

基于深度学习的 COVID-19 智能诊断系统

贾楠¹, 李燕², 郭静霞¹, 徐立¹, 白金牛¹

(1. 内蒙古科技大学 包头医学院计算机科学与技术学院, 内蒙古 包头 014040;

2. 包头市中心医院, 内蒙古 包头 014040)

摘要: 针对 2019 年 12 月在全球爆发的新冠肺炎, 传统的 RT-PCR 检测易出现假阴性或弱阳性并且检测时间长等问题, 设计了一种基于深度学习对胸部 X 光片辅助诊断新冠肺炎的方法; 采用 UNet 分割模型实现了肺部 ROI 区域的自动分割, 对分割后的影像应用 TrivialAugment 数据增强策略, 通过 MBCA-COVIDNET 模型实现胸部 X 光片三分类 (新冠肺炎、其它肺炎、正常) 任务, 该模型以 MobileNetV2 作为骨干网络, 并在其中加入坐标注意力机制 (CA); 利用 Hugging Face 和 Flask 开源软件搭建了 COVID-19 智能诊断系统; 实验结果表明 MBCA-COVIDNET 模型在 COVID-QU-Ex Dataset 测试集上取得了高达 97.98% 的准确率, 而模型的参数量和 MACs 仅有 2.23 M 和 0.33 G, 该智能诊断系统能够很好地辅助医生进行基于胸片的 COVID-19 诊断, 提升诊断的准确率以及诊断效率。

关键词: 胸片; 新冠肺炎; 深度学习; 分类模型; 坐标注意力机制

Intelligent Diagnosis System for COVID-19 Based on Deep Learning

JIA Nan¹, LI Yan², GUO Jingxia¹, XU Li¹, BAI Jinniu¹

(1. Baotou Medical College, Inner Mongolia University of Science and Technology, Baotou 014040, China;

2. Baotou Central Hospital, Baotou 014040, China)

Abstract: Aimed at the global outbreak of COVID-19 in December 2019, traditional RT-PCR tests are prone to the questions of false negative, weak positive results and long detection, A deep learning-based method for assisting in the diagnosis of COVID-19 on chest X-ray was proposed. The UNet segmentation model is used to accomplish the automatic segmentation of the lung ROI area, the segmented image is enhanced by the Trivial Augment data enhancement strategy, and the X-ray images of the chest are classified into three categories (normal, COVID-19, and other pneumonia). The model adds coordinate attention mechanism (CA) to MobileNetV2 as the backbone network; the open-source software such as Hugging Face and Flask is adopted to construct the COVID-19 intelligent diagnosis system; The experimental results show that the MBCA-COVIDNET model reaches the accuracy of 97.98% on the COVID-QU-Ex Dataset test set, even though the model's parameters and MACs are 2.23 M and 0.33 G respectively. The intelligent diagnostic system can help doctors diagnose COVID-19 based on the chest X Ray (CXR), and improve the accuracy and efficiency of the diagnosis.

Keywords: CXR; COVID-19; deep learning; classification model; CA

0 引言

自 2019 年 12 月以来, 新冠肺炎仍在全球蔓延。所以快速、准确地识别出新冠肺炎对于患者的及时医治以及控制疫情的传播具有非常重要的意义。目前新冠肺炎检测的金标准是逆转录聚合酶链反应 (RT-PCR)^[1], 但在实践中常常会因为采样质量或者病毒的载量出现假阴性或弱阳性结果, 从而导致较高的复检率, 容易延误治疗, 增加其他人员感染的风险。利用胸部 X 光片 (CXR) 也可以很好地进行新冠肺炎的辅助诊断。CXR 是最常见的诊断放射学检查之一, CXR 成像比 CT 成像更容易获得。因为 CT 扫描仪价格高昂且维护成本较高, 而 CXR 系统相对较为便宜, 且在乡镇一级的医院也普遍存在。但想通过 CXR 影像来识别患

者是否感染新型冠状病毒, 非常依赖放射科医生的经验。即便是训练有素的放射科医生也很容易犯错误, 因为其它类型的肺炎和新冠肺炎往往具有相似的视觉特征。因此, 用深度学习的方法构建一个模型支持放射科医生的决策过程, 加速新冠筛查, 帮助减少诊断错误具有非常重要的意义。

深度学习通过使用卷积神经网络可实现对医学影像特征的自动提取, 具有强大的表征能力^[2]。因此将深度学习应用于针对胸片的新冠肺炎检测也是当下研究的热点。Narin 等^[3]运用 ResNet50、ResNet101、ResNet152、InceptionV3 和 Inception-ResNetV2 五个模型, 通过五折交叉验证对胸片进行训练, 实现了 COVID-19、正常 (健康)、病

收稿日期: 2022-12-02; 修回日期: 2023-01-11。

基金项目: 内蒙古自治区高等学校科学研究项目 (NJZY22050); 包头医学院研究基金项目 (BYJJ-KCRH202206)。

作者简介: 贾楠 (1984-), 男, 内蒙古包头人, 硕士, 讲师, 主要从事医学图像处理方向的研究。

通讯作者: 白金牛 (1967-), 男, 内蒙古鄂尔多斯人, 硕士, 教授, 主要从事人工智能在医学图像中的应用方向的研究。

引用格式: 贾楠, 李燕, 郭静霞, 等. 基于深度学习的 COVID-19 智能诊断系统[J]. 计算机测量与控制, 2023, 31(4): 96-103.

病毒性肺炎和细菌性肺炎四分类任务。实验结果显示 ResNet50 模型获得了最高的分类准确率。Wang 等^[4] 基于 DenseNet-121 构造了一个深度学习管道, 除了可区分由 COVID-19 引起的病毒性肺炎和其它类型的肺炎外, 可以对新冠肺炎的严重程度进行评估。Wang 等^[5] 利用神经网络搜索技术构建了 COVID-Net 模型, 实现了正常 (Normal)、其它肺炎 (Non-COVID)、新冠肺炎 (COVID-19) 的三分类检测任务, 并取得了不错的结果。Siddhartha 等^[6], 提出了一种名为 COVIDLite 的方法, 该方法在图像预处理阶段运用白平衡与限制对比度自适应直方图均衡化相结合, 在神经网络构建中又采用了深度可分离卷积, 最终实现了 Normal、Non-COVID、COVID-19 的三分类准确率为 97.12%。Oh 等^[7] 提出了一种基于 patch 的深度神经网络架构, 可以在小数据集上稳定地训练, 从而解决了新冠肺炎数据集较少的问题。

本文将构建一个轻量化的神经网络模型, 该模型在满足检测精度的前提下, 参数量要尽可能的小。同时为了使所训练的模型不是仅仅停留在实验室, 而是能够真正地运用到临床辅助诊断中, 本文还将搭建一套方便医院影像工作人员使用的 web 系统。

1 智能诊断系统结构及原理

本系统原理如图 1 所示。包括如下 5 个模块, 1) 数据集的整理和划分; 2) 分割模型; 3) 图像预处理; 4) 分类模型; 5) 模型部署。

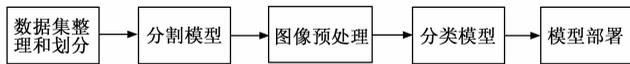


图 1 智能诊断系统原理图

具体说明如下: 1) 数据集整理和划分是将收集到影像数据按照类别进行整理, 一个类别对应一个文件夹, 文件夹中存入相应类别的影像。然后将整理后的影像按照一定比例划分为训练集、验证集和测试集; 2) 用训练集训练一个肺部分割模型, 实现胸部 X 光片的肺部区域的分割; 3) 对分割后的肺部影像进行图像预处理, 扩充训练样本的数量, 提高模型的泛化能力; 4) 基于处理后的影像数据训练一个新冠肺炎检测三分类模型; 5) 将分割模型和分类模型进行组合部署, 满足实际应用需求。

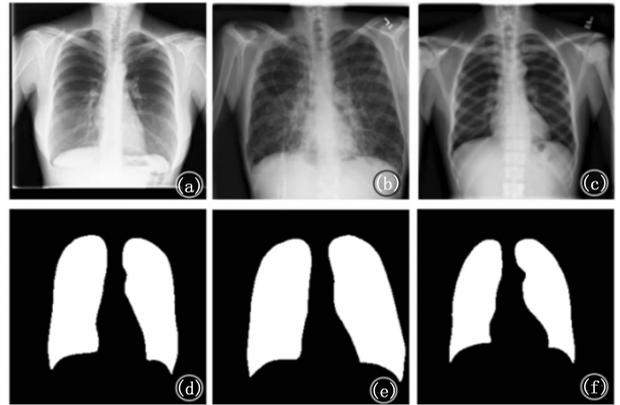
2 COVID-19 深度学习检测技术与方法

2.1 数据集介绍

本研究训练模型使用的是类别为 COVID-19、Non-COVID、Normal 的三分类数据集。一方面此类公开数据集较多; 另一方面这种分类方法, 有助于临床医生对 COVID-19 进行初筛分诊, 如果检测为 COVID-19 可通过 RT-PCR 进一步进行确诊。同时 COVID-19 和其它肺炎采取的治疗方案不同, 如此分类还有助于医生快速采取相应的治疗策略。数据集详细说明如下:

COVID-QU-Ex Dataset^[8-12], 是本文训练模型的数据

集。该数据集来源于 kaggle 网站, 包含 33 920 张胸部 x 光片, 其中 COVID-19 为 11 956 张, Non-COVID 为 11 263 张, Normal 为 10 701 张, 并且所有影像均包含肺部边缘轮廓分割掩膜。本研究将该数据集进行训练集、验证集和测试集的划分。影像示例如图 2 所示, 数据集详细统计见表 1。



(a) ~ (c) 分别为新冠肺炎患者、其它肺炎患者和正常人胸部 x 光片; (d) ~ (f) 分别为新冠肺炎患者、其它肺炎患者和正常人肺部边界区域。

图 2 COVID-QU-Ex Dataset 影像示例

表 1 COVID-QU-Ex Dataset 详细统计

图像类别	图像大小	训练集个数	验证集个数	测试集个数
COVID-19	256 * 256	7 658	1 903	2 395
Non-COVID	256 * 256	7 208	1 802	2 253
Normal	256 * 256	6 849	1 712	2 140
总数		21 715	5 417	6 788

此外为了验证模型的泛化能力和稳定性, 本论文还用 COVID CXR Image Dataset^[13-15] 对模型进行了测试。该数据集也来源于 kaggle 网站, 数据集包括 Normal、Non-COVID 和 COVID-19 患者的胸部 X 光片的前后方 (PA) 视图。共有 1823 张 CXR 影像, 包含 536 张 COVID-19 影像, 619 张 Non-COVID 影像和 668 张 Normal 影像。数据集中 COVID-19 病例的年龄范围为 18~75 岁。

2.2 检测方法

本研究利用深度学习对 CXR 影像进行分类的工作流程如图 3 所示。首先利用 COVID-QU-Ex Dataset 数据集中的肺部胸片以及肺部分割掩膜, 训练一个 UNet 分割模型, UNet 模型在医学图像分割领域表现优异, 该模型能够很好地实现 CXR 影像肺部区域 (ROI 区域) 的自动分割。其次将自动提取后的 ROI 区域送入图像分类模型, 最终实现 Normal、Non-COVID、COVID-19 的三分类。

本研究比较了当下流行的 10 种主流深度学习分类模型。详细的实验数据见文中第 3 部分。在以上 10 种模型的对比实验中, 发现针对 COVID-QU-Ex Dataset 数据集, 同系列的模型并不是模型深度越深, 模型精度越高。而是随着模型深度增加到一定程度, 精度不增反而略微下降。以

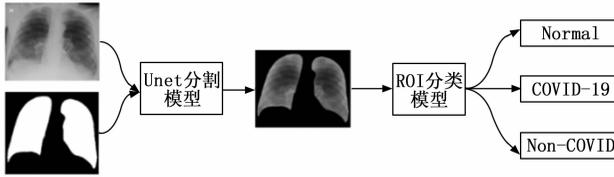


图 3 工作流程图

RestNet 系列为例, 模型在验证集上的准确率如下: Rest-Net18 为 0.947、RestNet34 为 0.948、RestNet50 为 0.946、RestNet101 为 0.946。图 4、图 5 分别为 RestNet 系列在训练集和验证集上直观展示。

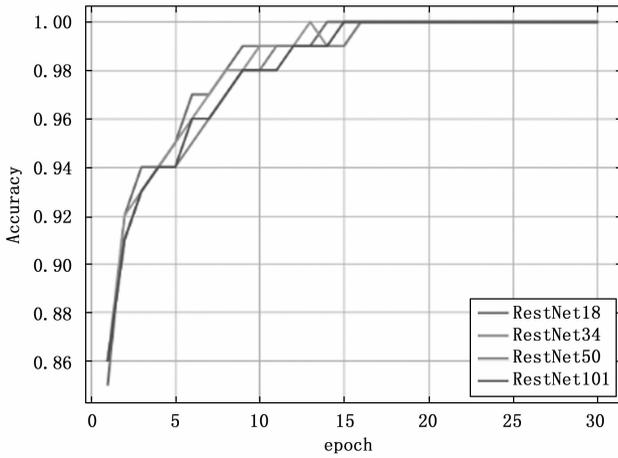


图 4 训练集准确率

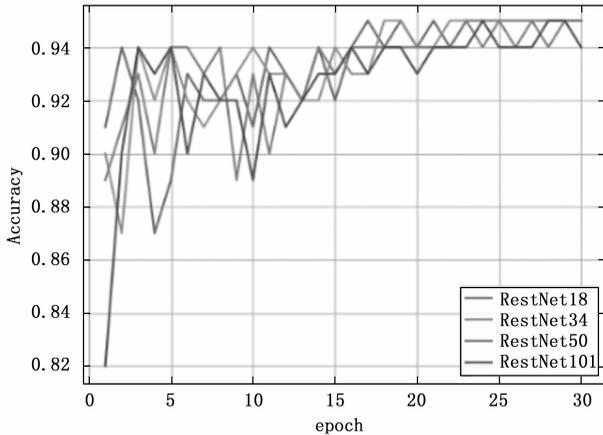


图 5 验证集准确率

受此启发, 针对该分类任务, 构建模型时可以不用设计过深的网络, 同样也能达到较好的分类效果, 这样就可以控制模型的大小, 方便模型在终端部署。为此本文以 MobileNetV2^[16] 为基础, 增加了 CA (coordinate attention)^[17] 注意力机制, 构建了一个新的模型用于 CXR 影像新冠肺炎检测任务, 并将其命名为 MBCA-COVIDNET。

2.3 MBCA-COVIDNET 模型结构

该模型的整体结构如图 6 所示。

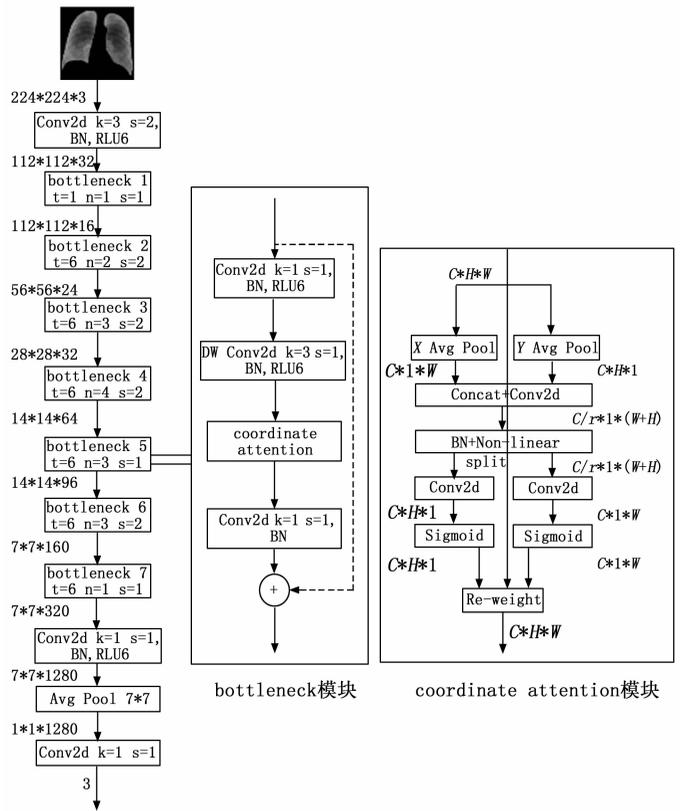


图 6 MBCA-COVIDNET 网络结构示意图

图 6 中, k 代表卷积核大小; 卷积中的 s 代表卷积步距; BN 代表批归一化; RLU6、Non-Linear、Sigmoid 代表非线性激活函数; t 代表隐藏层的扩展因子; n 代表 bottleneck 的重复次数; 图中左侧 MBCA-COVIDNET 中包含 7 个 bottleneck 层, 需要注意的是每个 bottleneck 层中标识出的 s 是对应的第一个 bottleneck 模块中的卷积步距, 其它 bottleneck 模块 s 均为 1; Avg Pool 代表全局平均池化; DW Conv2d 代表深度可分离卷积; C 代表通道数; H 代表特征图的高度; W 代表特征图的宽度; r 为代表特征图的缩放因子。

2.3.1 CA 模块

CA 是将位置信息嵌入到通道注意力中。与通道注意力机制不同, 传统的通道注意力机制是通过二维全局平均池化将每个通道上对应的空间信息 ($H \times W$) 压缩为 1 个具体的数值, 最终维度变为 $1 \times 1 \times C$ 的向量, 这种操作会损失物体的空间信息。而 CA 将通道注意力分解为两个一维特征编码的过程, 分别沿着两个空间方向聚合特征。这种方法可以在一个空间方向上捕获长距离依赖, 同时在另一个空间方向上保留精确的位置信息, 它们可以互补地应用到输入特征图来增强感兴趣的目标表示。具体实现过程如图 6 所示, 首先将输入特征图分别在宽度和高度两个方向进行全局平均池化, 分别获得在宽度和高度两个方向的特征图 z^w 和 z^h , 公式如下所示:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < w} x_c(h, i) \quad (1)$$

式中, W 为特征图的宽度, $x_c(h, i)$ 为第 c 通道在高度 h 处的第 i 个输入特征值, $z_c^h(h)$ 为第 c 通道在高度 h 处的输出。

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (2)$$

式中, H 为特征图的高度, $x_c(j, w)$ 为第 c 通道在宽度 w 处的第 j 个输入特征值, $z_c^w(w)$ 为第 c 通道在宽度 w 处的输出。

然后将特征图 z^w 的最后两个维度调换, 再将其与特征图 z^h 拼接在一起并送入卷积模块, 卷积核为 1×1 , 输出维度降低为原来的 C/r , r 为通道的缩放比率, 并将经过批量归一化处理的特征图 F_1 经过非线性激活函数得到形如 $1 \times (W+H) \times C/r$ 的特征图 f , $f \in R^{C/r \times (H+W)}$, 公式如下所示:

$$f = \delta(F_1([\cdot^h, z^w])) \quad (3)$$

式中, z^h 为高度方向上的特征图, z^w 为宽度方向上的特征图, $[\cdot, \cdot]$ 为在空间维度上的拼接操作, F_1 为对输入特征进行 1×1 卷积并进行批归一化操作, δ 为非线性激活函数。

接着, 沿着空间维度将 f 切分为两个单独的张量 $f^h \in R^{C/r \times H}$ 和 $f^w \in R^{C/r \times W}$ 再利用两个 1×1 卷积将特征图 f^h 和 f^w 变换到和输入 x 同样的通道数。

$$g^h = \sigma(F_h(f^h)) \quad (4)$$

$$g^w = \sigma(F_w(f^w)) \quad (5)$$

式中, f^h 为沿高度方向将 f 切分的特征, f^w 为沿宽度方向将 f 切分的特征, F_h 、 F_w 表示分别对 f^h 和 f^w 进行 1×1 卷积, 并将特征通道数从 C/r 变回为 C , σ 为 sigmoid 函数。

最后对 g^h 和 g^w 进行拓展, 作为注意力权重与输入相乘, CA 模块的最终输出可以表述如下式:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

式中, $g_c^h(i)$ 为沿着高度方向的最终注意力权重, $g_c^w(j)$ 为沿着宽度方向的最终注意力权重, $x_c(i, j)$ 为输入的特征, $y_c(i, j)$ 为输入特征与注意力权重相乘之后的最终输出结果。

2.3.2 改进 MobileNetV2 结构

为了进一步提高 MobileNetV2 模型分类的准确率, 在原模型的每个 bottleneck 模块中加入了 CA 模块, 加入位置如图 6 所示。CA 模块中 BN 层后的非线性激活函数采用 RELU6 激活函数。

2.4 数据预处理和数据增强

图像预处理, 本实验将所有的影像缩放到 224×224 , 数据集中的绝大部分图像都比该尺寸要大, 也有个别图像的尺寸小于 224×224 , 为将图像扩大至 224×224 , 实现输入模型的图像尺寸统一, 采用了最近邻插值。

为了更好地训练模型, 提高模型的精度, 在实验中也尝试了医学影像中常用的数据增强方法, 如直方图均衡化、Gamma 变换、TrivialAugment^[18]等。

1) 直方图均衡化: 直方图表示的是图片灰度值的分布, 直方图均衡化是将原图经过某种变换, 得到一幅灰度直方图为均匀分布的新图像的方法。其基本思想是对在图像中像素个数多的灰度级进行展宽, 而对像素个数少的灰

度级进行缩减, 从而达到清晰图像的目的。

2) Gamma 变换: Gamma 变换是对输入图像灰度值进行非线性操作, 使输出图像灰度值与输入图像灰度值呈指数关系。Gamma 变换就是用来图像增强, 其提升了暗部细节, 通过非线性变换, 让图像从曝光强度的线性响应变得更接近人眼感受的响应, 对过曝或过暗的图像进行矫正。

3) TrivialAugment: TrivialAugment 是一种自动增强策略。它不像 AutoAugment 和 RandAugment 需要搜索空间, 其不需要任何搜索, 整个方法非常简单, 每次随机选择一个图像增强操作, 然后随机确定它的增强幅度, 并对图像进行增强。TrivialAugment 的图像增强集合和 RandAugment 基本一样, 只不过其定义了一套更宽的增强幅度, 目前 torchvision 中已经实现了 TrivialAugmentWide。

2.5 模型评估指标

针对胸片的三分类任务, 采用如下 7 个评估指标对模型进行评估: 正确率 (Accuracy)、精度 (Precision)、灵敏度 (Sensitivity)、F1Score、特异度 (Specificity)、参数量 (Parameters)、计算量 (MACs)。

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

Accuracy 为所有被分类正确的影像数量与所有影像数量的比值。

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

Precision 为被分类为正例的影像中, 实际为正例的比例。

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

Sensitivity 为影像中所有正例被分类正确的比例, 用来衡量模型对正例的识别能力。

$$F1 \text{ Score} = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \quad (10)$$

F1 Score 兼顾了分类模型的精确率和召回率, 可以看作是模型精确率和召回率的一种调和平均, 它的最大值是 1, 最小值是 0。

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

Specificity 为影像中所有的负例被分类正确的比例, 衡量模型对负例的识别能力。

公式 (7) ~ (11) 中 TP 表示实际上是正例, 预测结果也是正例的数量; TN 表示实际上是负例, 预测结果也是负例的数量; FP 表示实际上是负例, 预测结果为正例的数量; FN 表示实际上是正例, 预测结果为负例的数量。

Parameters 为模型内部总的参数数量, 用来衡量模型的大小; MACs 为乘加累积次数, 1MACs 等于 1 个乘法和 1 个加法。

3 实验结果与分析

3.1 实验步骤

本研究所有实验均基于 Python3.9.5 和 pytorch1.11.0

的深度学习的框架实现，具体硬件设备配置：显卡为 NVIDIA-AGeForce RTX 3080 Laptop GPU，显存 16 G，电脑内存 32 G，CPU 为 AMD Ryzen 9 5900HX。

针对 COVID-QU-Ex Dataset 数据集的影像分类任务，对比了如下几个模型，RestNet 系列 (RestNet18、RestNet34、RestNet50、RestNet101)、EfficientNet 系列 (EfficientNetB0、EfficientNetB3)、MobileNetV2、MobileNetV3、SwinTransformer、ConvNext_small 以及新设计的 MBCA-COVIDNET 模型。所有这些模型训练的参数设置如下：训练迭代次数 (epoch) 30，因为 30 次模型已经收敛，再继续训练，模型精度没有继续提升；batchsize 为 64，因为受限于实验硬件设备所限，该大小可使得所有基线模型在当前 GPU (16 G) 下训练；初始学习率为 5×10^{-4} ，由于所有基线模型都采用迁移学习，都用到在 ImageNet1K 上的预训练权重，

所以初始学习率不易过大；优化器选用效果较好的 AdamW，weight_decay 设置为 5×10^{-2} ，AdamW 是在 Adam 的基础上引入了 L2 正则化，可有效地减小过拟合，针对该训练数据集，实验证明 AdamW 要优于 SGD；学习率下降策略采用 warmup+Cosine 的下降策略，热身训练为 1 个 epoch。实验证明该下降策略相比等间隔调整学习率 (StepLR) 能够带来精度的提升。同时所有模型训练时均不采用任何图像增强技术，只是将数据集中的图像大小缩放为 224×224 作为模型输入。

3.2 实验结果

基于上述统一设置的训练参数，训练得到的 10 个神经网络模型在 COVID-QU-Ex Dataset 测试集上的性能表现如表 2 所示，各模型的混淆矩阵如图 7 所示。从表 2 中可以看出 MBCA-COVIDNET 的模型的正确率为 97.02%，是所有

表 2 各深度学习模型在 COVID-QU-Ex 数据集上的性能对比

Model	Class	正确率/%	精确度/%	灵敏度/%	特异度/%	F1-score/%	Parameters/M	MACs/G
RestNet18	COVID-19	96.48	99.45	98.5	99.7	98.97	11.18	1.82
	Non-COVID		95.06	95.74	97.53	95.4		
	Normal		94.69	95	97.55	94.84		
RestNet34	COVID-19	95.96	99.08	98.87	99.5	98.97	21.29	3.68
	Non-COVID		93.49	95.56	96.69	94.51		
	Normal		95.13	93.13	97.81	94.12		
RestNet50	COVID-19	95.98	99.16	98.71	99.54	98.93	23.51	4.12
	Non-COVID		93.33	95.65	96.6	94.48		
	Normal		95.27	93.27	97.87	94.26		
RestNet101	COVID-19	96.15	99.37	98.66	99.66	99.01	42.51	7.85
	Non-COVID		94.01	95.47	96.98	94.73		
	Normal		94.86	94.07	97.65	94.46		
EfficientNetB0	COVID-19	96.41	99.58	98.62	99.77	99.1	4.01	0.40
	Non-COVID		93.97	96.76	96.91	95.34		
	Normal		95.52	93.55	97.98	94.52		
EfficientNetB3	COVID-19	96.95	99.71	98.83	99.84	99.27	10.7	0.99
	Non-COVID		95.72	96.18	97.86	95.95		
	Normal		95.21	95.65	97.78	95.43		
MobileNetV2	COVID-19	96.55	99.25	99.08	99.59	99.16	2.23	0.32
	Non-COVID		94.91	95.92	97.44	95.41		
	Normal		95.28	94.39	97.85	94.83		
MobileNetV3 large	COVID-19	96.54	99.41	99.04	99.68	99.22	4.21	0.23
	Non-COVID		94.15	96.49	97.02	95.31		
	Normal		95.89	93.79	98.15	94.83		
SwinTransformer_small	COVID-19	96.20	98.83	98.33	99.36	98.58	48.84	8.52
	Non-COVID		94.84	95.47	97.42	95.15		
	Normal		94.71	94.58	97.57	94.64		
ConvNext_small	COVID-19	96.99	99.18	99.33	99.04	99.64	49.46	8.7
	Non-COVID		95.31	96.54	97.64	95.92		
	Normal		96.18	95.19	98.26	95.68		
MBCA-COVIDNET	COVID-19	97.02	99.41	99.21	99.68	99.31	2.67	0.33
	Non-COVID		96.13	95.87	98.08	96.0		
	Normal		95.30	95.79	97.83	95.54		

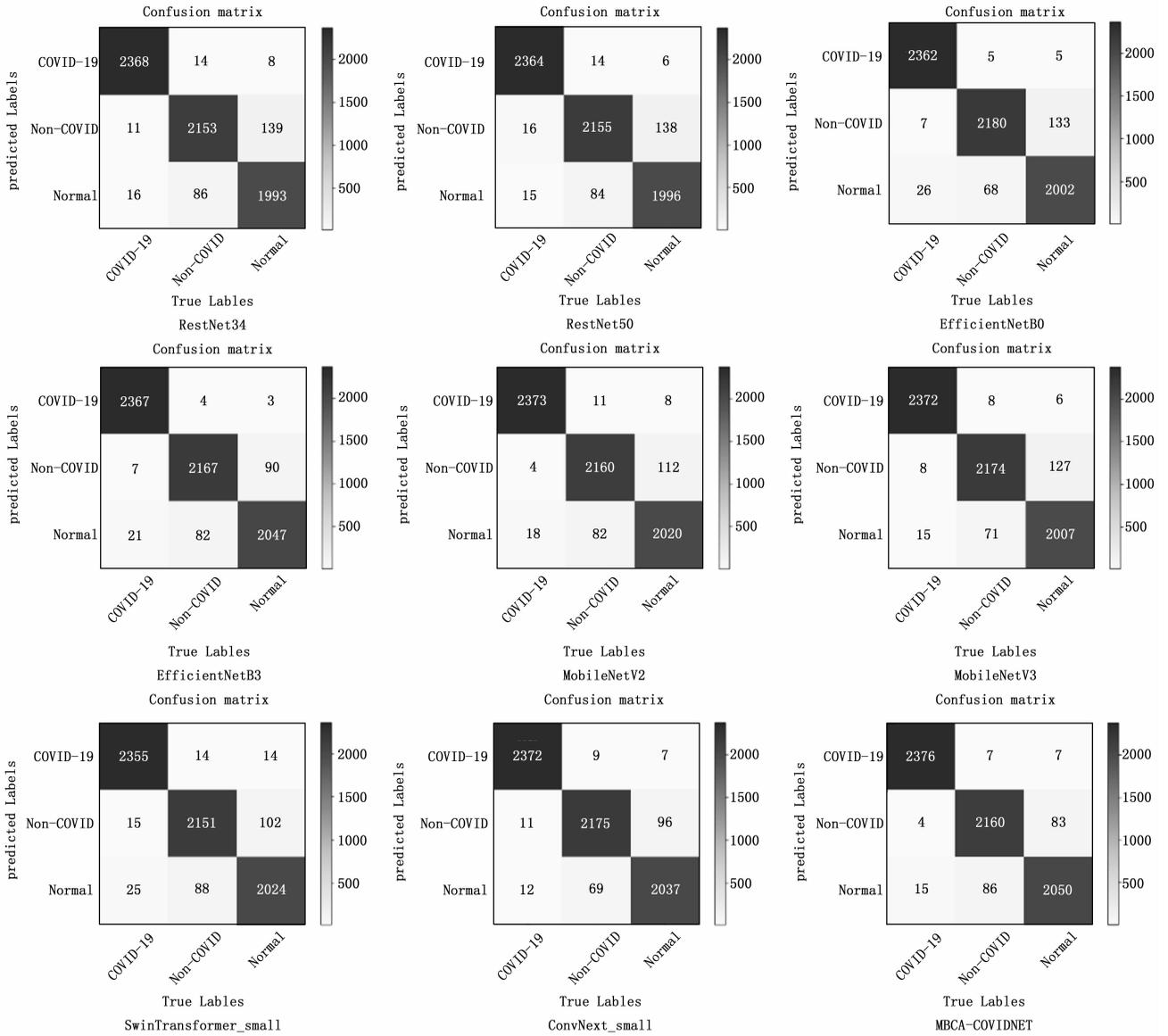


图 7 各深度学习模型在 COVID-QU-Ex 数据集上的混淆矩阵

模型最高的, ConvNext_small 模型的正确率为 96.99%, 仅从正确率这个指标看前者只比后者仅提升了 0.03%。但是对比模型的参数量可以发现 MBCA-COVIDNET 模型的参数量仅为 2.67 M, 而 ConvNext_small 模型的参数量为 49.46 M, 前者的参数量远小于后者。对比 MACs 指标, 前者为 0.33 G, 后者为 8.7 G, 前者的计算量也远小于后者。

对比 MBCA-COVIDNET 模型和 MoblieNetV2 模型, 可以发现正确率前者比后者提升了 0.47%, 参数量前者仅比后者多 0.44 M, 计算量前者比后者多 0.01。同时也证明了 MoblieNetV2 加入 CA 注意力机制的有效性。

通过观察 MBCA-COVIDNE 模型在测试集上的混淆矩阵, 可以发现, 2 395 例 COVID-19 测试影像中正确预测 2 376 例, 4 例预测为 Non-COVID 类别, 15 例预测为 Normal 类别。预测正确率为 99.41%, 灵敏度为 99.21%, 特异度为 99.68%, F1Score 为 99.31%; 2 253 例 Non-COVID

ID 测试影像中正确预测 2 160 例, 7 例预测为 COVID-19 类别, 86 例预测为 Normal 类别。预测正确率为 96.13%, 灵敏度为 95.87%, 特异度为 98.08%, F1Score 为 96.0%; 2 140 例 Normal 测试影像中正确预测 2 050 例, 7 例预测为 COVID-19 类别, 83 例预测为 Non-COVID 类别。预测正确率为 95.30%, 灵敏度为 95.79%, 特异度为 97.83%, F1Score 为 95.54%。实验证明该模型对 COVID-19 影像分类正确率要高于其它 2 个类别。而其它 2 个类别的影像分类正确率低的原因, 主要是模型将部分 Non-COVID 类别影像预测为 Normal 类别, 将部分 Normal 类别影像预测为 Non-COVID 类别造成的。

3.3 消融实验

为了更好地训练一个轻量化的模型, 本研究做了如下 3 个消融实验。

- 1) 针对 CA 模块中的非线性激活函数, 对比了 RELU、

RELU6、SiLU、LeakyReLU、Mish 5 个激活函数，结果表明 RELU6 激活函数效果更好一些，具体情况详见表 3。

表 3 CA 模块中不同激活函数的性能对比

编号	激活函数	正确率 / %
1	RELU6	97.02
2	Mish	96.94
3	RELU	96.83
4	LeakyReLU	96.70
5	SiLU	96.69

2) 为了验证在 COVID-QU-Ex 数据集上 CA 注意力模块优于其它注意力模块，以 MobileNetV2 为基础，对比了加入 SE 模块、CBAM 模块以及 CA 模块后的效果，结果表明 MobileNetV2 + CA 的正确率要优于 MobileNetV2 + CBAM 和 MobileNetV2 + SE，而且针对该数据集 MobileNetV2 + CBAM 和 MobileNetV2 的正确率差别不大，具体情况如表 4 所示。

表 4 MobileNetV2 中加入不同注意力模块的性能对比

编号	模型	正确率 / %
1	MobileNetV2	96.55
2	MobileNetV2 + SE	96.83
3	MobileNetV2 + CBAM	96.54
4	MobileNetV2 + CA	97.02

3) 本文也尝试了两种不同的数据增强策略，第一种用到了 Albumentations^[19] 库中 RandomGamma、RandomBrightnessContrast、CLAHE、Blur、MotionBlur、MedianBlur、HorizontalFlip、ShiftScaleRotate 数据增强方法的组合；第二种用的是 TrivialAugment。同时训练模型的迭代次数也从之前的 30 次增加到了 50 次，使得模型更好的收敛。实验结果表 TrivialAugment 数据增强的效果更优，详见表 5。

表 5 不同数据增强策略的性能对比

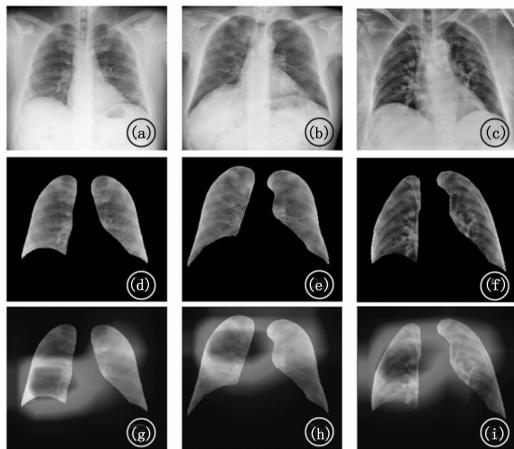
编号	数据增强策略	正确率 / %
1	Albumentations 中多种方法的组合	97.44
2	TrivialAugment	97.98

3.4 类激活图可视化

由于神经网络模型需要大量的数据来训练，是以数据驱动的方式创建的，因此通常认为神经网络是一个“黑盒”，其做出的分类缺乏可解释性。为了对 MBCA-COVID-NET 模型分类决策进行解释，本研究利用 Grad-CAM^[20] 技术，在胸片上进行了类激活图可视化，如图 8 所示。从图中可以看出该模型关注的影像区域较为合理，覆盖了肺部感染区域，能够辅助医生发现病灶，有助于医生对患者的诊断和治疗。

4 结束语

本文利用 MobileNet 网络结构并与 CA 注意力模块相结



(a) ~ (c) 为数据集中新冠肺炎患者原始胸部 X 光片样例；(d) ~ (f) 为新冠肺炎患者肺部边界区域；(g) ~ (i) 为新冠肺炎患者类激活图。

图 8 MBCA-COVIDNET 新冠影像类激活图

合，构建了一个针对胸片的 COVID-19 检测模型 MBCA-COVIDNET。该模型在 COVID-QU-Ex Dataset 数据集上的三分类任务中，采用了 TrivialAugment 图像预处理、迁移学习以及一系列训练技巧，最终在测试集上取得了 97.98% 的准确率，该模型比之前较先进的 COVIDLite 模型提升了 0.8 个百分点，同时该模型的参数量和计算量也均处于较低水平，易于在各种硬件上进行部署。

为了更好地演示该模型，利用 Hugging Face^[21] 开源软件开发了一个 COVID-19 智能辅助诊断演示系统。该开源软件可以非常容易的将训练好的模型托管到 Hugging Face Spaces 中，只需编写一个 app.py 文件，就可以很方便地随时随地进行系统的演示，软件界面如图 9 所示。



图 9 Hugging Face 系统界面

同时，为了使训练好的模型能够真正地应用于临床实践，而不是停留在实验室阶段，利用 Flask 开发了一个 COVID-19 的 web 应用程序，Flask 是一个使用 Python 编写的轻量级 Web 应用框架。该系统能够辅助医院工作人员，快速的判断 CXR 影像的类别。具体操作步骤如下：1) 点击选择文件按钮，加载一张待检测的影像；2) 点击预测按钮，系统会使用训练好的模型，对该影像进行预测，并且给出预测为各个类别的概率。经过实际测试该系统在普通的笔记本电脑上就可以部署并且能够流畅运行，具有极佳的用户体验。系统界面如图 10 所示。

本研究的局限性：1) 训练好的模型未能在更多的数据集上做测试，特别是当前新冠病毒进行了多次变异，针对



图 10 COVID-19 深度学习检测系统界面

感染变异后新冠病毒患者的胸部 X 光片, 该模型的泛化能力有待进一步验证; 2) 训练模型的数据集来源单一, 没能做到多中心; 3) 训练模型完全依赖于胸部 X 光片未能结合患者的临床相关数据; 4) 该模型最终功能只实现对胸部 X 光片进行新冠肺炎、其它肺炎、正常三分类任务, 未能指出新冠肺炎感染的严重程度。

下一步研究工作可在现有研究的基础上, 收集、标注新冠肺炎感染区域并训练一个感染区域分割模型, 从而实现利用胸部 X 光片对新冠肺炎感染严重程度进行量化分析。

参考文献:

- [1] WANG W, XU Y, GAO R, et al. Detection of SARS-CoV-2 in different types of clinical specimens [J]. *Jama*, 2020, 323 (18): 1843–1844.
- [2] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis [J]. *Medical Image Analysis*, 2017, 42: 60–88.
- [3] NARIN A, KAYA C, PAMUK Z. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks [J]. *Pattern Analysis and Applications*, 2021, 24 (3): 1207–1220.
- [4] WANG G, LIU X, SHEN J, et al. A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest X-ray images [J]. *Nature Biomedical Engineering*, 2021, 5 (6): 509–521.
- [5] WANG L, LIN Z Q, WONG A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images [J]. *Scientific Reports*, 2020, 10 (1): 1–12.
- [6] SIDDHARTHA M, SANTRA A. COVIDLite: A depth-wise separable deep neural network with white balance and CLAHE for detection of COVID-19 [EB/OL]. *arXiv preprint arXiv*, 2020-6-19. <https://arxiv.org/abs/2006.13873>.
- [7] OH Y, PARK S, YE J C. Deep learning COVID-19 features on CXR using limited training data sets [J]. *IEEE Transactions on Medical Imaging*, 2020, 39 (8): 2688–2700.
- [8] TAHIR A M, CHOWDHURY M E H, KHANDAKAR A, et al. COVID-19 infection localization and severity grading from chest X-ray images [J/OL]. *Computers in Biology and Medi-*

cine [2021-12-10]. <https://doi.org/10.1016/j.compbio.2021.105002>.

- [9] TAHIR A M, CHOWDHURY M E H, QIBLAWEY Y, et al. COVID-QU-Ex dataSet [DB/OL]. (2021-07-08) [2022-11-20]. <https://doi.org/10.34740/kaggle/dsv/3122958>.
- [10] 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/OL]. *Computers in Biology and Medicine*, 2021 (5). <https://doi.org/10.1016/j.compbio.2021.104319>.
- [11] DEGERLI A, AHISHALI M, YAMAC M, et al. COVID-19 infection map generation and detection from chest X-ray images [J]. *Health Information Science and Systems*, 2021, 9 (1): 1–16.
- [12] CHOWDHURY M E H, RAHMAN T, KHANDAKAR A, et al. Can AI help in screening viral and COVID-19 pneumonia? [J]. *IEEE Access*, 2020, 8: 132665–132676.
- [13] KERMANY D, ZHANG K, GOLDBAUM M. Large dataset of labeled optical coherence tomography (oct) and chest x-ray images [J]. *Mendeley Data*, 2018, 3: 10–17632.
- [14] COHEN J P, MORRISON P, DAO L, et al. COVID-19 image data collection: prospective predictions are the future [EB/OL]. *arXiv preprint arXiv* [2020-06-22]. <https://arxiv.org/abs/2006.13873>.
- [15] CHEN Z H. Mask-RCNN detection of COVID-19 pneumonia symptoms by employing stacked autoencoders in deep unsupervised learning on low-dose high resolution CT [J/OL]. *IEEE Dataport* [2020-10-01]. <http://dx.doi.org/10.21227/4kcm-m312>.
- [16] 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*, 2018: 4510–4520.
- [17] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13713–13722.
- [18] MÜLLER S G, HUTTER F. TrivialAugment: Tuning-free yet state-of-the-art data augmentation [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 774–782.
- [19] BUSLAEV A, IGLOVIKOV V I, KHVEDCHENYA E, et al. Albumentations: fast and flexible image augmentations [J]. *Information*, 2020, 11 (2): 125.
- [20] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization [C] // *Proceedings-16th IEEE International Conference on Computer Vision (ICCV 2017)*, Venice, Italy: IEEE Computer Society, 2017: 618–626.
- [21] FACE H. The AI community building the future [DB/OL]. (2022-06-26) [2022-11-23]. <https://github.com/huggingface>.