机视觉领域,这种基于自注意力机制的架构在图像分类任务中更是取得了与CNN相当的结果^[16].与传统的CNN不同,Transformer的自注意力机制可以实现图像的全局特征提取,因此,越来越多的研究人员致力于设计出基于Transformer的模型,以挑战更广泛、更复杂的计算机视觉任务.Liu等^[5]深入探讨Transformer的适用性,提出一种应用于计算机视觉领域的通用主干网络,将其命名为滑动窗口转换器(shifted windows Transformer, Swin Transformer).Swin Transformer在编码器中引入窗口多头自注意力机制(windows multi-head self-attention, W-MSA)与滑动W-MSA(shifted W-MSA, SW-MSA),增强

了模型的局部特征提取能力,更灵活地完成多尺度的特征提取.此外,Swin Transformer 的计算复杂度与输入图像的尺寸线性相关,因此与 ViT^[16]相比更为轻量.

2 DL-CTNet 模型

COVID-19 在肺部 CT 影像中的医学特征为磨玻璃影,通常出现在肺腔周围;病情加重时,还会出现铺路石征;以及出现局部血管扩张等现象^[24].图 1 所示为 COVID-19 的部分医学特征,其中,各种典型的病灶特征已用橘红色标记.

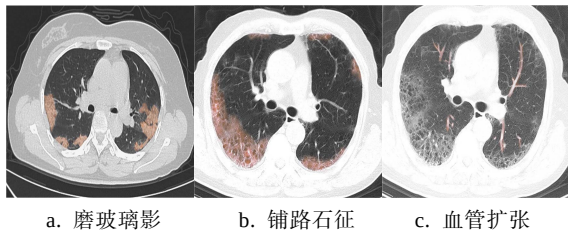


图 1 COVID-19 的 CT 图像医学特征

从特征分布的角度看,COVID-19 的 CT 图像包含许多局部特征,如铺路石征和局部血管扩张等;该图像还包含一些全局特征,如大面积的磨玻璃影.因此,基于 CT 影像的 COVID-19 分类任务非常考验模型的局部和全局特征提取能力.

针对以上问题,本文结合现有的 CNN 与 Transformer 架构提出 2 种模块:动态双路径多尺度特征融合(dynamic dual-path multi-scale feature fusion, D-DMFF)模块,局部与全局特征拼接(local and global feature concatenation, LGFC)模块;在此基础上,提出一种能够融合局部和全局特征的轻量级模型 DL-CTNet,其结构如图 2 所示.其中,预处理后的 CT 图像首先经过 DL-CTNet 的第 1 层,即空洞率为 2 的空洞卷积,以增加局部感受野、丰富特征信息;然后经过 2 个 D-DMFF 模块,以捕捉轮廓信息和纹理信息;再经过 3 个连续的 LGFC 模块,融合特征图中的局部信息与全局信息;经过一层自适应池化层压缩空间特征,对特征图进行展平处理;最后经过 2 个全连接层,整合类间区分信息,利用分类器输出分类标签.

2.1 D-DMFF 模块

由于在不同的 CT 图像样本中,COVID-19 局部病灶特征存在较大的差异,因此非常考验分类模型的局部表征能力.本课题组在之前的工作中^[28]提

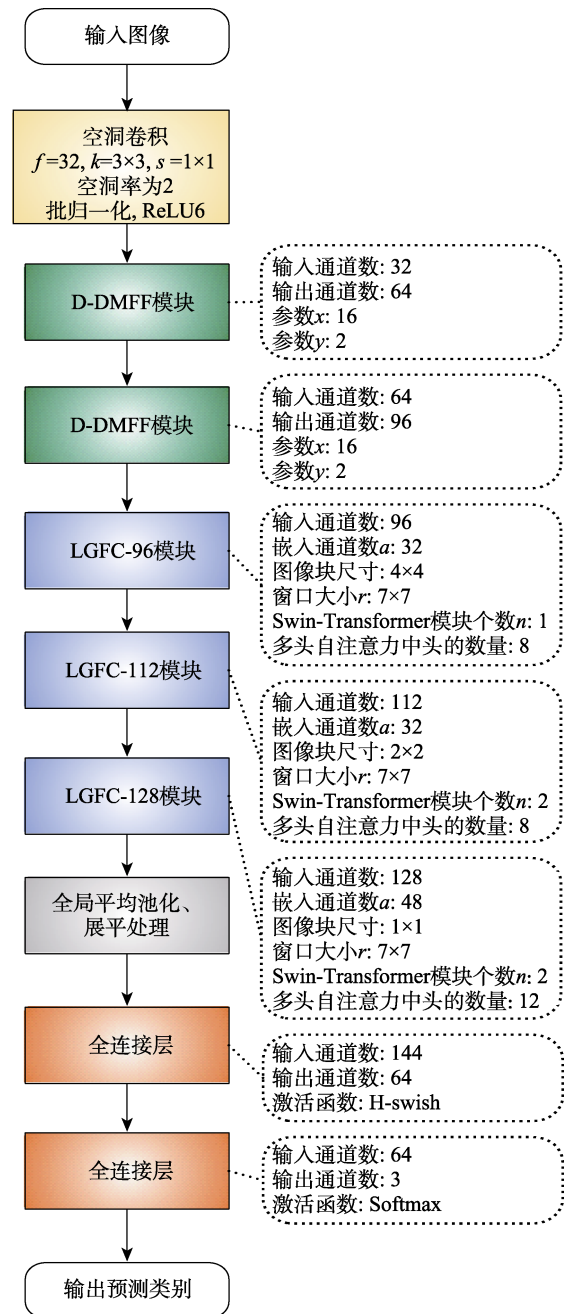


图 2 DL-CTNet 结构

出一种 DMFF 模块,并在 COVID 患者的胸部 X 射线图像识别任务中验证了该模块的表征能力. DMFF 模块是一种轻量高效的卷积模块,包含 2 条代表着不同感受野的深度可分离卷积^[14]分支,具有较强的局部信息提取能力,其出色的表征能力得益于多尺度空间特征提取模式.为了使 DMFF 模块可以更灵活地改变特征图维度,本文提出 D-DMFF 模块,其结构如图 3 所示.其中, h, w 分别表示输入特征图的高度和宽度; c 表示特征图的通道数; f, k, s 分别表示卷积核的个数、尺寸和步长.与 DMFF 模块不同, D-DMFF 中包含 2 个需

要人工预设的参数 x 与 y , 调整参数 x 可改变模块通道数提升幅度, 以控制模块的宽度; 调整参数 y 可设置模块中深度卷积的步距, 以调节下采样。

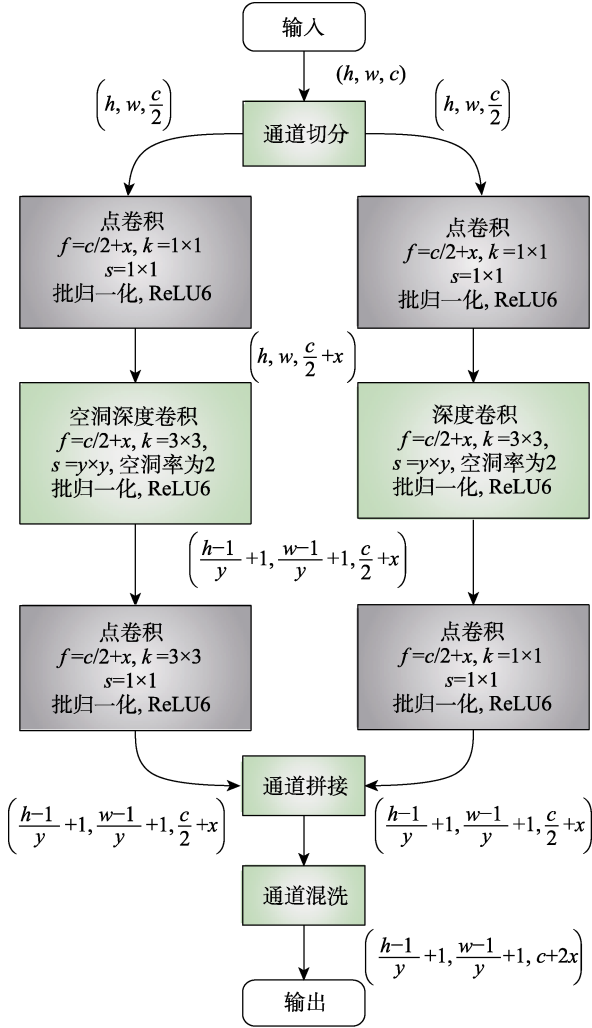


图 3 D-DMFF 模块结构

D-DMFF 模块包含 2 个支路, 分别使用空洞深度卷积层与深度卷积层提取局部特征, 并采用通道拼接的方法将 2 条支路提取到的特征进行融合。在模块内的深度卷积层和空洞卷积层设置前、后各一层点卷积层, 前者旨在提升特征图通道维度, 丰富特征信息; 后者旨在融合通道间的信息。在模块输出特征图前进行一次通道混洗, 使通道充分融合。本文中, D-DMFF 模块将作为 LGFC 模块与 DL-CTNet 的子模块。

2.2 LGFC 模块

为了提高自动识别模型的局部和全局特征提取能力, 控制模型的参数量和计算量, 本文提出一种基于 D-DMFF 模块与 Swin Transformer 模块的并行结构模块, 即 LGFC 模块, 其结构如图 4 所示。

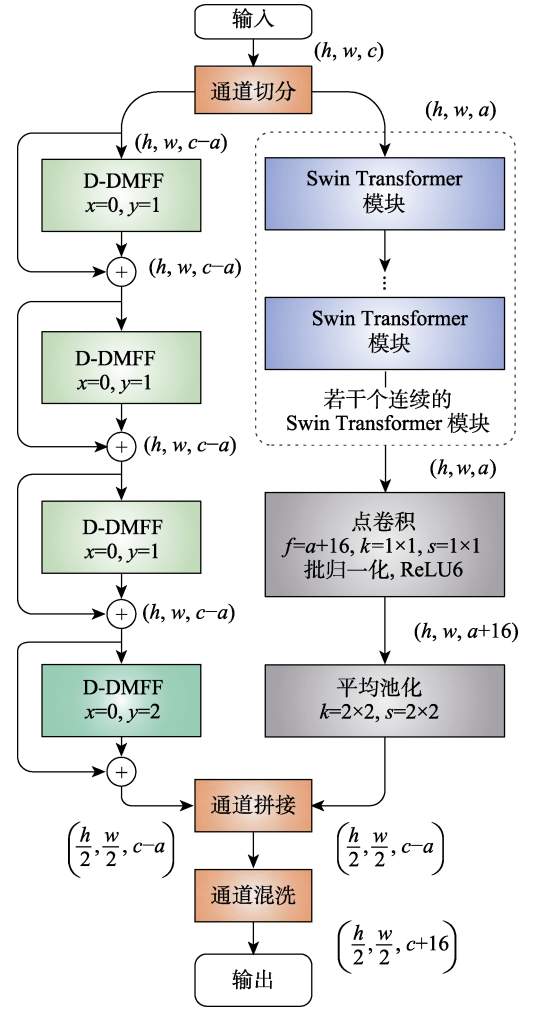


图 4 LGFC 模块结构

(1) LGFC 模块将输入的特征图在通道域上按一定比例进行切分, 得到 2 组通道数不同、空间维度相同的特征图。通往支路 1 的特征图的大小为 $h \times w \times a$, 首先经过若干个连续的 Swin Transformer 模块, 利用窗口自注意力机制完成全局特征提取; 然后经过一个点卷积层丰富特征信息; 最后经过一个池化核大小为 2、步长为 2 的平均池化层压缩空间特征, 提高特征提取效率。通往支路 2 的特征图的形状为 $h \times w \times (c-a)$, 经过 4 个引入跳跃连接的 D-DMFF 模块, 以残差学习完成局部特征提取并压缩空间信息。(2) 将上述 2 组特征提取后的特征图进行通道拼接, 得到形状为 $\frac{h}{2} \times \frac{w}{2} \times (c+16)$ 的特征图。(3) 在 LGFC 模块的输出之前, 对特征图进行通道混洗, 保持特征图内各通道的信息交流。

2.3 模型复杂度分析

设输入特征图的大小为 $h_i \times w_i \times d_i$, 卷积核大

小为 $d_i \times k^2 \times d_o$, 输出特征图的大小 $h_o \times w_o \times d_o$, 记此特征提取过程为变换 1. 若采用标准卷积完成变换 1, 则产生的参数量 P_{std} 和计算量 F_{std} 分别为

$$P_{std} = d_i \times k^2 \times d_o \quad (1)$$

$$F_{std} = h_o \times w_o \times d_i \times k^2 \times d_o \quad (2)$$

若采用深度卷积或空洞率为 2 的空洞深度卷积完成变换 1, 则产生的参数量 P_{dw} 和计算量 F_{dw} 分别为

$$P_{dw} = d_i \times k^2 \quad (3)$$

$$F_{dw} = h_o \times w_o \times d_i \times k^2 \quad (4)$$

设输入特征图和输出特征图的大小均为 $h \times w \times d$, 记此特征提取过程为变换 2. 若采用 Swin Transformer 模块完成变换 2, 其中, W-MSA 中的窗口大小为 r , 多层感知机(multi-layer perceptron, MLP)中全连接层的内置维度为 $4 \times d$, 则产生的参数量 P_{swin} 和计算量 F_{swin} 分别为

$$P_{swin} = 12 \times d^2 + 2 \times r^2 \times d \quad (5)$$

$$F_{swin} = 12 \times h \times w \times d^2 + 2 \times h \times w \times r^2 \times d \quad (6)$$

设输入特征图的大小为 $h \times w \times d$, 输出特征图的大小为 $\frac{h}{2} \times \frac{w}{2} \times (d+16)$, 记此特征提取过程为变换 3. 若采用 LGFC 模块完成变换 3, 设支路 1 中的通道数为 a , Swin Transformer 模块的个数为 n , 根据式(3)~式(6)可知, 其支路 1 产生的参数量 P_{LGFC_1} 、计算量 F_{LGFC_1} , 以及其支路 2 产生的参数量 P_{LGFC_2} 、计算量 F_{LGFC_2} 分别为

$$P_{LGFC_1} = n \times (12 \times a^2 + 2 \times r^2 \times a) + a^2 + 16 \times a \quad (7)$$

$$F_{LGFC_1} = n \times (12 \times h \times w \times a^2 + 2 \times h \times w \times r^2 \times a) + h \times w \times a \times (a+16) \quad (8)$$

$$P_{LGFC_2} = 4 \times (d-a)^2 + 36 \times (d-a) \quad (9)$$

$$F_{LGFC_2} = 4 \times h \times w \times (d-a)^2 + \frac{117}{4} \times h \times w \times (d-a) \quad (10)$$

由式(7)~式(10)可知, LGFC 模块完成变换 3 所产生的参数量 P_{LGFC} 和计算量 F_{LGFC} 分别为

$$P_{LGFC} = (12 \times n + 5) \times a^2 + (2 \times n \times r^2 - 20) \times a - 8 \times a \times d + 4 \times d^2 + 36 \times d \quad (11)$$

$$F_{LGFC} = \left(2 \times n \times r^2 - 8 \times d - \frac{53}{4} \right) \times h \times w \times a + (12 \times n + 5) \times h \times w \times a^2 + 4 \times h \times w \times d^2 + \frac{117}{4} \times h \times w \times d \quad (12)$$

若采用 4 层 3×3 标准卷积完成变换 3, 设此 4

层标准卷积层输出通道均为 $(d+16)$, 第 1 层卷积核步长为 2, 其余 3 层步长为 1, 由式(1)(2)可知, 其产生的参数量 P_{conv3_4} 和计算量 F_{conv3_4} 分别为

$$P_{conv3_4} = 36 \times d^2 + 1008 \times d + 6912 \quad (13)$$

$$F_{conv3_4} = (9 \times d^2 + 252 \times d + 1728) \times h \times w \quad (14)$$

根据式(11)~式(14)可知, 相对于采用 4 层 3×3 标准卷积层, 采用 LGFC 模块完成变换 3 的参数减少量 ΔP_{LGFC} (即 P_{conv3_4} 与 P_{LGFC} 的差值), 以及计算减少量 ΔF_{LGFC} (即 F_{conv3_4} 与 F_{LGFC} 的差值)分别为

$$\Delta P_{LGFC} = 32 \times d^2 + 972 \times d - (12 \times n + 5) \times a^2 -$$

$$(2 \times n \times r^2 - 20) \times a + 8 \times a \times d + 6912,$$

$$\Delta F_{LGFC} = \left(5 \times d^2 + \frac{891}{4} \times d + 1728 \right) \times h \times w -$$

$$\left(2 \times n \times r^2 - 8 \times d - \frac{53}{4} \right) \times h \times w \times a -$$

$$(12 \times n + 5) \times h \times w \times a^2.$$

模块所用的滑动窗口大小 r 均为 7, 且每个 LGFC 模块中的 h, w, d, n, a 均为已知变量, 因此可求得各模块的参数减少量 ΔP_{LGFC} 与计算减少量 ΔF_{LGFC} , 计算结果如表 1 所示. 可以看出, DL-CTNet 中的 3 个 LGFC 模块对应的 ΔP_{LGFC} 和 ΔF_{LGFC} 均大于 0, 说明 LGFC 模块可有效地降低模型的复杂度.

表 1 各 LGFC 模块的参数减少量与计算减少量对比

模块	ΔP_{LGFC}	ΔF_{LGFC}
LGFC-96	0.40M	230.96
LGFC-112	0.51M	64.70
LGFC-128	0.63M	16.80

为了分析 DL-CTNet 模型的复杂度, 本文通过实验测试了该模型与一些深度学习模型在 CT 分类实验中的参数量与计算量, 结果如表 2 所示, 其中, 设输入图像尺寸为 224×224 像素. 可以看出, DL-CTNet 模型的复杂度远低于一些经典的深层 CNN 模型, 如 VGGNet19, ResNet50, DenseNet121 等, 也远低于目前较先进的计算机视觉模型, 如 Swin Transformer-Tiny, ConvNext-Tiny; 与轻量级模型相比, DL-CTNet 的计算量是 ShuffleNetV2 的 269%, MobileNetV2 的 127%, 而 ShuffleNetV2 和 MobileNetV2 的参数量分别是 DL-CTNet 的 605% 与 1 062%. 因此, DL-CTNet 模型具有较低的复杂度, 尤其是在参数量方面占有极大的优势.

表2 DL-CTNet 与一些深度学习模型的复杂度对比

模型	计算量	参数量
VGGNet19 ^[9]	19 646.97M	139.59M
ResNet50 ^[10]	4 119.88M	23.51M
DenseNet121 ^[13]	2 881.60M	6.62M
ShuffleNetV2 ^[29]	147.79M	1.26M
MobileNetV2 ^[30]	312.85M	2.23M
InceptionV3 ^[22]	5 729.24M	21.79M
Swin Transformer-Tiny ^[5]	4 495.90M	27.47M
ConvNext-Tiny ^[31]	4 466.01M	27.82M
DL-CTNet	398.19M	0.21M

3 实验结果分析

3.1 数据集

本文实验使用2个开源CT数据集:(1) 中国胸部CT图像调查联盟(China consortium of chest CT image investigation, CC-CCII)数据集^[32], 用于DL-CTNet模型的消融实验和有效性验证实验;(2) COVIDxCT数据集^[25], 用于DL-CTNet模型的有效性验证补充实验. 表3所示为上述2个数据集的CT图像样本分布情况.

表3 实验数据集样本分布情况

数据集	样本		样本总量
	类别	数量	
CC-CCII ^[32]	COVID-19	638	4 658
	普通肺炎	1 392	
	正常	2 628	
COVIDxCT ^[25]	COVID-19	94 548	194 922
	普通肺炎	40 291	
	正常	60 083	

从CC-CCII数据集中随机选取531幅COVID-19阳性图像、1 147幅普通肺炎图像和2 212幅正常图像作为训练集, 其余的107幅COVID-19阳性图像、245幅普通肺炎图像和416幅正常图像作为测试集. 本文按照Gunraj等^[25]提供的方案, 对COVIDxCT数据集的训练集和测试集进行划分, 得到的训练集包含88 530幅COVID-19阳性图像、32 896幅普通肺炎图像和47 838幅正常图像, 其余的6 018幅COVID-19阳性图像、7 395幅普通肺炎图像和12 245幅正常图像作为测试集. 实验数据集的部分示例如图5所示. 可以看出, COVID-19与普通肺炎的CT图像类间相似度较高.

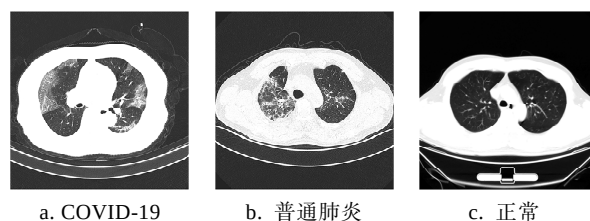


图5 实验数据集部分图像示例

本文采用准确率 A , 精度 P , 灵敏度 S_e , 特异度 S_p 和平衡分数 F_1 评估指标衡量模型的性能, 根据模型预测正确的正样本(true positive, TP), 预测错误的正样本(false negative, FN), 预测正确的负样本(true negative, TN)和预测错误的负样本(false positive, FP), 则5个评估指标的计算公式为

$$A = \frac{TP + TN}{TP + TN + FP + FN},$$

$$P = \frac{TP}{TP + FP},$$

$$S_e = \frac{TP}{TP + FN},$$

$$S_p = \frac{TN}{TN + FP},$$

$$F_1 = \frac{2 \times TP}{2 \times TP + FP + FN}.$$

3.2 预处理与参数设置

采取4步预处理对训练集进行数据增强:(1) 将CT图像在 $[-60^\circ, 60^\circ]$ 进行随机旋转;(2) 将分辨率固定为 256×256 像素;(3) 进行中心裁剪, 使其分辨率变为 224×224 像素;(4) 其图像对比度增强至原来的140%.

实验中, 大部分模型的初始学习率为0.001, 为提高收敛质量, 个别模型的初始学习率需作调整, 例如, InceptionV3的初始学习率设为0.005; VGGNet19, Swin Transformer-Tiny, ConvNext-Tiny, MobileNetV2, ShuffleNetV2等模型的初始学习率设为0.00001. 本文根据数据集样本数量调整模型的训练周期, 对于CC-CCII数据集, 每组模型遍历150次数据集; 而对于COVIDxCT数据集, 每组模型遍历30次数据集. 为加快模型收敛速度, 实验模型均采用Adam优化器. 实验平台为Ubuntu 20.04.4 LTS操作系统, Intel(R) Xeon(R) Silver 4214 CPU @ 2.20GHz, GeForce RTX 2080Ti, CUDA 11.1.74, cuDNN版本号为8.0.05, 模型搭建框架为Pytorch. 所有模型的训练与测试在同一配置环境下进行.

3.3 模型有效性验证与结果分析

DL-CTNet模型的全局特征提取能力得益于

LGFC 模块中的 Swin Transformer 模块, 而 Swin Transformer 模块的个数会影响模型的性能与复杂度. 为了确定 3 个 LGFC 模块支路 1 中 Swin Transformer 模块的个数, 本文共设计 9 个模型(模型 1~模型 9)作为对照组, 该组中的各个模型除 Swin Transformer 模块个数不同以外, 其他设置均与 DL-CTNet 相同. 记 LGFC-96, LGFC-112, LGFC-128 这 3 个模块的 Swin Transformer 模块个数分别为 x_1 , x_2 和 x_3 , 各模型的参数设置、复杂度以及在 CC-CCII 数据集上的准确率如表 4 所示.

表 4 Swin Transformer 模块实验结果对比

模型	x_1	x_2	x_3	计算量	参数量	准确率/%
1	1	1	1	368.33M	0.17M	92.057
2	1	1	2	374.67M	0.20M	91.016
3	1	2	1	380.42M	0.18M	91.536
DL-CTNet	1	2	2	398.19M	0.21M	93.880
4	1	2	3	399.12M	0.24M	92.708
5	1	2	4	400.03M	0.26M	91.927
6	2	1	1	416.70M	0.18M	93.480
7	2	1	2	423.04M	0.21M	91.536
8	2	2	1	428.79M	0.19M	92.708
9	2	2	2	435.13M	0.22M	91.146

从表 4 可以看出, 在复杂度方面, 10 个模型的参数量差异较为明显, 模型 5 的参数量是模型 1 的 153%; 在准确率方面, 10 个模型的准确率最大值与最小值的差距达 2.864%; DL-CTNet 的复杂度在对照组的 9 个模型位居中等, 但准确率高达 93.880%. 从性能与复杂度两方面综合考虑, 本文最终选用参数组 $\{x_1 = 1; x_2 = 2; x_3 = 2\}$.

与 Transformer 相比, CNN 具有更强的平移不变性与局部相关性. CNN 层数越多, 模型对像素空间的敏感性就越高, 但增加模型深度也会导致过拟合. 为了确定 3 个 LGFC 模块支路 2 中 D-DMFF 模块的个数, 本文共设计 5 个模型(模型 10~模型 14)作为对照组, 该组中的各个模型除了 D-DMFF 模块个数不同外, 其他设置均与 DL-CTNet 相同. 记 LGFC 模块的 D-DMFF 模块个数为 z , 各模型的 z 值以及在 CC-CCII 数据集上准确率如表 5 所示.

从表 5 可以看出, 随着 LGFC 模块中 D-DMFF 模块个数的增加, DL-CTNet 模型对 CT 图像的分类准确率呈上升趋势; 该组模型的性能差异明显, 准确率最大值与最小值的差距高达 3.255%; 当 $z = 4$ 时, 模型性能最优; 当 $z > 4$ 时, 随着 z 的增大, 模型的性能呈轻微下降趋势. 因此, 本文最终 LGFC

表 5 不同 D-DMFF 模块个数下各模型实验结果对比

模型	z	准确率/%
10	1	90.625
11	2	92.969
12	3	91.667
DL-CTNet	4	93.880
13	5	93.750
14	6	93.620

模块中的 D-DMFF 模块数量设置为 4.

为了说明 LGFC 模块各支路之间特征提取模式的差异, 本文从 CC-CCII 测试集中分别抽取 COVID-19 的 CT 图像样本以及普通肺炎的 CT 图像样本各一幅, 并制作样本在最后一个 LGFC 模块的类激活映射(class activation mapping, CAM)热力图, 如图 6 和图 7 所示.

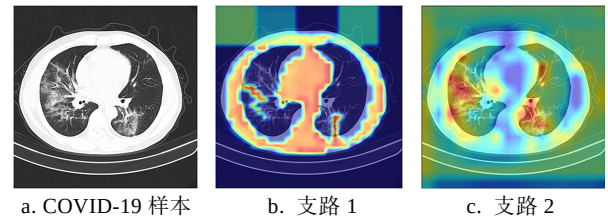


图 6 COVID-19 样本在最后一个 LGFC 模块中的 CAM 图

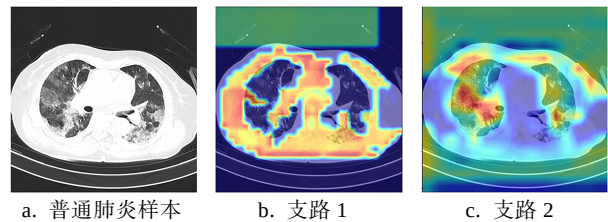


图 7 普通肺炎样本在最后一个 LGFC 模块中的 CAM 图

为了验证 LGFC 模块各支路的有效性, 本文设置一组消融实验, 其中, 保留 LGFC 模块支路 1 得到 LGFC1-CTNet, 保留 LGFC 模块支路 2 得到 LGFC2-CTNet, 并对比 LGFC1-CTNet, LGFC2-CTNet 和 DL-CTNet 在 CC-CCII 数据集上的分类性能. 该实验中的 2 个对比模型除了 LGFC 模块的结构不同外, 其他设置均与 DL-CTNet 相同. LGFC 模块消融实验结果如图 8 所示.

从图 8 可以看出, LGFC1-CTNet 和 LGFC2-CTNet 的分类准确率均低于 DL-CTNet, 说明与仅采用其中一种特征提取模式相比, 同时采用局部和全局特征提取模式更有利于提高模型的性能; 在训练的过程中, LGFC1-CTNet 的准确率波动幅度

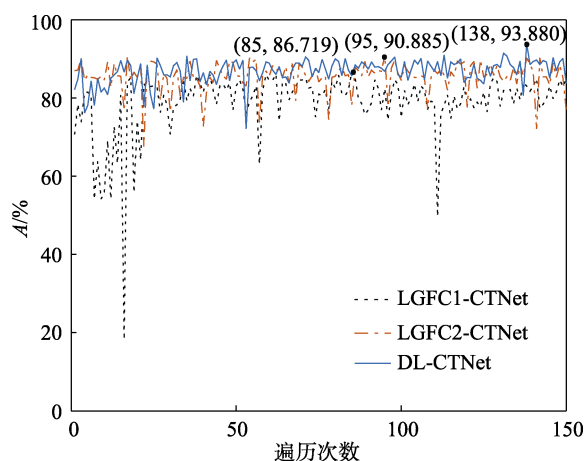


图 8 LGFC 模块消融实验结果

较大(尤其在训练阶段的前期), 相比之下, LGFC2-CTNet 和 DL-CTNet 的分类准确率曲线相对平稳; 由于 LGFC1-CTNet 的特征提取架构以 Transformer 为主, 与 CNN 相比, Transformer 的平移不变性较弱, 因此识别性能不稳定; 反观 LGFC2-CTNet 与 DL-CTNet, 其特征提取架构均以 CNN 为主, 平移不变性和局部相关性较强, 因此准确率曲线相对平稳。

为了说明 DL-CTNet 对 COVID-19 的 CT 图像识别的有效性, 本文使用 VGGNet19, ResNet50, DenseNet121 等经典 CNN, ShuffleNetV2, MobileNetV2 等轻量级模型, 以及 Swin Transformer, ConvNext 等目前较先进的模型在 CC-CCII 数据集上进行实验, 结果如表 6 所示。

表 6 不同模型在 CC-CCII 数据集上实验指标结果对比

%

模型	A	P	S_e	S_p	F_1
VGGNet19 ^[9]	91.016	91.625	86.401	97.455	91.193
ResNet50 ^[10]	91.276	91.737	87.024	97.046	91.376
DenseNet121 ^[13]	92.318	92.105	86.358	94.279	92.032
InceptionV3 ^[22]	90.625	90.494	81.080	95.344	89.959
MobileNetV2 ^[30]	91.927	91.717	85.774	96.712	91.799
ShuffleNetV2 ^[29]	92.448	92.258	86.318	97.062	92.315
Swin Transformer-Tiny ^[5]	93.099	93.349	85.595	96.317	92.690
ConvNext-Tiny ^[31]	88.542	87.877	77.499	95.617	87.860
DL-CTNet	93.880	93.764	88.341	97.679	93.750

注: 粗体表示最优值。

从表 6 可以看出, DL-CTNet 的各项指标均高于其他对比模型; 在 3 个经典 CNN 中, DenseNet121 的分类准确率较优, 但仍低于 DL-CTNet 约 1.562%, 且复杂度也远高于 DL-CTNet, 其参数量与计算量为 DL-CTNet 的 3 152%和 724%; 在 2 个轻量级 CNN 中, ShuffleNetV2 的分类准确率较优, 但仍低于 DL-CTNet 约 1.432%, 且在复杂度方面, 虽然 ShuffleNetV2 的计算量是 DL-CTNet 的 37%, 但是其参数量为 DL-CTNet 的 605%; 在 2 个目前较先进的模型中, Swin Transformer-Tiny 的分类准确率较高, 但仍比 DL-CTNet 低 0.781%, 且其参数量和计算量为 DL-CTNet 的 13 081%和 1 129%; DL-CTNet 的隐藏层设计借鉴了 ShuffleNetV2 的基本单元, 采用通道切分—特征提取—通道拼接的模型结构, 因此有效实现轻量化; 在特征提取方面, DL-CTNet 采用 D-DMFF 模块和 Swin Transformer 模块作为特征提取模块, 使得 DL-CTNet 在特征提取过程中可兼顾局部特征和全局特征, 因此 DL-CTNet 的识别性能比 ShuffleNetV2, Swin Transformer-Tiny 等模型优越。

本文收集了一些自动识别 COVID-19 的 CT 图像的深度学习方法, 并对这些方法与 DL-CTNet 的样本分布进行实验, 结果如表 7 所示。可以看出, 在 4 种对比方法中, Alshazly 等^[23]提出的 CovidDenseNet

表 7 5 种方法样本分布实验结果对比

方法	样本分布		准确率/%	参数量
	类别	数量		
CovidResNet ^[23]	COVID-19	2 168	82.46	9.84M
	普通肺炎	1 247		
	正常	758		
CovidDenseNet ^[23]	COVID-19	2 168	82.87	1.44M
	普通肺炎	1 247		
	正常	758		
Sohan 等 ^[20]	COVID-19	1 230	75.00	6.98M
	正常	1 230		
Cohen 等 ^[21]	COVID-19	1 230	68.54	21.79M
	正常	1 230		
本文	COVID-19	638	93.88	0.21M
	普通肺炎	1 392		
	正常	2 628		

参数量最少,但仍为 DL-CTNet 的 686%; DL-CTNet 识别准确率最高,且采用的样本总量最充足.由于表 7 中各模型的准确率数据是在不同数据样本分布、不同实验环境下得到的,为了更直观、更严谨地对比以上识别方法的效果,更深入地验证 DL-CTNet 的有

效性,本文测试了以上的各种方法在 COVIDxCT 数据集的表现. COVIDxCT 数据集的 CT 图像样本相对充足,有利于降低样本不足造成的误差.实验中,各个模型均在同一环境与参数设置下训练,各模型的性能指标以及混淆矩阵分别如表 8 和图 9 所示.

表 8 5 种方法在 COVIDxCT 数据集上的实验结果性能指标对比

方法	A	P	S_e	S_p	F_1
CovidResNet ^[23]	98.079	98.104	97.905	99.130	98.079
CovidDenseNet ^[23]	98.160	98.181	97.990	99.178	98.166
Sohan 等 ^[20]	94.376	94.694	94.431	97.489	94.451
Cohen 等 ^[21]	98.273	98.294	98.364	99.239	98.278
本文	98.613	98.614	98.507	99.261	98.612

注:粗体表示最优值.

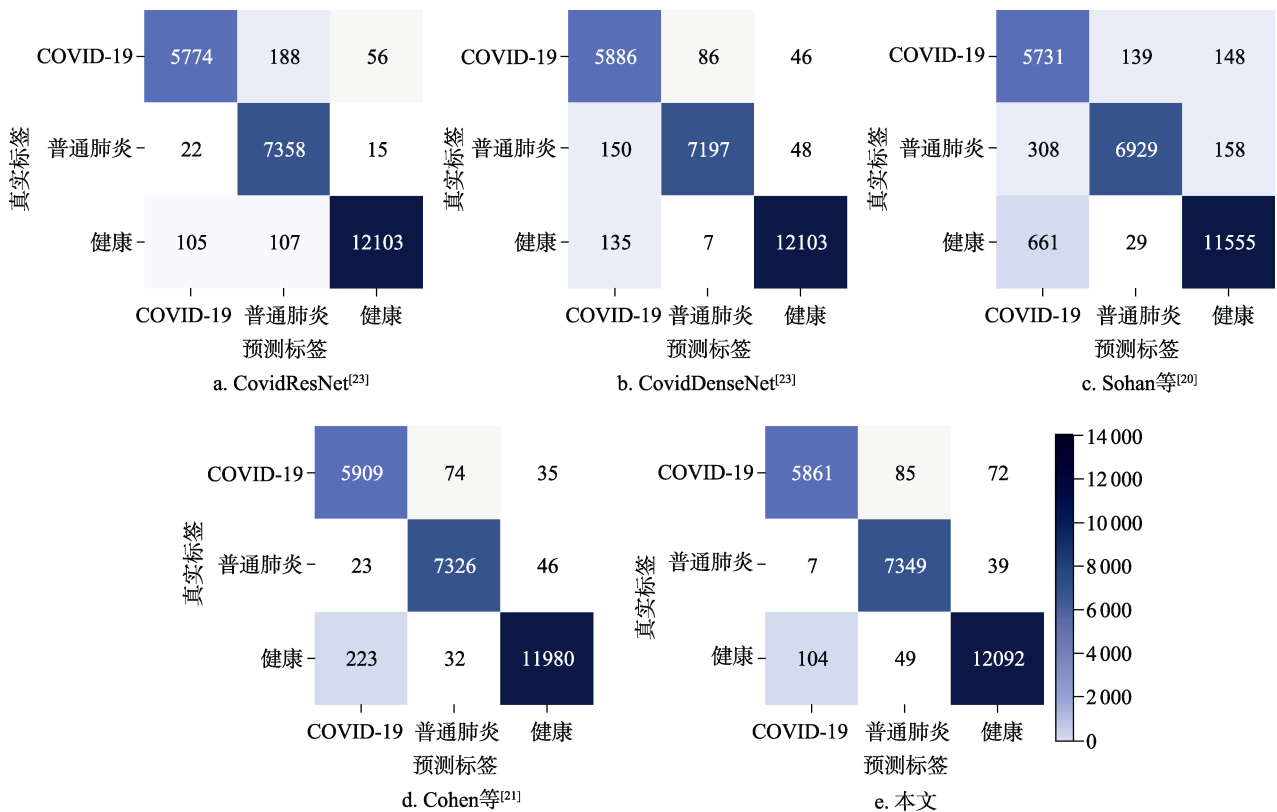


图 9 DL-CTNet 与 4 种深度学习识别方法在 COVIDxCT 数据集上混淆矩阵结果

从表 8 可以看出,本文方法的各项性能指标均优于其他方法;4 种对比方法中, Cohen 等^[21]的方法分类效果最佳,所用模型 InceptionV3 可实现基于 CNN 的多尺度特征提取,其分类准确率高达 98.273%,但仍比本文方法低 0.340%,且其参数量大约是本文方法的 104 倍.观察图 9d 和图 9e 可进一步对比 2 种方法的性能, Cohen 等的方法对于 COVID-19 漏诊 109 次,误诊 256 次,共出错 365 次;而本文方法对于 COVID-19 漏诊 157 次,误诊 111

次,共出错 268 次.因此,本文提出的 DL-CTNet 模型能更轻量高效、更准确地识别 COVID-19 的 CT 图像.

4 结 语

基于对 CNN 与 Transformer 架构的特性分析,针对 CT 图像的特点,本文提出 D-DMFF 模块以及可实现局部特征与全局特征提取的 LGFC 模块;

在此基础上, 提出一种基于 CNN 与 Transformer 架构的轻量级模型 DL-CTNet. 在 2 个 CT 图像数据集上验证了 DL-CTNet 的有效性. 实验结果表明, DL-CTNet 的识别准确率、精度、灵敏度、特异度、 F_1 值等指标均高于其他模型, 且具有更低的计算复杂度. 虽然 DL-CTNet 在实验中表现出色, 但仍需要更深入的临床测试. 后续将努力实现模型的可解释性, 从而更好地辅助医务人员诊断肺炎.

参考文献(References):

- [1] Xu X W, Jiang X G, Ma C L, *et al.* A deep learning system to screen novel coronavirus disease 2019 pneumonia[J]. *Engineering*, 2020, 6(10): 1122-1129
- [2] Xie X Z, Zhong Z, Zhao W, *et al.* Chest CT for typical coronavirus disease 2019 (COVID-19) Pneumonia: Relationship to Negative RT-PCR Testing[J]. *Radiology*, 2020, 296(2): E41-E45
- [3] Ye Z, Zhang Y, Wang Y, *et al.* Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review[J]. *European Radiology*, 2020, 30(8): 4381-4389
- [4] Rodriguez-Morales A J, Cardona-Ospina J A, Gutiérrez-Ocampo E, *et al.* Clinical, laboratory and imaging features of COVID-19: a systematic review and meta-analysis[J]. *Travel Medicine and Infectious Disease*, 2020, 34: Article No.101623
- [5] Liu Z, Lin Y T, Cao Y, *et al.* Swin transformer: hierarchical vision transformer using shifted windows[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 9992-10002
- [6] Wang W, Tian J G, Zhang C W, *et al.* An improved deep learning approach and its applications on colonic polyp images detection[J]. *BMC Medical Imaging*, 2020, 20(1): Article No.83
- [7] Rahimzadeh M, Attar A. A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2[J]. *Informatics in Medicine Unlocked*, 2020, 19: Article No.100360
- [8] Wang W, Li Y T, Zou T, *et al.* A novel image classification approach via dense-MobileNet models[J]. *Mobile Information Systems*, 2020, 2020: Article No.7602384
- [9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[OL]. [2022-06-16]. <https://arxiv.org/abs/1409.1556>
- [10] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C] //Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 770-778
- [11] Chollet F. Xception: deep learning with depthwise separable convolutions[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 1800-1807
- [12] He K M, Zhang X Y, Ren S Q, *et al.* Identity mappings in deep residual networks[C] //Proceedings of the 14th European Conference on Computer Vision. Heidelberg: Springer, 2016: 630-645
- [13] 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. Los Alamitos: IEEE Computer Society Press, 2017: 2261-2269
- [14] Howard A G, Zhu M L, Chen B, *et al.* MobileNets: efficient convolutional neural networks for mobile vision applications[OL]. [2022-06-16]. <https://arxiv.org/abs/1704.04861>
- [15] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need[C] //Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM Press, 2017: 6000-6010
- [16] Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: transformers for image recognition at scale[OL]. [2022-06-16]. <https://arxiv.org/abs/2010.11929>
- [17] Gheflati B, Rivaz H. Vision transformers for classification of breast ultrasound images[C] //Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. Los Alamitos: IEEE Computer Society Press, 2022: 480-483
- [18] Chen J Y, He Y F, Frey E C, *et al.* ViT-V-Net: vision transformer for unsupervised volumetric medical image registration[OL]. [2022-06-16]. <https://arxiv.org/abs/2104.06468>
- [19] Jia Q R, Shu H. BiTr-Unet: a CNN-transformer combined network for MRI brain tumor segmentation[C] //Proceedings of the 7th International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Heidelberg: Springer, 2022: 3-14
- [20] Sohan M F, Basalamah A, Solaiman M. COVID-19 detection using machine learning: a large scale assessment of x-ray and CT image datasets[J]. *Journal of Electronic Imaging*, 2022, 31(4): Article No.041212
- [21] Cohen M W, Gilo O, David L. A computer aided medical classification system of COVID-19 CT lung scans using convolution neural networks[J]. *Computer-Aided Design & Applications*, 2022, 19(3): 522-533
- [22] Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the Inception architecture for computer vision[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 2818-2826
- [23] Alshazly H, Linse C, Abdalla M, *et al.* COVID-Nets: deep CNN architectures for detecting COVID-19 using chest CT scans[J]. *PeerJ Computer Science*, 2021, 7: Article No.e655
- [24] Fan X L, Feng X F, Dong Y Y, *et al.* COVID-19 CT image recognition algorithm based on transformer and CNN[J]. *Displays*, 2022, 72: Article No.102150
- [25] Gunraj H, Wang L D, Wong A. COVIDNet-CT: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest CT images[J]. *Frontiers in Medicine*, 2020, 7: Article No.608525
- [26] Touvron H, Cord M, Douze M, *et al.* Training data-efficient image transformers & distillation through attention[C] //Proceedings of the 38th International Conference on Machine Learning. Lille: PMLR Press, 2021, 139: 10347-10357
- [27] Szegedy C, Liu W, Jia Y Q, *et al.* Going deeper with convolutions[C] //Proceedings of the IEEE Conference on Computer

- Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 1-9
- [28] Wang Wei, Hu Yiyang, Wang Xin, *et al.* DD-CovidNet model for X-ray images recognition of coronavirus disease 2019[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(11): 1649-1657(in Chinese)
(王威, 胡亿洋, 王新, 等. 针对新型冠状病毒肺炎 X 射线图像识别的 DD-CovidNet 模型[J]. 计算机辅助设计与图形学学报, 2021, 33(11): 1649-1657)
- [29] Ma N N, Zhang X Y, Zheng H T, *et al.* ShuffleNet V2: practical guidelines for efficient CNN architecture design[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 122-138
- [30] Sandler M, Howard A, Zhu M L, *et al.* MobileNetV2: inverted residuals and linear bottlenecks[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 4510-4520
- [31] Liu Z, Mao H Z, Wu C Y, *et al.* A ConvNet for the 2020s[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 11966-11976
- [32] Zhang K, Liu X H, Shen J, *et al.* Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography[J]. Cell, 2020, 181(6): 1423-1433.e11