

## 基于轻量级RG-DenseNet的COVID-19 CT图像分类

张子宇<sup>1</sup>, 赵可辉<sup>2</sup>, 牛慧芳<sup>3</sup>, 张志强<sup>1</sup>, 周连田<sup>4</sup>

1. 山东中医药大学智能与信息工程学院, 山东 济南 250000; 2. 山东中医药大学第二附属医院特检科, 山东 济南 250000; 3. 山东省药品不良反应检测中心, 山东 济南 250000; 4. 菏泽市中医医院碎石科, 山东 菏泽 247000

**【摘要】目的:** 基于轻量级RG-DenseNet构建COVID-19 CT图像分类模型。**方法:** 以DenseNet121为基础, 添加通道和空间注意力机制模块减少无关特征的干扰, 将DenseNet中的Bottleneck模块替换为前激活的RG-beneck2模块减少模型参数的同时保持精度尽可能不变。构建RG-DenseNet模型, 在COVIDx CT-2A数据集上进行3分类实验。**结果:** RG-DenseNet准确率为98.93%、精确率为98.70%、召回率为98.97%、特异性为99.48%、F1分数为98.83%。**结论:** RG-DenseNet与原模型DenseNet121相比在保持准确度仅降低0.01%的情况下, 减少92.7%的参数量和计算量, 轻量化效果显著, 具有实际应用价值。

**【关键词】** RepGhost; DenseNet; COVID-19; 深度学习; 图像分类

**【中图分类号】** R318; TP391

**【文献标志码】** A

**【文章编号】** 1005-202X(2023)12-1494-08

### COVID-19 classification on CT image using lightweight RG DenseNet

ZHANG Ziyu<sup>1</sup>, ZHAO Kehui<sup>2</sup>, NIU Huifang<sup>3</sup>, ZHANG Zhiqiang<sup>1</sup>, ZHOU Liantian<sup>4</sup>

1. College of Intelligence and Information Engineering, Shandong University of Traditional Chinese Medicine, Jinan 250000, China; 2. Special Inspection Department, the Second Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan 250000, China; 3. Shandong Province Adverse Drug Reaction Testing Center, Jinan 250000, China; 4. Department of Lithotripsy, Heze Traditional Chinese Medicine Hospital, Heze 247000, China

**Abstract: Objective** To construct a COVID-19 CT image classification model based on lightweight RG DenseNet. **Methods** A RG-DenseNet model was constructed by adding channel and spatial attention modules to DenseNet121 for minimizing the interference of irrelevant features, and replacing Bottleneck module in DenseNet with pre-activated RG beneck2 module for reducing model parameters while maintaining accuracy as much as possible. The model performance was verified with 3-category classification experiments on the COVIDx CT-2A dataset. **Results** RG-DenseNet had an accuracy, precision, recall rate, specificity, and F1-score of 98.93%, 98.70%, 98.97%, 99.48%, and 98.83%, respectively. **Conclusion** Compared with the original model DenseNet121, RG-DenseNet reduces the number of parameters and the computational complexity by 92.7%, while maintaining an accuracy reduction of only 0.01%, demonstrating a significant lightweight effect and high practical application value.

**Keywords:** RepGhost; DenseNet; COVID-19; deep learning; image classification

### 前言

新冠肺炎(COVID-19)由新型冠状病毒(SARS-CoV-2)感染引起<sup>[1]</sup>。据世界卫生组织官网的数据显示,截至

2023年7月24日,全球COVID-19累计确诊病例达768 237 024例,累计死亡病例6 951 664例。COVID-19病患的常见症状是发热、咳嗽、呼吸短促、咽喉痛、头痛、腹泻等<sup>[2]</sup>,严重者会呼吸衰竭甚至休克。在COVID-19出现的3年后,虽然许多国家的病例总体下降,但是由于SARS-CoV-2是RNA病毒,其单链结构极易变异<sup>[3]</sup>,变异病毒导致部分国家的死亡和住院人数激增。诊断COVID-19的方法主要包括逆转录聚合酶链反应(RT-PCR)、等温核酸扩增试验、抗体试验、血清试验和医学成像<sup>[4-5]</sup>。目前最主流的检测方法是通过RT-PCR进行核酸检测<sup>[6]</sup>。因为COVID-19

**【收稿日期】** 2023-08-05

**【基金项目】** 中国药品监管科学研究行动计划第二批重点项目(2022SDADRKY06)

**【作者简介】** 张子宇, 硕士, 研究方向: 计算机视觉、图像处理与分析, E-mail: 1454681376@qq.com

**【通信作者】** 赵可辉, 教授, 主任医师, 研究方向: 心脏超声, E-mail: zhaokh1202@sina.com

在感染的早期就可以传播,所以核酸检测非常适合用于无症状感染者的早期筛查。但是核酸检测的时间周期较长和假阴性率较高间接导致病毒变异疾病进展和恶化。对病毒变异体的研究表明,病毒的变异降低敏感性导致核酸检测的假阴性率很高。面对这种状况可以选择使用放射学检查进行辅助诊断。放射学检查是检测疑似 COVID-19 病例最重要的工具之一<sup>[7]</sup>,包括 CT 和 X 射线检查。CT 比 X 射线更敏感,甚至在临床症状出现和 PCR 检测出阳性之前,在 CT 上就可以看到肺纤维化<sup>[8-9]</sup>。虽然医学成像不建议用于明确诊断,但由于其他方法的局限性,它可以用于早期 COVID-19 诊断<sup>[10-11]</sup>。通过医学影像学检查作为辅助诊断手段,在 COVID-19 的各阶段疗程中起着重要的作用。一些有早发 COVID-19 症状的患者在 CT 扫描中发现 SARS-CoV-2 感染<sup>[12]</sup>。同时,他们的 RT-PCR 检测结果为阴性,几天后重复两次检测,RT-PCR 证实了 CT 扫描的诊断结果。COVID-19 的病灶以云雾状和磨玻璃样改变多见,病灶较局限,呈斑片状、亚段性或节段性分布为主,密度不均,病灶绝大多数位于肺的周边、末梢支气管和肺泡部位。如果病毒性肺炎还未得到临床有效抑制,可表现为双肺弥漫性病变,少数可呈“白肺”、实变影为主<sup>[13-14]</sup>。与使用核酸试剂检测相比,使用 CT 扫描能更加有效地筛选出无症状的病人,但这种方法要投入到实际应用中仍有局限性。人体本身的肺部特征具有不透明

度和复杂性,有部分肺炎患者可能伴随其他呼吸系统相关的疾病,这都会极大地干扰肺炎类型的判断<sup>[15]</sup>,而且即便是经验极其丰富的医生配合上现代精密诊断仪器,也会在稍不注意时发生漏诊或误诊的现象,而且每张 CT 影像都需要医生花费大量时间去审查,这将导致医疗资源更加紧张。不仅是针对肺炎 CT 影像,几乎所有的医学图像都面临着类似的挑战。为了应对当下常用的医学图像分类网络参数多、模型大、计算量大、GPU 内存占用高和训练时间长等情况,本文提出一种轻量级的 RG-DenseNet 网络模型来对 COVID-19 CT 图像进行分类。

## 1 相关方法

### 1.1 卷积块注意力模块 (Convolutional Block Attention Module, CBAM)

CBAM 是由 Woo 等<sup>[16]</sup>在 2018 年提出的用于前馈卷积神经网络简单而有效的注意力模块。CBAM 包含 2 个独立的子模块:通道注意力模块(Channel Attention Module, CAM)和空间注意力模块(Spatial Attention Module, SAM)。具体结构如图 1 所示,卷积层输出的结果,会先通过一个 CAM,得到加权结果之后,再经过一个 SAM,最终进行加权得到结果。这一模块能够通过很低的参数量和计算量的代价来提高模型在分类和检测上的精度。

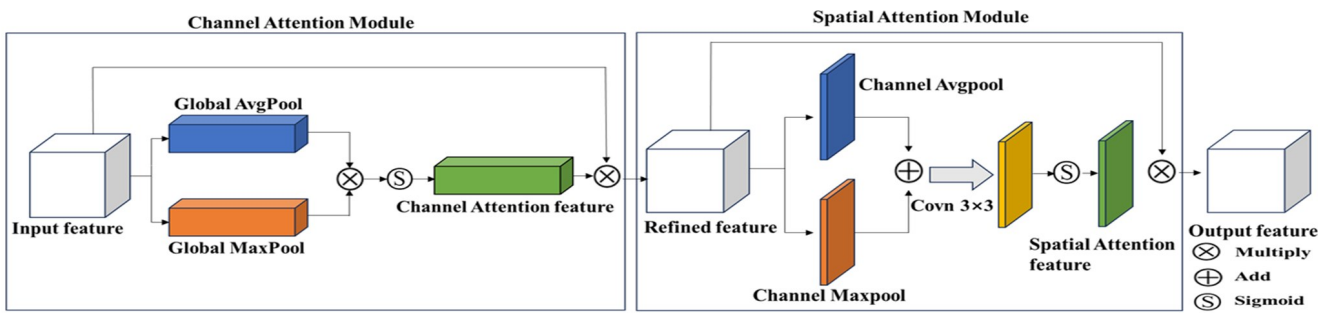


图1 卷积块注意力模块结构  
Figure 1 Convolutional block attention module

### 1.2 RepGhost 模块

GhostNet<sup>[17]</sup>是由诺亚方舟实验室在2020年提出的一个轻量级网络,由大量 Ghost 模块组成。Ghost 模块使用的是特征重复技术,特征重复技术是指通过简单连接来自不同层的一些特征图,来额外获得一些其他的特征。为了减少网络计算量,Ghost 模块将传统的卷积分为两部分进行,先通过 1×1 卷积生成通道数较小的特征图以减少计算量然后在得到特征

图的基础上通过深度可分离卷积 (Depthwise Convolution) 再进一步减少计算量,生成新的特征图,最后使用 Concat 操作将生成的新特征图与原始特征图合并起来,得到最终输出的特征图,具体结构如图 2 所示。Concat 操作是一种 0 参数量,0 浮点运算量的操作,它将数据在通道数上进行相加。但是,Concat 操作在硬件设备上的计算成本是不可忽略的。RepGhost 模块提出一种更有效的方法,通过重参数

化来生成和融合不同的特征图。Chen 等<sup>[18]</sup>发现,在硬件设备上由于内存复制导致 Concat 操作比 Add 操作效率低得多,Add 操作是进行特征图的相加,通道数不变,会损失一些特征但能节省大量推理时间。如图 2 所示,与 Ghost 模块不同,RepGhost 模块去掉了低效的 Concat 操作,信息融合过程由 Add 操作以隐含的方式执行,将支路过程添加 BN-ReLU 的方式来提升精度。Ghost bottleneck (G-beneck) 和 RepGhost bottleneck (RG-beneck1) 结构如图 3 所示。G-beneck 主要由两个堆叠的 Ghost 模块组成,第一个 Ghost 模块用作扩展层,增加通道数,第二个 Ghost 模块减少通道数,与 shortcut 路径匹配,使用 shortcut 连接两个 Ghost 模块的输入和输出,借鉴了 MobileNetV2 的结构,第二个 Ghost 模块后不适用 ReLU 层,因为深度卷积后再加 ReLU 层可能会导致效果变差,具体结构如图 3 所示,差异的地方标红色。RepGhost 在 Ghost 的基础上对 G-beneck 进行修改,将 Concat 操作改 Add,在支路上添加 BN 层,这样能控制最后 output 的通道数大幅度减小,使得模型需要训练的参数量减小。SBlock1 块为 shortcut 层。受

DenseNet<sup>[19]</sup>的启发,将原本 RG-beneck1 结构中的后激活 Conv-BN-ReLU 结构改编为前激活 BN-ReLU,将 SBlock1 (shortcut 层)变成 BN-Conv 的 SBlock2,改变完成的结果如 RG-beneck2 所示。深度残差网络中的身份映射证明<sup>[20]</sup>,前激活能提升网络的训练和测试速度,还能在一定程度上提升精度。

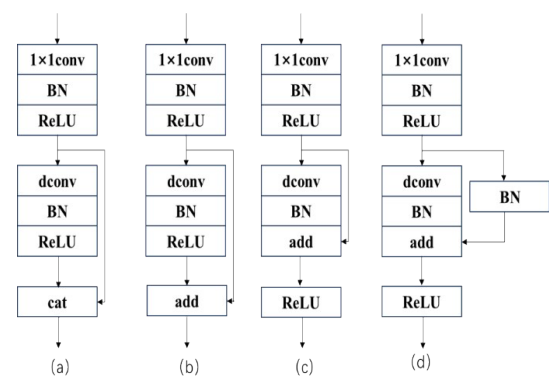


图 2 RepGhost 模块结构  
Figure 2 RepGhost module structure

a: 原始 Ghost 模块; b~d: 由 Ghost 模块向 RepGhost 模块改进;  
dconv: 深度卷积

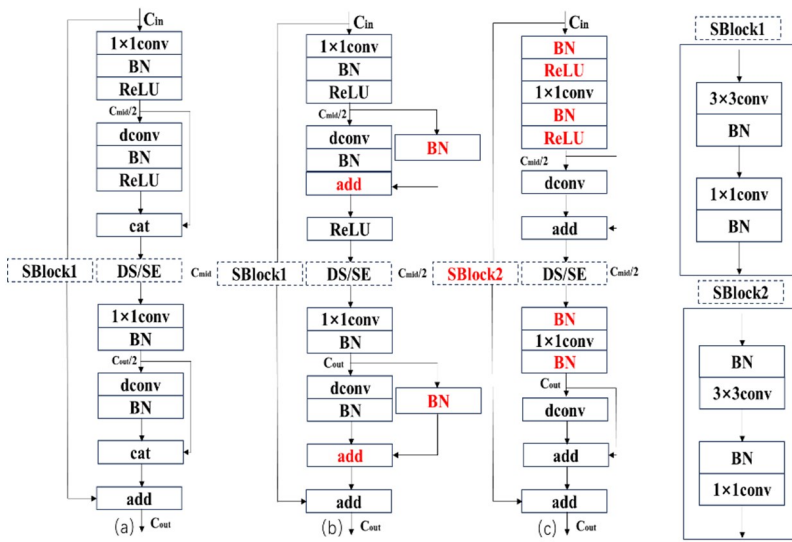


图 3 RG-beneck2 模块结构  
Figure 3 RG-beneck2 module structure

a: G-beneck; b: RG-beneck1; c: RG-beneck2; SBlock1 和 SBlock2: 直连块 (shortcut)。DS: 下采样层; SE 下采样层。虚线中的块仅在必要时插入。模块不同处用红色标注

1.3 RG-DenseNet

组成 DenseNet121 结构为卷积层、DenseBlock、Transition layers、池化层和全连接层。其中 DenseBlock 由数个前激活 [1×1 conv, 3×3 conv] 组成,此处的 [1×1 conv, 3×3 conv] 即为 Bottleneck 结构,具体结构如图 4 所示,  $k$  表示增长率,用来固定

中间和输出的通道数。本文构建一种轻量级的 RG-DenseNet 模型,以 DenseNet121 网络为基础结构。将原本的 DenseBlock 中的 Bottleneck 替换成 RepGhost 模块,来达到轻量化的目的。每 2 个 RepGhost 模块组可以组成更高效的 RG-beneck2,使用 RG-beneck2 代替 2 个 Dense Bottleneck,组成的新



模块称为RG-DenseBlock。为了尽可能地减少模型的参数量与计算量,本文设计的模型中CBAM模块主要添加在stage1的池化层之前和stage4的最后一个RG-beneck2之后。将Bottleneck替换为RG-beneck2后由于RG-beneck2舍弃Concat操作,导致通道数大幅度下降,能够有效减少模型的参数量和计算量。将网络的输入通道数减少为32,增长率 $k=16$ ,RG-DenseNet具体网络结构如图5所示。

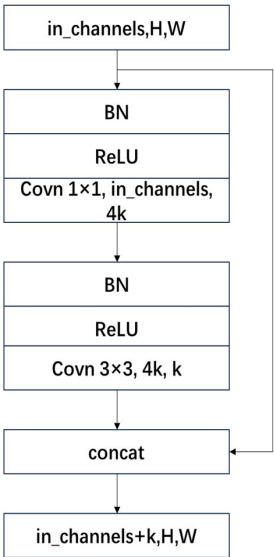


图4 Bottleneck 结构  
Figure 4 Bottleneck block

## 2 实验过程

### 2.1 实验环境

实验使用 Windows 11 操作系统,PyTorch 深度学习开发框架,用 Python 作为开发语言。实验采用的 CPU 为 Intel Core i9-12900h, GPU 为 NVIDIA GeForce RTX 3070Ti(Laptop)。

### 2.2 数据集

COVIDx CT 是一个大型的 COVID CT 图像公开数据集,该数据集包含中国生物信息中心(CNCB)收集的 1 489 名患者的胸部 CT 扫描图像(共 104 009 张)<sup>[21]</sup>。在此基础上研究人员为了提升数据集的多样性添加了世界各地患者的 CT 图像并提出 COVIDx CT-2 数据集。本研究使用 COVIDx CT-2 数据集集中的 COVIDx CT-2A<sup>[22]</sup>,COVIDx CT-2 数据集包含多个国家 3 745 名患者的 CT 扫描图像(共 194 922 张),分为正常(健康志愿者)CT 图像、普通肺炎 CT 图像和 COVID-19 CT 图像 3 类。该数据集可以在 Kaggle 上公开获取,链接如下: <https://www.kaggle.com/datasets/hgunraj/covidxct>。具体数据如表 1 所示。

### 2.3 评价指标

由于使用的 COVIDx CT-2A 数据集中数据是根据患者进行划分,重新划分可能导致训练集与测试集中有 CT 图像来源于同一患者,影响准确度,为了保证实验的客观准确使用原数据集的划分进行 3 分

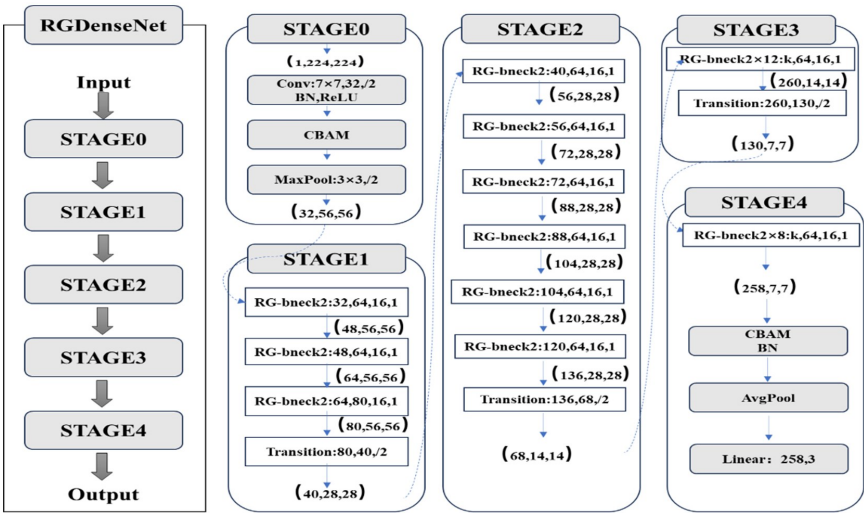


图5 RG-DenseNet 结构图  
Figure 5 RG DenseNet structure

类实验。为了评估本文所提出的方法,选用混淆矩阵、准确率(Accuracy)、精确率(Precision)、召回率(Recall)、特异度(Specificity)、F1 分数作为性能指标。混淆矩阵中真正例(TP)、假正例(FP)、真反例

(TN)和假反例(FN)是计算性能指标的主要组成部分。TP 代表真实值是 positive,模型认为是 positive 的数量;FN 代表真实值是 positive,模型认为是 negative 的数量;FP 代表真实值是 negative,模型认为

表1 COVIDx CT-2A数据集的图像(患者数)分布情况  
Table 1 Distribution of images (number of patients) in the COVIDx CT-2A dataset

数据集	正常	普通肺炎	COVID-19	总计
训练集	35 996(321)	25 496(558)	82 286(1 958)	143 778(2 837)
验证集	11 842(126)	7 400(190)	6 244(166)	25 486(482)
测试集	12 245(126)	7 395(125)	6 018(175)	25 658(426)
总计	60 083(573)	40 291(873)	94 548(2 299)	194 922(3 745)

是 positive 的数量;TN 代表真实值是 negative,模型认为是 negative 的数量。准确率、精确率、召回率、特异性和F1分数分别定义在式(1)~式(5)中:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$
 (1)

$$\text{Precision} = \frac{TP}{TP + FP}$$
 (2)

$$\text{Recall} = \frac{TP}{TP + FN}$$
 (3)

$$\text{Specificity} = \frac{TN}{TN + FP}$$
 (4)

$$F1 = \frac{2\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$
 (5)

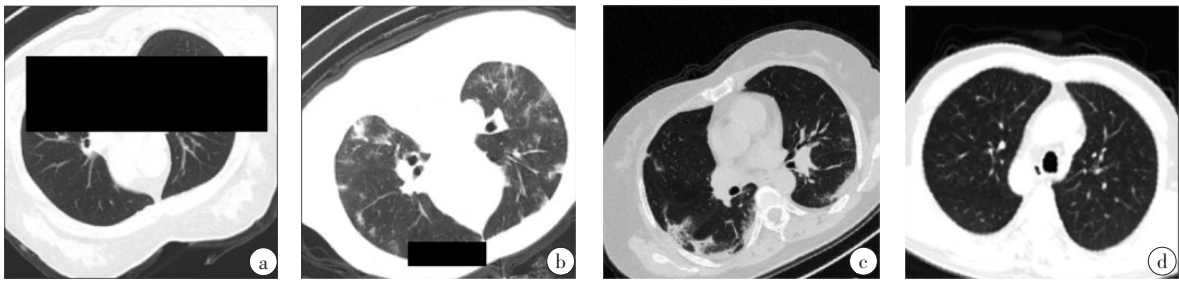


图6 预处理后图像

Figure 6 Preprocessed images

a~d 分别代表在预处理后的图像数据集中随机抽取的4张肺部CT图像

参数设置 epochs(训练周期)为 30, batch sizes(批量大小)为 64。使用 Adam 优化器,学习率为  $5 \times 10^{-4}$ , beta 系数(计算梯度以及梯度平方的运行平均值)为 (0.9, 0.999), 维持数值稳定性的小实数 eps 为  $10^{-8}$ , 权重衰减(L2 惩罚)为 0.000 1。学习率衰减策略使用余弦退火策略(CosineAnnealingLR), 每个 batch 迭代都进行学习率更新,  $T\text{-max} = \text{epochs} \times \text{iters}$ , 其中, iters 表示完成一个 epoch 所需的迭代次数,  $T\text{-max}$  代表总迭代次数,  $\text{iters} = 2\ 247$ ,  $\text{epochs} = 30$ ,  $T\text{-max} = 67\ 410$ 。

3 结果分析

3.1 消融实验

2.4 数据预处理及实验参数设置

数据集图像大小为  $512 \times 512$ , 由于图像来源于世界各地的患者且使用不同仪器获得, 所以图像有一定的差距, 必须先进行一定的预处理, 本文选择将  $512 \times 512$  的原始图像按中心裁剪成  $340 \times 380$  的大小, 这样能很好地减少其余部分在缩放中的影响并保存图像中肺部区域, 对图像调整到模型需要的  $224 \times 224$  进行归一化并进行数据增强。本文使用以下 4 种数据增强方式。(1)随机水平翻转: 以 50% 的概率对输入图像进行水平翻转; (2)随机垂直翻转: 以 50% 的概率对输入图像进行垂直翻转; (3)随机旋转: 以  $-30^\circ \sim 30^\circ$  随机进行图像旋转; (4)随机擦除(RandomErasing): 以 50% 的概率对图像的随机部位选择一个矩形范围进行擦除, 擦除区域与输入图像的比例范围(scale)为 0.02~0.33, 擦除区域的纵横比范围(value)为 0.3~3.3。随机擦除能将随机选中区域的图片信息遮盖<sup>[23]</sup>, 强迫模型学习该区域外的特征进行识别, 能够在一定程度上避免模型陷入局部最优, 提高模型的泛化能力。图 6 所示的肺部 CT 图像就是数据预处理后的图像。

为了证明改进后的 RG-DenseNet 网络模型和选取的图像增强方法对提升网络的性能以及减少参数量和计算量有贡献, 设计消融实验。本文 8 组实验使用不同的组合, 组合方式包括: CBAM 模块的插入位置(CBAM 在 stage1 前/后)、4 种图像增强方式、以及在网络中使用 RG-beneck1 或 RG-beneck2。主要评判标准为网络模型的参数量与准确率。实验 1 使用 DenseNet 模型, 实验 2~8 使用 RG-DenseNet 模型。

消融实验结果如表 2 所示, 实验 2~8 与实验 1 对比可以看出当使用 RG-beneck1 或 RG-beneck2 替换 Bottleneck 时, 都可以使模型的参数量与参数大小下降 92.7% 以上。实验 8 和实验 6 对比可以看出用

RG-beneck2 比 RG-beneck1 能提升准确度且增加的参数大小仅有 0.05 MB。添加 CBAM 模块后会使参数量有小幅度的增加。由实验 7 和实验 8 对比可知在 stage1 前添加 CBAM 会对模型有负作用,由实验 6 和实验 5 对比可知在 stage4 后添加 CBAM 比在 stage1 前添加 CBAM 增加 0.06 MB 的参数大小,但提升 0.2% 的准确度。由实验 4 与实验 2 对比发现使用图

像增强 1、2、3 比未使用图像增强准确率高 1.15%,实验 5 与实验 4 对比发现使用图像增强 1、2、3、4 比使用图像增强 1、2、3 准确率高 0.09%。实验证明通过在 stage4 后添加 CBAM 模块和使用 RG-beneck2 替换 DenseNet121 中的 Bottleneck,可以使得 RG-DenseNet 模型能够在减少参数量的前提下保持准确度与 DenseNet121 相似。

表 2 消融实验结果  
Table 2 Ablation experiment results

网络模型	编号	CBAM在 stage1 前	CBAM在 stage4 后	RG-beneck1	RG-beneck2	图像增强 1、2、3	图像增强 4	准确率/%	参数量	参数大小/MB
DenseNet	1					√	√	98.94	6 950 659	26.51
RG-DenseNet	2	√		√				97.47	478 448	1.83
	3				√			98.20	492 528	1.88
	4	√		√		√		98.62	478 448	1.83
	5	√		√		√	√	98.71	478 448	1.83
	6		√	√		√	√	98.91	494 704	1.89
	7	√	√		√	√	√	98.85	509 139	1.94
	8		√		√	√	√	98.93	508 784	1.94

√表示实验使用了该方法,图像增强 1、2、3、4 分别代表使用随机水平翻转、随机垂直翻转、随机旋转、随机擦除

3.2 对比实验

为了证明本文提出的 RG-DenseNet 模型的优越性,设计了对比实验。该实验将 RG-DenseNet、原网络 DenseNet121 和近年来表现较好的轻量化网络如 MobileNetV2、MobileNetV3、ShuffleNetV2 做对比。对比实验中除 DenseNet121 外,其余模型用的硬件设备、超参数、迭代次数、图像增强方法完全一致。DenseNet121 由于设备显存不够, batchsize 设置为 32 进行实验,其余指标与其他模型实验一致。实验结果如表 3 和表 4 所示。

由表 3、表 4 可知,本文提出的 RG-DenseNet 的准确率为 98.93%、精确率为 98.7%、召回率为 98.97%、特异性为 99.48%、F1 分数为 98.83%。其中准确率与 DenseNet121 相近,仅相差 0.01%。在准确率、精确率、召回率、特异性和 F1 分数这些指标上优于 MobileNetV2、MobileNetV3、ShuffleNetV2。RG-DenseNet 的参数量为 508 784,参数大小为 1.94 MB,最终的模型保存大小为 2.67 MB;参数量和参数大小较 DenseNet121 减少 92.7%,模型保存大小减少 90.1%,计算量减少 92.7%。RG-DenseNet 的 GPU 内存占用只有 DenseNet121 的 53.5%,训练时间减少 1 h 30 min 17 s,测试时间减少 23.7 s。这是由于本文提出的 RG-DenseNet 将原本的

DenseBlock 换成 RepGhost 模块,其中 DenseBlock 中的 3×3 卷积被替换成了深度可分离卷积(Depthwise Conv),深度可分离卷积在各个通道上进行卷积再使用 1×1 卷积,相较于常规卷积有效减少了参数量。RepGhost 模块首先通过 1×1 的卷积生成通道数较小的特征图以减少的计算量,然后在得到的特征图基础上,通过深度可分离卷积再进一步减少计算量并且生成新的特征图,最后将两组特征图拼接到一起,有效减少提取特征所需的参数量与计算量。DenseNet121 中使用的 Concat 不断进行通道数维度上的连接,虽然能够获得大量特征,但必然会导致计算量和显存用量大幅度的提高,而且 Concat 操作后提取的中间特征图中存在大量的冗余。本文将 DenseBlock 之间的 Concat 操作替换为 add 操作,减少通道数,虽然 add 操作会损失提取到的部分特征,但由于卷积提取特征过程中大部分都是无用的冗余。且 RG-beneck2 中保留了 Ghost 的方法由 2 个 RepGhost 模块组成用第一个 RepGhost 模块进行通道数的增加,第二个 RepGhost 模块进行通道数的减少来和 shortcut 进行路径匹配,用更少的参数量与计算量获得更多的特征。由此可见,本文提出的 RG-DenseNet 模型大幅减少参数量、计算量、GPU 的内存占用和模

表3 模型精度对比(%)

Table 3 Comparison of model performance (%)

网络模型	类别	准确率	精确率	召回率	特异性	F1分数
DenseNet121	正常	-	99.37	98.68	99.43	99.02
	非新冠肺炎	-	99.43	99.68	99.77	99.55
	新冠肺炎	-	97.50	98.59	99.23	98.04
	平均	98.94	98.77	98.98	99.48	98.87
MobileNetV2 <sup>[24]</sup>	正常	-	99.44	98.48	99.49	98.96
	非新冠肺炎	-	99.24	99.49	99.69	99.36
	新冠肺炎	-	97.24	98.85	99.14	98.04
	平均	98.86	98.64	98.94	99.44	98.79
MobileNetV3 <sup>[25]</sup>	正常	-	99.57	98.33	99.61	98.95
	非新冠肺炎	-	98.82	99.59	99.52	99.20
	新冠肺炎	-	97.09	98.60	99.09	97.84
	平均	98.76	98.49	98.84	99.41	98.66
ShuffleNetV2 <sup>[26]</sup>	正常	-	99.10	98.54	99.19	98.82
	非新冠肺炎	-	99.47	99.08	99.79	99.27
	新冠肺炎	-	97.25	98.85	99.14	98.04
	平均	98.77	98.61	98.82	99.37	98.71
RG-DenseNet	正常	-	99.52	98.62	99.57	99.07
	非新冠肺炎	-	99.31	99.70	99.72	99.50
	新冠肺炎	-	97.28	98.60	99.15	97.94
	平均	98.93	98.70	98.97	99.48	98.83

表4 模型大小及测试时间对比

Table 4 Comparisons of model size and testing time

网络模型	参数量	参数大小/MB	计算量/MB	GPU内存使用/MB	模型保存大小/MB	训练时间	测试时间/s
DenseNet121	6 950 659	26.51	2 886.79	9 860	27.11	7 h 9 min 8 s	91.5
MobileNetV2	2 227 139	8.50	304.27	6 818	8.73	6 h 16 min 19 s	81.2
MobileNetV3	3 875 073	14.78	250.87	5 854	14.99	6 h 8 min 27 s	72.5
ShuffleNetV2	1 255 379	4.79	140.21	2 882	4.95	5 h 34 min 29 s	74.2
RG-DenseNet	508 784	1.94	209.32	5 282	2.67	5 h 38 min 51 s	67.8

型保存大小,有效缩短模型的训练与测试时间,达到轻量化的目的。

和其余轻量化模型相比, RG-DenseNet 的计算量与 ShuffleNetV2 相比高 49.2%, GPU 内存使用高 83.3%, 但参数量、参数大小只有 ShuffleNetV2 的 40.5%, 准确率比 ShuffleNetV2 高 0.16%。ShuffleNetV2 的训练时间是最短的, RG-DenseNet 的训练时间比 ShuffleNetV2 多 4 min 22 s, 相差不算太大。与 MobileNetV2、MobileNetV3 相比, RG-DenseNet 准确度分别高 0.07% 和 0.17%; 参数量和参数大小分别为 MobileNetV2 和 MobileNetV3 的 22.8% 和 13.1%; 计算量分别为

MobileNetV2 和 MobileNetV3 的 68.8% 和 83.4%; GPU 内存占用比 MobileNetV2 和 MobileNetV3 分别减少 1 536 MB 和 572 MB; 模型保存大小分别为 MobileNetV2 和 MobileNetV3 的 30.6% 和 17.8%。训练时间比 MobileNetV2 和 MobileNetV3 分别减少 37 min 28 s 和 29 min 36 s, 测试时间较 MobileNetV2 和 MobileNetV3 分别减少 13.4 s 和 4.7 s。

RG-DenseNet 各项模型大小指标都低于 MobileNetV2、MobileNetV3, 计算量和 GPU 占用高于 ShuffleNetV2, 其余指标都低于 ShuffleNetV2, 并且 RG-DenseNet 在相同条件下精度指标与 DenseNet121



近乎一致,高于 MobileNetV2、MobileNetV3 和 ShuffleNetV2,证明本文提出的轻量级模型精度高于常用轻量级模型,参数量低,GPU 占用相似,训练时间与测试时间短,具有实际应用价值。

#### 4 结论与展望

本文提出一种用于 COVID-19 辅助诊断的轻量级神经网络,称为 RG-DenseNet。该网络是基于 DenseNet121 改进得来的,在最后的平均池化层与全连接层之前添加 CBAM 模块,并将 DenseNet 中的 Bottleneck 模块每两层替换为一个 RG-beneck2 模块。对于模型在轻量化后精度降低的问题,提出旋转和随机擦除的方法进行图像增强。使用 COVIDx CT-2A 数据集进行消融实验与对比实验,相较于 MobileNetV2、MobileNetV3,以更少的参数量和计算量获得更高的准确度,训练速度与 ShuffleNetV2 相近。相较于原模型 DenseNet121, RG-DenseNet 在保持准确度几乎不变的前提下,减少 92.7% 的参数量、53.5% 的 GPU 内存使用量和 21.0% 的训练时间。结果表明 RG-DenseNet 具有参数量和计算量少、准确度高的特点,能够明显减少显存占用,提升训练与验证速度。接下来的研究工作是扩充数据集以及对模型和图像增强方法进行优化,提高模型的精度。

#### 【参考文献】

- [1] Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China[J]. N Engl J Med, 2020, 382(8): 727-733.
- [2] Singhal T. A review of coronavirus disease-2019 (COVID-19)[J]. Indian J Pediatr, 2020, 87(4): 281-286.
- [3] 曹广文. 新型冠状病毒进化、相关流行特征和特异性防控中的关键问题[J]. 上海预防医学, 2020, 32(9): 697-703.  
Cao GW. Key issues in the evolution, related epidemic characteristics and specific prevention and control of novel coronavirus[J]. Shanghai Journal of Preventive Medicine, 2020, 32(9): 697-703.
- [4] 宋兰, 宋伟, 隋昕, 等. 北京协和医院普通型新型冠状病毒肺炎患者的初诊临床特征及 CT 影像学表现[J]. 中国医学科学院学报, 2020, 42(3): 376-382.  
Song L, Song W, Sui X, et al. Preliminary study on clinical features and CT findings of common-type coronavirus disease 2019 patients in Peking Union Medical College Hospital [J]. Acta Academiae Medicinae Sinicae, 2020, 42(3): 376-382.
- [5] Ghaderzadeh M, Eshraghi MA, Asadi F, et al. Efficient framework for detection of COVID-19 omicron and delta variants based on two intelligent phases of CNN models[J]. Comput Math Methods Med, 2022. Doi: 10.1155/2022/4838009.
- [6] Kovács A, Palásti P, Veréb D, et al. The sensitivity and specificity of chest CT in the diagnosis of COVID-19[J]. Eur Radiol, 2021, 31(5): 2819-2824.
- [7] Rocklöv LJ. The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus[J]. J Travel Med, 2021, 28(7): taab124.
- [8] Yang W, Shaman J. COVID-19 pandemic dynamics in India and impact of the SARS-CoV-2 Delta (B.1.617.2) variant[J]. medRxiv, 2021. Doi: 10.1101/2021.06.21.21259268.
- [9] Ghaderzadeh M, Asadi F, Jafari R, et al. Deep convolutional neural network-based computer-aided detection system for COVID-19 using multiple lung scans: design and implementation study[J]. J Med Internet Res, 2021, 23(4): e27468.
- [10] ACR. ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected Covid-19 infection [EB/OL]. <https://www.acr.org/Advocacy-and-Economics/ACR-PositionStatements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-Covid-19-Infection> (accessed on 31.05.20).
- [11] 张振华, 吉祥, 张劲松, 等. 基于 AI 技术的新型冠状病毒肺炎 CT 影像特点分析[J]. 医疗卫生装备, 2020, 41(5): 6-8.  
Zhang ZH, Ji X, Zhang JS, et al. Analysis of CT imaging characteristics of novel coronavirus pneumonia based on AI technology[J]. Chinese Medical Equipment Journal, 2020, 41(5): 6-8.
- [12] 梁琪. 新型冠状病毒肺炎影像学检查、诊断及医院内感染预防与控制: 湖南省放射学专家共识[J]. 中南大学学报(医学版), 2020, 45(3): 221-228.  
Liang Q. Imaging examination, diagnosis, and control and prevention of nosocomial infection for coronavirus disease 2019: Expert consensus of Hunan radiologist [J]. Journal of Central South University (Medical Sciences), 2020, 45(3): 221-228.
- [13] 向容, 刘翔雄, 吴勇. 新型冠状病毒肺炎的 CT 表现[J]. 中国 CT 和 MRI 杂志, 2022, 20(7): 70-72.  
Xiang R, Liu XC, Wu Y. Application of chest CT in patients with novel coronavirus pneumonia[J]. Chinese Journal of CT and MRI, 2022, 20(7): 70-72.
- [14] 高亭, 徐仰玲, 何小鹏, 等. 40 例新型冠状病毒肺炎的流行病学及临床特征分析[J]. 中国呼吸与危重症监护杂志, 2020, 19(2): 148-153.  
Gao T, Xu YL, He XP, et al. Epidemiological and clinical characteristics of 40 patients with coronavirus disease 2019 outside Hubei [J]. Chinese Journal of Respiratory and Critical Care Medicine, 2020, 19(2): 148-153.
- [15] 陈海霞. 新型冠状病毒肺炎患者肺部病变的影像学表现[J]. 实用医药杂志, 2020, 37(8): 715-718.  
Chen HX. Imaging manifestations of lung lesions in patients with novel COVID-19[J]. The Journal of Practical Medicine, 2020, 37(8): 715-718.
- [16] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.
- [17] Han K, Wang Y, Tian Q, et al. GhostNet: more features from cheap operations [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1580-1589.
- [18] Chen C, Guo Z, Zeng H, et al. RepGhost: a hardware-efficient ghost module via Re-Parameterization [J]. arXiv preprint arXiv: 2211.06088, 2022.
- [19] 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.
- [20] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks [C]//Computer Vision-ECCV 2016: 14th European Conference, Proceedings, Part IV 14. Springer International Publishing, 2016: 630-645.
- [21] Gunraj H, Wang LD, Wang A. COVIDNet-CT: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest CT images[J]. Front Med, 2020, 7: 608525.
- [22] Aboutaleb H, Abbasi S, Shafiee MJ, et al. COVID-Net CT-S: 3D convolutional neural network architectures for COVID-19 severity assessment using chest CT images [J]. arXiv preprint arXiv: 2105.01284, 2021.
- [23] Zhong Z, Zheng L, Kang G, et al. Random erasing data augmentation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 13001-13008.
- [24] Sandler M, Howard A, Zhu M, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 4510-4520.
- [25] Howard A, Sandler M, Chu G, et al. Searching for mobileNetV3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2019: 1314-1324.
- [26] Ma N, Zhang X, Zheng HT, et al. ShuffleNet v2: practical guidelines for efficient CNN architecture design [C]//Proceedings of the European Conference on Computer vision (ECCV). 2018: 116-131.

(编辑:薛泽玲)