

基于轻量化注意力残差网络的食管鳞癌识别方法

王金铭^{1,2}, 李鹏², 梁燕³, 孙玮^{1,2}, 宋杰³, 冯亚东³, 赵凌霄^{2*}¹中国科学技术大学生命科学与医学部生物医学工程学院, 江苏 苏州 215163;²中国科学院苏州生物医学工程技术研究所, 江苏 苏州 215163;³东南大学附属中大医院消化内科, 江苏 南京 210009

摘要 食管鳞癌(ESCC)是我国常见的消化道恶性肿瘤之一。临床上,窄带成像联合放大内镜(NBI-ME)能够显示出食管粘膜层的微血管形态变化,是诊断ESCC的重要手段。针对ESCC识别模型难以兼顾识别准确率和推理效率的问题,提出一种融合注意力机制的轻量化残差网络(CALite-ResNet)对食管NBI-ME图像进行分类。从多家医院采集到206例患者共11468张NBI-ME图像作为本研究数据集。实验结果表明,ESCC识别的准确率和敏感度分别在图像级别达到96.39%和95.70%,在病人级别达到95.70%和94.62%,单张食管图像的平均预测时间为16.42 ms。因此,CALite-ResNet模型对ESCC具有较高的识别准确率和较快的推理效率,能够为ESCC的临床辅助诊断提供有效帮助,具备一定的临床意义与应用价值。

关键词 图像处理;轻量化网络;注意力机制;食管鳞癌;窄带成像;放大内镜

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP220856

Esophageal Squamous Cell Carcinoma Recognition Based on Lightweight Residual Networks with an Attention Mechanism

Wang Jinming^{1,2}, Li Peng², Liang Yan³, Sun Wei^{1,2}, Song Jie³, Feng Yadong³, Zhao Lingxiao^{2*}¹*School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China, Suzhou 215163, Jiangsu, China;*²*Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Science, Suzhou 215163, Jiangsu, China;*³*Department of Gastroenterology, Zhongda Hospital, Southeast University, Nanjing 210009, Jiangsu, China*

Abstract Esophageal squamous cell carcinoma (ESCC) is one of the most common malignant digestive tract tumors in China. Clinically, narrowband imaging combined with magnifying endoscopy (NBI-ME) can be used to investigate the morphological changes of microvessels in the esophageal mucosa and serves as an important means of diagnosing ESCC. To solve the ESCC recognition model's difficulty in considering both the recognition accuracy and reasoning efficiency, a lightweight residual network (CALite-ResNet) with an integrated attention mechanism is proposed to classify esophageal NBI-ME images. The dataset for this study comprises 11468 NBI-ME images of 206 patients collected from multiple hospitals. The experimental results show that the accuracy and sensitivity of the ESCC recognition is 96.39% and 95.70% at the image level, and 95.70% and 94.62% at the patient level, respectively, and the average prediction time of a single esophageal image is 16.42 ms. Therefore, the CALite-ResNet model has a higher recognition accuracy and faster reasoning efficiency for ESCC recognition, and a certain clinical significance and application value, thereby making it effective for use in the auxiliary clinical diagnosis of ESCC.

Key words image processing; lightweight network; attention mechanism; esophageal squamous cell carcinoma; narrow band imaging; magnifying endoscopy

收稿日期: 2022-03-02; 修回日期: 2022-04-18; 录用日期: 2022-05-05; 网络首发日期: 2022-05-15

基金项目: 江苏省重点研发计划项目(BE2019710)、苏州市科技计划项目(SYS2019008)、常州市科技计划项目(CE20195001)

通信作者: hitic@sibet.ac.cn

1 引言

食管癌是发病率居于全球第 7 位、死亡率居于全球第 6 位的消化道恶性肿瘤,如何降低它的发病率和死亡率已成为全球亟待解决的问题^[1]。在我国,食管癌的发生以食管鳞癌(ESCC)为主,其发病率占食管癌总发病率的 90% 以上^[2]。ESCC 患者的预后与诊断时的肿瘤分期密切相关^[3-4]。开展 ESCC 的筛查及早诊早治是目前提高 ESCC 治疗效果和治愈率的有效途径。消化内镜检查是诊断与明确早期 ESCC 的最主要方式。常规白光内镜下,早期 ESCC 及癌前病变起病隐匿,易漏诊。然而,窄带成像联合放大内镜(NBI-ME)可以更清楚地显示食管黏膜层的微血管形态结构,有助于早期 ESCC 的精准诊断^[5]。临床实践中早期 ESCC 的诊断结果极度依赖于内镜医师的主观判断和技能经验,对于缺乏经验的医师来说极易造成误诊漏诊,同时高强度的阅片工作会严重影响诊断效率。

近年来,人工智能(AI)技术的蓬勃发展为 ESCC 的辅助诊断提供了可能性^[6-9]。传统的 AI 辅助诊断大多基于特征工程和模式识别技术^[10-11]。随着深度学习技术的快速发展和日益成熟,越来越多基于深度学习的辅助诊断模型被应用于 NBI-ME 下 ESCC 的筛查和诊断,并取得了优异的成果^[12]。Zhang 等^[13]采用卷积神经网络(CNN)的监督学习方式对 NBI-ME 图像进行早期食管癌检测,证实了 AI 辅助诊断技术应用于食管早癌筛查的可行性。Xue 等^[14]利用 CNN 模型提取 NBI-ME 图像的特征信息并使用支持向量机(SVM)作为分类器进行 ESCC 微血管的分型与识别。其后,该团队又设计了一种基于全卷积神经网络的 ESCC 微血管分型识别算法,平均交并比(MIoU)达到了 77.8%^[15]。Ohmori 等^[16]使用较大数据集训练 AI 辅助诊断系统来识别 ESCC,其诊断性能与内镜专家无显著差异。Nakagawa 等^[17]建立了一个基于 Single Shot MultiBox Detector(SSD)模型的 ESCC 浸润深度预测系统,其诊断性能可与专家媲美。Everson 等^[18]提出一种结合深度监督和嵌入式类激活图的深度学习方法,利用 7046 张 NEI-ME 图像训练网络模型,对 ESCC 的识别准确率可达 93.7%。Guo 等^[19]专门开发了一种基于 SegNet 架构的 Computer Aided Diagnosis(CAD)系统来自动检测与识别 ESCC 病变区域,该系统在临床环境中可以实时应用。腾讯研发的基于 AI 的食管癌辅助筛查系统^[20]能够分析和识别内窥镜图像,协助临床医师判断食管病变位置,识别准确率可达 90%。上述研究工作表明,深度学习技术在 ESCC 的辅助诊断上具有很大潜力。

目前,基于深度学习的 ESCC 识别系统仍存在难以兼顾准确率和效率的问题,且此类系统的训练依赖大量数据。由于医疗数据获取与标注的时间、人力成本较高,则充分利用有限数据资源成为解决问题的关

键所在。为此,本文提出一种融合注意力机制的轻量化残差网络(CALite-ResNet)。相较于传统 CNN, CALite-ResNet 能够更好地兼顾 ESCC 识别的准确率与推理效率,同时这种轻量化网络参数量较小,训练时所需的数据也较少。主要研究工作如下:1)借鉴深度残差网络(ResNet50^[21])的主干结构,使用 GhostModule^[22]和自校准卷积(SCConv)^[23]设计出轻量化残差结构;2)添加 Coordinate Attention(CA)注意力机制^[24]自适应细化提取特征,在不影响实效性的前提下提高网络对关键特征的提取筛选能力;3)采用 Class-Balanced(CB)损失函数^[25]平衡样本类别权重,提高对 ESCC 识别的敏感度;4)在多中心数据集上完成模型评估和验证工作。

2 方法设计

2.1 网络模型设计

CALite-ResNet 的主干以 ResNet50 为基准,具体结构如图 1 所示。其中,GSCBottleneck 轻量化残差模块由 GhostModule 和 SCConv 操作构成,以减小模型参数量、提升推理效率。引入 CA 注意力机制构建 CA-GSCBottleneck 模块来增强模型对特征的学习与表达能力,提高模型的识别精度。用这两个模块替代 ResNet50 的原始残差结构,可以在不损失模型精度的情况下减小模型参数量,以降低硬件资源要求,利于模型推广应用到临床实践中。

2.1.1 轻量化残差模块

一般来说,由传统卷积方式组成的 CNN 计算成本高,且产生的特征图会有较多冗余情况。GhostModule^[22]可以替代部分卷积,实现轻量且高效的网络模型。图 2 显示了 GhostModule 的结构。GhostModule 先通过 1×1 卷积操作生成本质特征图,再使用深度可分离卷积对本质特征图进行线性变换来获得冗余特征图,最后将这两种特征图进行级联操作,得到与传统卷积层输出相同尺寸和通道大小的特征图。

SCConv^[23]改进了传统的卷积特征转换过程,通过内部通信显著扩展每个卷积层的感受野,丰富输出特征的多样性。SCConv 不引入额外的参数和复杂性,但在实际使用中仍会延长算法的推理时间。为保证 ESCC 识别的高效性,使用 GhostModule 替换 SCConv 一个分支中的 3×3 卷积,减小参数量,使其更轻量化。改进后的 SCConv 结构如图 3 所示。

具体流程如下:按通道维度, $C \times H \times W$ 大小的特征图 \mathbf{X} 被拆分为 $\{\mathbf{X}_1, \mathbf{X}_2\}$ 并分别传入不同支路中。 \mathbf{X}_1 所在支路首先进行一次下采样(Down)操作,其次经过 Γ_1 卷积操作得到特征作为自校准参考,并对其进行上采样(Up)操作将尺寸由 $C/2 \times H/4 \times W/4$ 恢复为 $C/2 \times H \times W$,再将该特征图与原特征图 \mathbf{X}_1 相加,

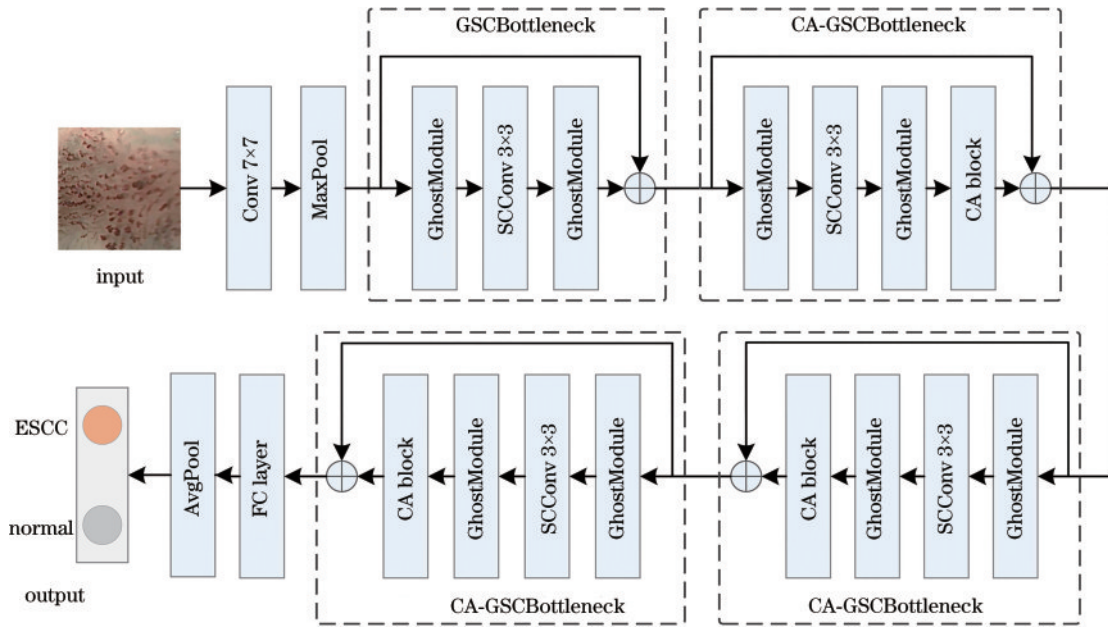


图1 CALite-ResNet结构

Fig. 1 Architecture of CALite-ResNet

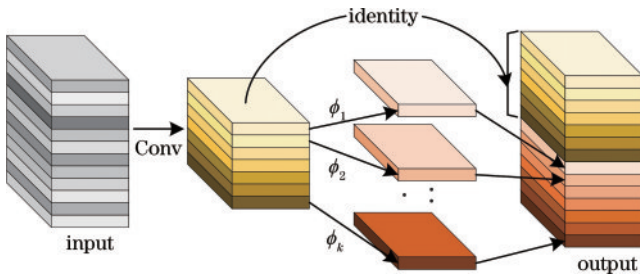


图2 GhostModule结构示意图

Fig. 2 Schematic of GhostModule

激活后通过矩阵乘积来校准特征 $\Gamma_2(X_1)$, 经 Γ_3 卷积操作输出特征 Y_1 。 X_2 所在支路通过一个 3×3 卷积常规操作 Γ_4 输出特征 Y_2 。 将此两路输出特征 Y_1, Y_2 在通道维度上级联合并最终输出 Y 。 计算过程如下:

$$Y_1 = \Gamma_3 \left\{ \sigma \left\{ \text{Up} \left\{ \Gamma_1 \left[\text{Down} (X_1) \right] \right\} + X_1 \right\} \cdot \Gamma_2 (X_1) \right\}, \quad (1)$$

$$Y_2 = \Gamma_4 (X_2), \quad (2)$$

$$Y = Y_1 + Y_2, \quad (3)$$

式中: σ 为 Sigmoid 激活函数。 在不改变模型结构的前

提下, SCConv 通过对输入特征分两路卷积、池化采样和校准处理等加强了特征图的内在通信, 有效扩增了卷积层的感受野。 这样有助于提高模型对 ESCC 的识别精度。

为降低参数量、增大网络模型的感受野, 使用 GhostModule 代替 ResNet50 中 Bottleneck 的 1×1 卷积, 使用上述改进后的 SCConv 替换 ResNet50 模型中 Bottleneck 的 3×3 卷积, 搭建出 GSCBottleneck 轻量化残差结构, 具体如图 4 所示。 这种结构设计可以达到既缩减模型参数量又不损失模型精度的效果。

2.1.2 注意力模块

深度学习中注意力机制的核心目的是从众多信息中选择对当前任务目标更关键的信息, 已被证实有助于各种计算机视觉任务。 Hou 等^[24] 提出一种新颖的高效 CA 注意力机制。 CA 注意力机制将位置信息嵌入通道注意力中, 使网络能够在更大的区域上捕获有效特征, 还可以避免产生大量的计算开销, 其结构如图 5 所示。

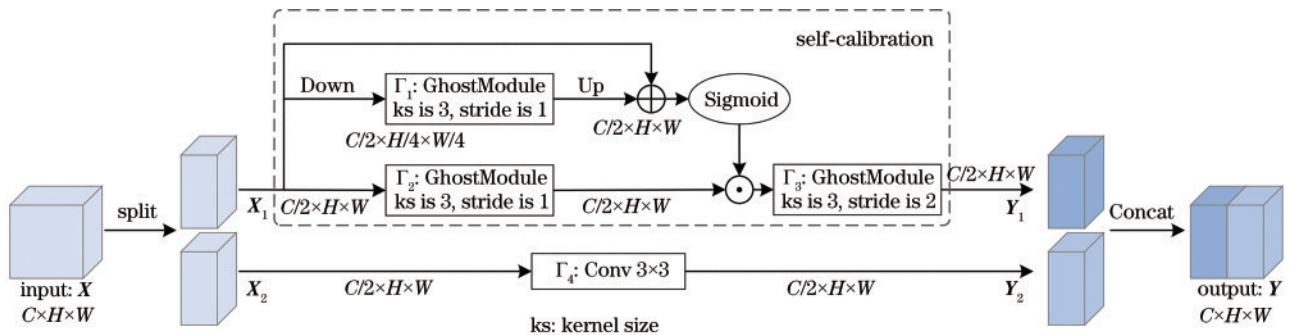


图3 改进后的 SCConv 结构示意图

Fig. 3 Schematic illustration of improved SCConv

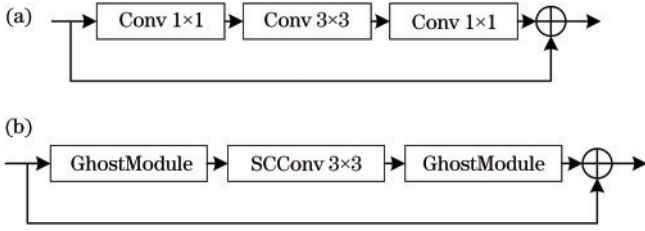


图 4 残差模块结构图。(a) ResNet50 的 Bottleneck 结构；
(b) GSCBottleneck 结构

Fig. 4 Diagrams of residual block structures. (a) Bottleneck structure of ResNet50; (b) GSCBottleneck structure

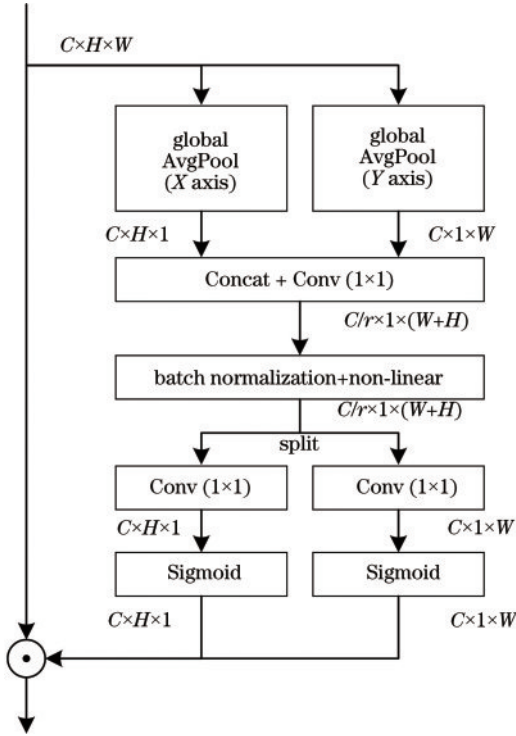


图 5 CA 注意力机制结构
Fig. 5 Structure of CA attention mechanism

CA 模块作为一个计算单元,可以将任何中间张量 $\mathbf{X} = [x_1, x_2, \dots, x_c] \in \mathbf{R}^{C \times H \times W}$ 作为输入并输出一个同样尺寸且有着增强表示能力的输出 $\mathbf{Y} = [y_1, y_2, \dots, y_c]$ 。首先,对输入的 \mathbf{X} 使用尺寸为 $(H, 1)$ 和 $(1, W)$ 的池化核沿着水平坐标方向和垂直坐标方向对每个通道 c 进行编码:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i), \quad (4)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w). \quad (5)$$

上述两个公式聚合了高度 h 和宽度 w 方向上的特征信息,产生了一对方向感知特征图。其次,级联之前生成的两个特征图,通过共享的 1×1 卷积层进行 F_1 变换:

$$\mathbf{f} = \delta \{ F_1 [z^h, z^w] \}, \quad (6)$$

式中: $[\cdot, \cdot]$ 表示沿空间维度的级联运算; δ 表示非线性激活函数。生成的 $\mathbf{f} \in \mathbf{R}^{C/r \times (H+W)}$ 是对空间信息在水平方向和垂直方向的中间特征图,这里的 r 表示控制模块大小的下采样比例。

然后,沿空间维度将 \mathbf{f} 分成两个单独的张量 $\mathbf{f}^h \in \mathbf{R}^{C/r \times H}$ 和 $\mathbf{f}^w \in \mathbf{R}^{C/r \times W}$,再利用两个 1×1 卷积 F_h 与 F_w 将特征图 \mathbf{f}^h 和 \mathbf{f}^w 变换到与输入 \mathbf{X} 同样的通道数。最后,由 Sigmoid 函数激活后生成两个方向上的权重并重新加权输入的原始特征:

$$\mathbf{g}^h = \sigma [F_h(\mathbf{f}^h)], \quad (7)$$

$$\mathbf{g}^w = \sigma [F_w(\mathbf{f}^w)], \quad (8)$$

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j). \quad (9)$$

CA 注意力机制能够捕获跨通道信息、方向感知和位置感知信息,有助于模型更精准地定位和识别感兴趣目标。它灵活且轻量,很容易插入网络模型中。为了使网络能够根据不同区域特征的特点自适应分配权重,且避免产生较多额外的计算量,向 GSCBottleneck 模块中加入 CA 注意力机制,构建出 CA-GSCBottleneck 残差模块,具体如图 6 所示。该残差模块可提高模型对有效特征的提取能力。

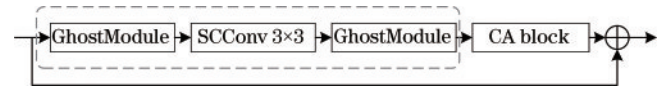


图 6 CA-GSCBottleneck 模块结构
Fig. 6 Structure of CA-GSCBottleneck module

2.2 损失函数

对 ESCC 识别的研究工作本质上是一种食管 NBI-ME 图像的二分类方法研究。在这项任务中,ESCC 的识别敏感度相比于准确率而言具有更重要的意义,对于提高 ESCC 的检出率至关重要。所用的实验数据集存在类别不平衡问题,导致二分类任务中模型准确率降低,故采用 CB 损失函数^[25]来优化模型参数。该函数基于每种类别有效样本数的加权策略来重新平衡损失,从而得到类别平衡的损失函数。本实验使用的 CB 损失函数是基于 Softmax Cross-Entropy 损失函数实现的,其公式如下:

$$L_{CB}(\mathbf{Z}, \mathbf{y}) = -\frac{1 - n_y}{N(n_y - 1)} \log \left[\frac{\exp(z_y)}{\sum_{j=1}^C z_j} \right], \quad (10)$$

式中: \mathbf{Z} 表示模型预测的各个类别的概率值 $\mathbf{Z} = [z_1, z_2, \dots, z_C]^T$, C 表示类别数; N 表示所有类别的有效样本数量; n_y 表示类别 y 的样本数量; $\frac{1 - n_y}{N(n_y - 1)}$ 表示

权衡类别的权重因子。

2.3 多数投票原则

每张食管 NBI-ME 图像的分类结果均可以通过

CALite-ResNet 模型预测,且所采集的同一病例下的所有样本图像均为同一所属类别。因此,为了评估病人个体的识别结果,采用多数投票原则对单个病人的多张图像样本预测结果进行融合。统计同一病例的所有样本图像的预测结果,选择该病例样本图像中输出最多的类别作为该患者的最终识别结果。原理如图 7 所示。

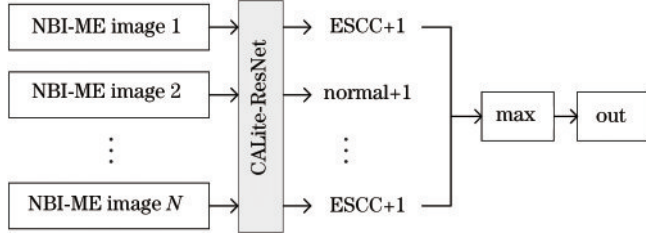


图 7 多数投票方法的原理示意

Fig. 7 Schematic of majority voting method

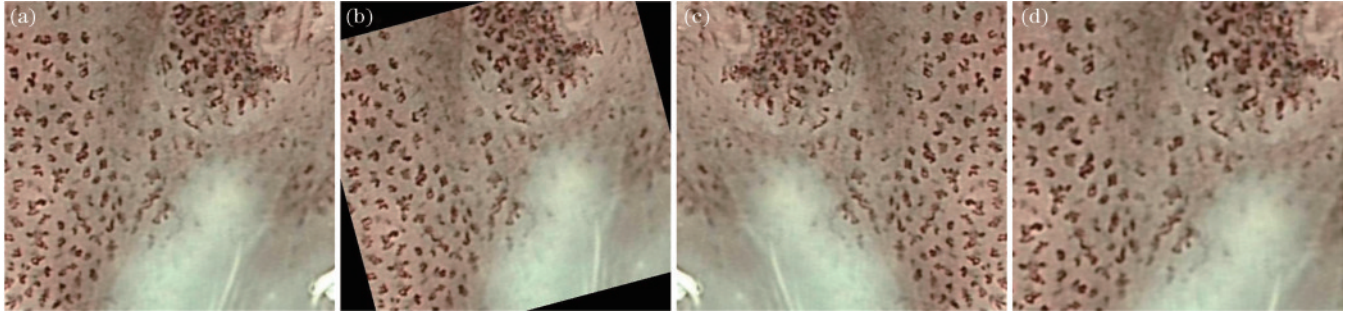


图 8 不同数据增强方式的效果示例图。(a)原始图片;(b)随机旋转;(c)水平翻转;(d)随机缩放

Fig. 8 Example diagrams of effects obtained using different data enhancement methods. (a) Origin image; (b) random rotation; (c) horizontal flip; (d) random scaling

3.2 实验细节

实验环境:64位Ubuntu 16.04操作系统,16 GB 显存的 Nvidia Tesla V100 显卡,Python 3.6 和 PyTorch 1.2.0 软件平台。

模型参数利用文献[26]中的方法实现初始化。原始图像预处理后统一尺寸为 256×256 ,将固定大小的图像输入模型中进行训练和测试。训练时初始学习率设为 0.0005,批处理大小设为 32,迭代次数(epoch)为 200。采用 CB 损失函数优化模型、Stochastic Gradient Descent (SGD) 优化器更新权重参数,动量参数设为 0.9,权重衰减参数设为 0.0005。

3.3 评价标准

采用准确率 (ACC)、敏感度 (SENS)、特异度 (SPEC)、精确率 (PRE) 和 F1 分数 (F1-score) 等多项常用的评价指标对模型性能进行定量评估,计算公式如下:

$$X_{\text{ACC}} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}, \quad (11)$$

$$X_{\text{SENS}} = \frac{T_p}{T_p + F_n}, \quad (12)$$

$$X_{\text{SPEC}} = \frac{T_n}{T_n + F_p}, \quad (13)$$

$$X_{\text{PRE}} = \frac{T_p}{T_p + F_p}, \quad (14)$$

$$X_{\text{F1-score}} = \frac{2 \times X_{\text{PRE}} \times X_{\text{SENS}}}{X_{\text{PRE}} + X_{\text{SENS}}}, \quad (15)$$

式中: T_p 表示正确识别出实际为 ESCC 图像的样本数量; T_n 表示正确识别出实际为正常食管图像的样本数量; F_n 表示实际为 ESCC 图像被错误识别为正常的样本数量; F_p 表示实际为正常食管图像被错误识别为 ESCC 的样本数量。上述评估指标的取值范围都在 $[0, 1]$ 之间,且值越接近 1 表示模型的性能越好。

为了更直观地评估模型的性能,引入受试者工作特征曲线 (ROC) 来表示敏感度与特异度之间的相互关系。ROC 曲线越靠拢左上角,该曲线线下面积 (AUC) 越大,代表模型综合性能越好。

3.4 实验结果分析

首先,利用 5 折交叉验证定量评估 CALite-ResNet 模型的性能和鲁棒性;其次,对比其他网络模型来验证所提模型的有效性和综合性能;最后,定量比较所

3 实验结果与分析

3.1 数据集描述

本研究所用数据集来自多家三甲医院的消化内科,共收集到 206 例患者的 11468 张 NBI-ME 图像,其中,包括 126 例 ESCC 患者的 7147 张病灶图像和 80 例患者的 4321 张正常食管图像。将数据集以病人级别按 8:1:1 的比例随机划分成训练集 (training)、验证集 (validation) 和测试集 (testing),并据此进行 5 折交叉验证来检验网络模型的泛化性和鲁棒性。

深度学习模型性能受训练数据集大小和数据分布的影响,数据不足易使模型产生过拟合现象。如何在有限医学数据集上提升模型的泛化性和稳健性,是 ESCC 识别任务需要解决的难题之一。本研究采用随机旋转 ($0^\circ \sim 45^\circ$)、水平翻转和随机缩放的增强方式对有限图像数据进行扩增。图 8 为不同数据增强方式的效果图示例。

提方法与相关研究方法的泛化性。

3.4.1 模型性能评估

为了使网络模型得到充分的训练并尽可能避免训练结果的偶然性,采用5折交叉验证策略来定量评价模型对ESCC的识别性能和鲁棒性。模型整体性能从两个方面进行评估:图像级别和病人级别,其结果分别如表1、2所示,表中的均值反映模型的有效性,标准差(SD)反映模型的稳定性。

由表1可知,CALite-ResNet模型在图像级别的识别准确率、敏感度、特异度、精准度和F1-score值的均值分别达到了96.39%、95.70%、96.06%、97.37%和0.9652。由表2可知,CALite-ResNet模型在病人级别的分类准确率、敏感度、特异度、精准度和F1-score值

的均值分别达到了95.70%、94.62%、89.24%、95.89%和0.9524。另外,从表中还可看出,在5折交叉验证中,每折的验证准确率相差不大,其标准差不高于1%。综上所述,所提模型具有一定的稳定性和泛化能力,且在基于NBI-ME图像的ESCC识别上具有良好的分类性能。

3.4.2 模型综合性能对比

本小节将CALite-ResNet与图像识别领域常见的网络模型进行比较,从模型参数量、预测时间和图像级别的识别准确率等方面综合衡量模型的性能,实验结果如表3所示。为保证实验的公平性,所有模型训练时设置相同参数,且在同一折数据集下完成训练和测试。

表1 5折交叉验证在图像级别的分类结果

Table 1 Classification results of 5-fold cross-validation at image level

Experiment	ACC / %	SENS / %	SPEC / %	PRE / %	F1-score
1-fold	96.30	94.77	97.76	97.59	0.9616
2-fold	94.65	93.95	96.59	98.71	0.9627
3-fold	97.09	97.81	95.63	97.86	0.9783
4-fold	96.82	96.81	96.77	97.78	0.9729
5-fold	97.07	95.14	93.53	94.91	0.9503
Mean±SD	96.39±0.9136	95.70±1.4090	96.06±1.4325	97.37±1.2887	0.9652±0.0097

表2 5折交叉验证在病人级别的分类结果

Table 2 Classification results of 5-fold cross-validation at patient level

Experiment	ACC / %	SENS / %	SPEC / %	PRE / %	F1-score
1-fold	95.83	94.44	88.89	96.88	0.9564
2-fold	96.15	95.00	90.00	93.33	0.9416
3-fold	95.24	93.75	87.50	96.43	0.9507
4-fold	95.65	95.45	90.90	96.15	0.9580
5-fold	95.65	94.44	88.89	96.67	0.9554
Mean±SD	95.70±0.2953	94.62±0.5731	89.24±1.1494	95.89±1.3039	0.9524±0.0059

表3 不同网络模型的综合性能对比

Table 3 Comprehensive performance comparison of different network models

Model	Params / 10 ⁶	Predicted time / ms	ACC / %	SENS / %	SPEC / %	F1-score
DenseNet ^[27]	12.49	19.40	91.98	91.04	94.57	0.9434
Xception ^[28]	20.81	8.31	92.47	91.49	95.19	0.9470
ResNet50 ^[21]	23.51	11.18	94.32	93.73	95.97	0.9604
ResNeXt50 ^[29]	22.98	13.61	94.48	93.89	96.12	0.9616
Res2Net50 ^[30]	23.01	17.05	94.73	94.06	96.59	0.9633
CALite-ResNet	16.66	16.42	94.65	93.95	96.59	0.9627

从表3可以看出,与ResNet50模型相比,CALite-ResNet模型的预测时间延长了5.24 ms,但其参数量比ResNet50模型减小了 6.85×10^6 ,敏感度提高了0.22个百分点。与其他残差网络相比,CALite-ResNet模型的识别准确率、敏感度和F1-score等评级指标优于ResNet50和ResNeXt50,虽然其准确率、敏感度等略低于Res2Net50模型,但参数量和预测时间较少。另

外,利用ROC曲线评估不同模型的整体识别性能,结果如图9所示。其中,除了Res2Net模型,CALite-ResNet模型的ROC曲线最趋向于左上角,且AUC值高达0.9865。与Res2Net50模型对比可看出,CALite-ResNet模型的ROC曲线与Res2Net50模型的非常接近,且其AUC值只略低了0.0006,整体表现性能和Res2Net50模型也基本相差不大。综上所述,CALite-

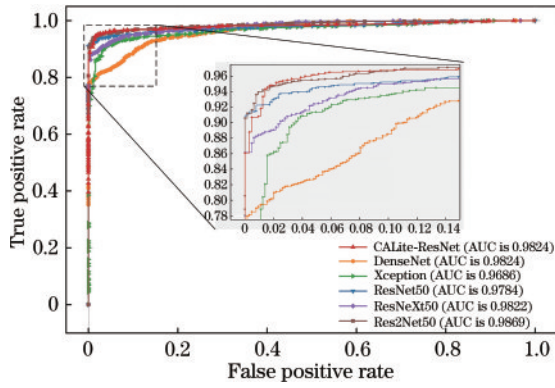


图 9 不同网络模型的 ROC 曲线

Fig. 9 ROC curves of different network models

ResNet 模型在基于 NBI-ME 图像的 ESCC 识别任务

中表现较好,并且在确保高识别精度的同时,大大减轻了硬件环境的负担。

3.4.3 与相关研究方法比较

本小节将所提方法与目前仅有的文献[18]中的基于 NBI-ME 图像的 ESCC 识别研究方法进行对比。除了采用所提多中心数据集以外,也基于文献[18]中公开的 ESCC 数据集进行补充对比,以充分验证所提方法的泛化性。从表 4 的图像级别定量结果可知,所提方法在 2 个数据集上的识别准确率、敏感度和 F1-score 均优于文献[18]方法。其中,准确率和敏感度分别高出 2.51(3)个百分点、2.21(1.92)个百分点。这表明所提方法具有较好的泛化性,能够在 NBI-ME 图像上表现出优异的 ESCC 识别能力,可以为内镜医师提供更加客观、准确的诊断结果。

表 4 与相关研究方法的比较

Table 4 Comparison with related research methods

Method	Proposed dataset			Open dataset		
	ACC / %	SENS / %	F1-score	ACC / %	SENS / %	F1-score
Reference [18]	94.58	95.60	0.9595	93.43	93.70	0.9415
Proposed method	97.09	97.81	0.9783	96.43	95.62	0.9681

3.5 消融实验

为了兼顾精度和效率,本研究设计了 GSCBottleneck 残差模块和 CA 注意力机制,同时采用了 CB 损失函数(CBL)训练策略。本小节消融实验将验证各个模块和训练策略对 ESCC 识别结果的影响。以 ResNet50 作为基准模型(Baseline),在保证数据集、超参数设置等其他实验条件相同的情况下,通过控制累加单一变量的方式,完成此次消融实验。实验结果如表 5 所示。从表 5 可知,GSCBottleneck 模块应用到

基准模型中,使得参数量明显减少,模型更加轻量化。随着 CA 注意力机制的加入,ESCC 识别的准确率提高了 0.87 个百分点。这表明所设计的轻量化残差模块和 CA 注意力机制可以有效优化模型性能。采用 CBL 训练策略后,模型的敏感度达到 97.81%,验证了 CBL 作用于食管 NBI-ME 图像识别任务上的有效性。

为深入研究 CA 注意力机制对模型性能的影响,又设计了模型中是否应添加注意力机制的消融实验。同时,还引入其他注意力模块进行量化对比,包括 Squeeze-and-Excitation (SE)^[31]、Convolutional Block Attention Module (CBAM)^[32]。该实验的基准模型(Baseline)是不引入注意力模块的轻量化残差网络,简称为 Lite-ResNet。表 6 中比较了 Lite-ResNet 模型添加不同注意力机制后对 ESCC 图像的识别结果。从表 6 可看出,模型加入 CA 注意力模块后识别准确率提高了 1.68 个百分点。通过与其他注意力模块的比较可知,CA 模块不仅更轻量化,而且对模型的准确率、敏感度、特异度和 AUC 值都有更好的表现结果。

表 5 消融实验结果对比

Table 5 Comparison of the results of ablation experiments

Method	Params / 10 ⁶	ACC / %	SENS / %	F1-score
Baseline	23.51	96.14	97.21	0.9712
+ GSCBottleneck	15.27	95.46	96.80	0.9663
+ GSCBottleneck+CA	16.66	96.33	97.33	0.9727
+ GSCBottleneck+CA+CBL	16.66	97.09	97.81	0.9783

表 6 不同注意力模块的实验结果比较

Table 6 Comparison of experimental results in different attention modules

Method	Params / 10 ⁶	Predicted time / ms	ACC / %	SENS / %	AUC
Baseline(Lite-ResNet)	15.27	16.48	95.41	96.20	0.9715
+ SE	17.27	23.79	95.98	96.61	0.9782
+ CBAM	17.27	21.35	96.69	97.33	0.9826
+ CA(CALite-ResNet)	16.66	16.42	97.09	97.81	0.9883

为了更直观地验证 CA 注意力机制对 ESCC 识别所起到的积极作用,借助 Grad-CAM^[33]绘制热力图,可

可视化出模型所关注的重点区域信息,以此来定性分析模型对关键特征的提取能力。选取 5 位患者的识别结

果作为示例,具体如图 10 所示,第 1 行为原始图像,第 2 行为 Lite-ResNet 模型的热力图,第 3 行为加入 CA 模块的 CALite-ResNet 模型的热力图。图 10 中呈现的微血管形态区域即为 ESCC 病变。从图 10 可清楚看到,CA 注意力机制能够帮助模型利用图像中的关

键信息学习到有效病变特征,提高模型对 ESCC 的表达能力。结合定量和定性评价结果可知,CA 注意力机制可以提升基于 NBI-ME 图像的 ESCC 识别模型的综合性能。

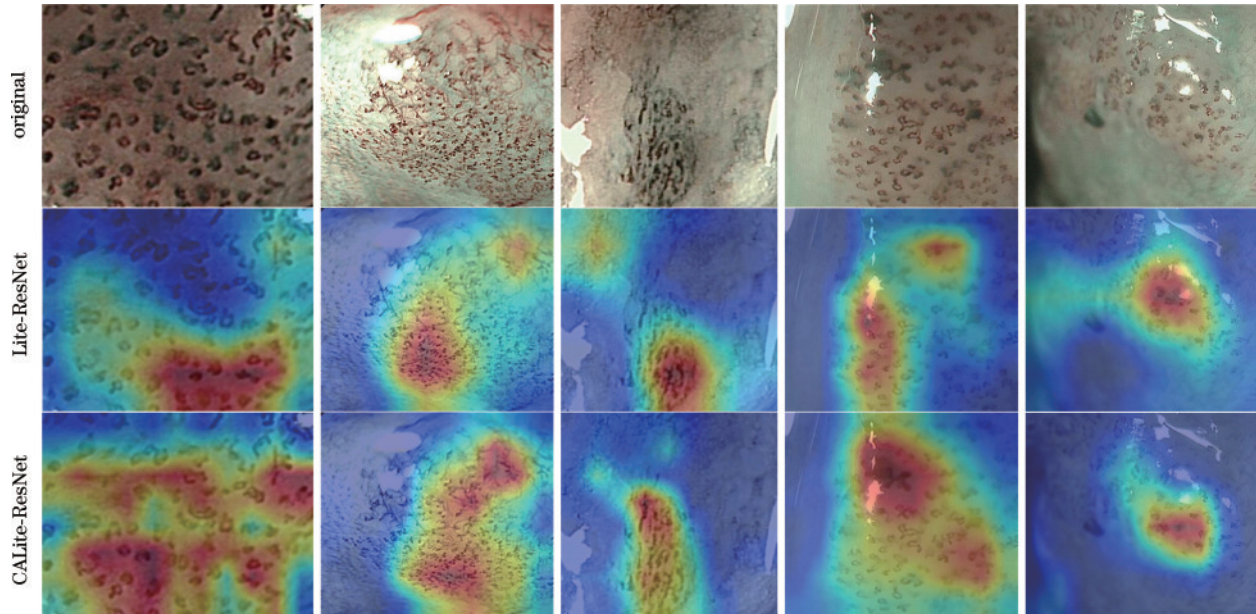


图 10 Grad-CAM 可视化
Fig. 10 Grad-CAM visualizations

4 结 论

提出 CALite-ResNet 对食管 NBI-ME 图像进行分类。该模型能够取得较高的 ESCC 识别准确率,同时兼顾推理效率。利用 GhostModule 和 SCConv 构建 GSCBottleneck 模块来代替 ResNet50 的原始残差结构,在不影响识别精度和预测时间的情况下,降低模型参数量。通过添加 CA 注意力模块,更有效地提取出图像关键信息来增强有用区域的特征描述,提升模型对 NBI-ME 图像中 ESCC 病灶的表征能力。使用 CB 损失函数,降低易分类的样本权重,让模型更加关注困难样本,从而提升整体分类性能。经实验对比可知,CALite-ResNet 具备一定的稳定性和泛化性,对 ESCC 识别的准确率、敏感度等各项评价指标也表现较佳。它以较小的参数量不仅降低了对硬件资源的要求,而且能够较好地完成对 ESCC 高效且精准的识别任务。在今后的研究工作中,将不断收集与扩充食管 NBI-ME 临床病例数据,提升模型的鲁棒性和泛化性,以实现一个更健壮模型并推广应用到 ESCC 的实际临床诊断中。

参 考 文 献

- [1] Sung H, Ferlay J, Siegel R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: a Cancer Journal for Clinicians, 2021, 71(3): 209-249.
- [2] Jia J P, Wang F, Wei C B, et al. The prevalence of dementia in urban and rural areas of China[J]. Alzheimer's & Dementia, 2014, 10(1): 1-9.
- [3] 国家消化内镜专业质控中心, 国家消化系统疾病临床医学研究中心(上海), 国家消化道早癌防治中心联盟, 等. 中国早期食管癌及癌前病变筛查专家共识意见(2019 年, 新乡)[J]. 中华消化内镜杂志, 2019, 36(11): 793-801. National Digestive Endoscopy Improvement System, National Clinical Research Center for Digestive Diseases (Shanghai), National Early Gastrointestinal-Cancer Prevention & Treatment Center Alliance, et al. China experts consensus on the protocol of early esophageal cancer and pre-cancerous lesion screening(2019, Xinxinag) [J]. Chinese Journal of Digestive Endoscopy, 2019, 36(11): 793-801.
- [4] 中华人民共和国国家卫生健康委员会. 食管癌诊疗规范(2018 年版)[J]. 中华消化病与影像杂志(电子版), 2019, 9(4): 158-192. National Health Commission of the People's Republic of China. Specification for diagnosis and treatment of esophageal cancer (2018 edition)[J]. Chinese Journal of Digestion and Medical Imageology (Electronic Edition), 2019, 9(4): 158-192.
- [5] Goda K, Tajiri H, Ikegami M, et al. Magnifying endoscopy with narrow band imaging for predicting the invasion depth of superficial esophageal squamous cell carcinoma[J]. Diseases of the Esophagus, 2009, 22(5): 453-460.

- [6] Liedlgruber M, Uhl A. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review[J]. *IEEE Reviews in Biomedical Engineering*, 2011, 4: 73-88.
- [7] 刘一鸣, 肖志勇. 基于特征融合的肝脏肿瘤自动分割方法[J]. *激光与光电子学进展*, 2021, 58(14): 1417001.
Liu Y M, Xiao Z Y. Automatic segmentation algorithm of liver tumor based on feature fusion[J]. *Laser & Optoelectronics Progress*, 2021, 58(14): 1417001.
- [8] 曹斌, 杨锋, 马金刚. 深度学习方法在肺结节诊断中的应用[J]. *激光与光电子学进展*, 2021, 58(16): 1600005.
Cao B, Yang F, Ma J G. Application of deep learning methods in diagnosis of lung nodules[J]. *Laser & Optoelectronics Progress*, 2021, 58(16): 1600005.
- [9] 赵欣, 王欣, 王洪凯. 融合注意力和 Inception 模块的脑白质病变端到端分割[J]. *光学学报*, 2021, 41(9): 0910002.
Zhao X, Wang X, Wang H K. End-to-end segmentation of brain white matter hyperintensities combining attention and Inception modules[J]. *Acta Optica Sinica*, 2021, 41(9): 0910002.
- [10] Min J K, Kwak M S, Cha J M. Overview of deep learning in gastrointestinal endoscopy[J]. *Gut and Liver*, 2019, 13(4): 388-393.
- [11] Rajkomar A, Dean J, Kohane I. Machine learning in medicine[J]. *The New England Journal of Medicine*, 2019, 380(14): 1347-1358.
- [12] Mori Y, Kudo S E, Mohmed H E N, et al. Artificial intelligence and upper gastrointestinal endoscopy: current status and future perspective[J]. *Digestive Endoscopy*, 2019, 31(4): 378-388.
- [13] Zhang C Z, Ma L, Uedo N, et al. Tu1217 the use of convolutional neural artificial intelligence network to aid the diagnosis and classification of early esophageal neoplasia. A feasibility study[J]. *Gastrointestinal Endoscopy*, 2017, 85(5): AB587-AB588.
- [14] Xue D X, Zhang R, Feng H, et al. CNN-SVM for microvascular morphological type recognition with data augmentation[J]. *Journal of Medical and Biological Engineering*, 2016, 36(6): 755-764.
- [15] Zhao Y Y, Xue D X, Wang Y L, et al. Computer-assisted diagnosis of early esophageal squamous cell carcinoma using narrow-band imaging magnifying endoscopy[J]. *Endoscopy*, 2019, 51(4): 333-341.
- [16] Ohmori M, Ishihara R, Aoyama K, et al. Endoscopic detection and differentiation of esophageal lesions using a deep neural network[J]. *Gastrointestinal Endoscopy*, 2020, 91(2): 301-309.
- [17] Nakagawa K, Ishihara R, Aoyama K, et al. Classification for invasion depth of esophageal squamous cell carcinoma using a deep neural network compared with experienced endoscopists[J]. *Gastrointestinal Endoscopy*, 2019, 90(3): 407-414.
- [18] Everson M, Herrera L G P, Li W, et al. Artificial intelligence for the real-time classification of intrapapillary capillary loop patterns in the endoscopic diagnosis of early oesophageal squamous cell carcinoma: a proof-of-concept study[J]. *United European Gastroenterology Journal*, 2019, 7(2): 297-306.
- [19] Guo L J, Xiao X, Wu C C, et al. Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos)[J]. *Gastrointestinal Endoscopy*, 2020, 91(1): 41-51.
- [20] 腾讯发布一个 AI 神器有望攻克食管癌早筛难题[J]. *信息与电脑(理论版)*, 2017(15): 5.
Tencent's release of an AI artifact is expected to overcome the problem of early screening of esophageal cancer[J]. *China Computer & Communication*, 2017(15): 5.
- [21] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [22] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1577-1586.
- [23] Liu J J, Hou Q B, Cheng M M, et al. Improving convolutional networks with self-calibrated convolutions [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10093-10102.
- [24] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.
- [25] Cui Y, Jia M L, Lin T Y, et al. Class-balanced loss based on effective number of samples[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9260-9269.
- [26] He K M, Zhang X Y, Ren S Q, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1026-1034.
- [27] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.
- [28] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1800-1807.
- [29] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA.

- New York: IEEE Press, 2017: 5987-5995.
- [30] 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.
- [31] Hu J, Shen L, Sun G, et al. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [32] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [33] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 618-626.