计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2025,34(2):216−224 [doi: 10.15888/j.cnki.csa.009752] [CSTR: 32024.14.csa.009752]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 CNN 与 Transformer 混合模型的肺炎辅助诊断①

贠　恺[1], 贾荣浩[1], 魏国辉[1], 赵　爽[2], 李学辉[1], 马志庆[1]

[1](山东中医药大学 智能与信息工程学院, 济南 250355)
[2](山东中医药大学 实验室管理处, 济南 250355)
通信作者: 马志庆, E-mail: mazhq126@163.com

摘　要: 肺炎是一种常见的呼吸系统疾病, 早期诊断对于有效治疗至关重要. 本研究提出了卷积神经网络 (CNN) 和 Transformer 结合的 CTFNet 混合模型, 旨在实现高效而准确的肺炎辅助诊断. 该模型融合了卷积分词器和聚焦线性注意力机制. 卷积分词器通过卷积操作实现更紧凑的特征提取, 并保留图像的关键局部特征降低计算复杂度, 提高模型的表达能力. 聚焦线性注意力机制缓解了 Transformer 的计算需求, 优化了注意力框架, 大幅提升了模型性能. 在 Chest X-ray Images 数据集上, CTFNet 在肺炎分类任务中表现出色, 达到了 99.32% 的准确率、99.55% 的精确率、99.55% 的召回率和 99.55% 的 *F*1 值. 较好的性能凸显了该模型在临床应用中的潜力. 为了评估 CTFNet 的泛化能力, 我们将其应用于 COVID-19 Radiography Database 数据集. 在该数据集中, CTFNet 被用于多个二分类任务均达到 98% 以上的准确率. 这些结果表明, CTFNet 在肺炎图像分类的各种任务中具有较好的泛化能力和可靠性.

关键词: 肺炎图像分类; 卷积神经网络; Transformer; 卷积分词器; 聚焦线性注意力机制

引用格式: 贠恺,贾荣浩,魏国辉,赵爽,李学辉,马志庆.基于 CNN 与 Transformer 混合模型的肺炎辅助诊断.计算机系统应用,2025,34(2):216–224. http://www.c-s-a.org.cn/1003-3254/9752.html

Pneumonia Assisted Diagnosis Based on Hybrid Model of CNN and Transformer

YUN Kai[1], JIA Rong-Hao[1], WEI Guo-Hui[1], ZHAO Shuang[2], LI Xue-Hui[1], MA Zhi-Qing[1]

[1](School of Intelligence and Information Engineering, Shandong University of Traditional Chinese Medicine, Jinan 250355, China)
[2](Laboratory Management Office, Shandong University of Traditional Chinese Medicine, Jinan 250355, China)

Abstract: Pneumonia is a prevalent respiratory disease for which early diagnosis is crucial to effective treatment. This study proposes a hybrid model, CTFNet, which combines convolutional neural network (CNN) and Transformer to aid in the effective and accurate diagnosis of pneumonia. The model integrates a convolutional tokenizer and a focused linear attention mechanism. The convolutional tokenizer performs more compact feature extraction through convolution operations, retaining key local features of images while reducing computational complexity to enhance model expressiveness. The focused linear attention mechanism reduces the computational demands of the Transformer and optimizes the attention framework, significantly improving model performance. On the Chest X-ray Images dataset, CTFNet demonstrates outstanding performance in pneumonia classification tasks, achieving an accuracy of 99.32%, a precision of 99.55%, a recall of 99.55%, and an *F*1-*score* of 99.55%. The impressive performance highlights the model's potential for clinical applications. The model is evaluated on the COVID-19 Radiography Database dataset for its generalization ability. In this dataset, CTFNet achieves an accuracy above 98% in multiple binary classification tasks. These results indicate that CTFNet exhibits strong generalization ability and reliability across various tasks in pneumonia image classification.

① 基金项目: 国家自然科学基金 (61702087); 山东省研究生教育质量提升计划 (SDYJG1943); 山东中医药大学科学研究基金 (KYZK2024Q30)
收稿时间: 2024-07-09; 修改时间: 2024-08-01; 采用时间: 2024-08-13; csa 在线出版时间: 2024-12-16
CNKI 网络首发时间: 2024-12-17

Key words: pneumonia image classification; convolutional neural network (CNN); Transformer; convolutional tokenizer; focused linear attention mechanism

肺炎是一种常见的呼吸系统疾病, 对公共卫生构成重大挑战[1]. 这种疾病由各种病原体引起, 细菌和病毒感染是主要原因[1]. 肺炎的症状很常见, 包括疼痛、咳嗽和呼吸急促. 因此及早识别和及时管理此类感染至关重要[2]. 胸部 X 光检查 (chest X-ray, CXR) 是识别肺炎的常用诊断工具, 在临床管理和流行病学研究中都发挥着关键作用. 然而传统的肺炎诊断方法严重依赖于医生的经验和主观判断, 可能导致效率低下和容易受到个人偏见的影响[3]. 深度学习技术的出现为解决这些挑战开辟了新的途径[4,5]. 随着深度学习的进步和广泛应用, 自动医学图像分类取得了显著改进. 这些进展有助于减少与肺炎相关的诊断错误和遗漏, 从而提高诊断的准确性[6]. 深度学习架构展现出有效的预测能力, 其表现优于医生的诊断[7].

卷积神经网络 (convolutional neural network, CNN) 是一种深度学习技术, 它模仿人类视觉系统的机制来识别图像中的模式和特征. 它们在医学图像分析中显示出巨大的潜力[8]. 这些网络利用多层卷积运算从图像中提取局部特征. Szepesi 等人[9]提出了一个类似于 VGG-16 的网络, 在卷积部分引入了 dropout, 并在每个卷积块内采用 ReLU 激活的批归一化. 该模型在 CXR 肺炎分类中取得了 97.2% 的准确率、97.3% 的召回率和 97.4% 的精确率. Abubeker 等人[10]引入了一个机器学习框架, 将密集连接的 B2-Net 架构与 CNN-160、ResNet-121 和 VGG-16 集成并重新训练, 在 CXR 肺炎分类中达到了 98.88% 的准确率、98.08% 的精确率和 98.08% 的召回率. Sharma 等人[11]提出了 COVDC-Net, 利用 MobileNetV2 和 VGG-16 的 ImageNet 预训练权重. 这些模型使用置信度融合方法进行融合, 在肺炎分类中达到了 96.48% 的准确率、96.59% 的精确率和 96.57% 的召回率.

Transformer 模型凭借其自注意力机制在处理长距离依赖关系方面表现出色[12], 这促成了它在自然语言处理领域的成功[13]. 它捕获全局上下文信息的能力为医学图像分析带来了新的见解[14–16]. Chen 等人[17]提出了用于肺炎分类的 BoT-ViTNet 深度学习网络模型, 该模型通过在 ResNet50 初始 3 个阶段的最后一个瓶颈块中引入多头注意力机制, 并在最后阶段集成结合了 Transformer 和 Bottleneck 结构的 TRT-ViT 块, 增强了 ResNet50 架构. 该模型实现了 98.91% 的准确率、97.80% 的精确率、98.76% 的召回率和 98.27% 的 *F*1 分数. Ukwuoma 等人[18]设计了一种架构, 新颖的特征提取框架与 VIT 模型相结合用于肺炎分类. 该架构首先应用集成技术和全局二阶池化来获得丰富的特征, 然后在对图像进行分块以进行位置嵌入后, 将其输入到 VIT 模型中进行分析. 该模型实现了 97.84% 的准确率、97.84% 的精确率和 97.76% 的召回率.

尽管 CNN 模型在图像识别方面取得了显著的成功, 但其在处理全局信息方面存在局限. Transformer 模型在图像处理任务中展现出处理图像全局信息的潜力[14], 然而 Transformer 在小样本数据集上的效果往往不尽如人意, 而传统 Softmax 注意力机制的高计算需求也限制了其在医学图像分类中的适用性[19]. 为了克服这些局限性, 本文提出了 CTFNet 模型, 将 CNN 的局部特征提取优势与 Transformer 的全局特征表示能力相结合. 在 CTFNet 中, 采用卷积分词器来提高模型在小规模数据集上的效果. 通过引入聚焦线性注意力机制, 优化了计算效率和分类性能, 从而从 CXR 图像中获得更准确的肺炎诊断.

1 模型及方法

在本研究中受 CoAtNet[20]和 MobileNetV2[21]模型的启发, 提出了一种名为 CTFNet 的深度学习模型, 旨在高效分析 CXR 图像并准确诊断肺炎. CTFNet 模型结合了 CNN 和 Transformer 的优点, 实现了对局部和全局特征的有效提取. 模型首先用卷积分词器对输入的 CXR 图像进行初步的特征提取, 卷积分词器能够保留图像中的细节特征. 随后这些特征被送入一系列 MBConv 模块和 FLA-Transformer 模块, 以进一步提取和整合特征信息. MBConv 模块具有卷积神经网络的空间归纳偏置和对数据增强的不敏感性. 特征图经过 MBConv 模块处理后, 能够更好地保留局部特征和细节信息. 然后这些特征图被传递到 FLA-Transformer 模块, 该模块利用聚焦线性注意力机制 (focused linear attention, FLA), 有效捕捉输入数据的全局关系, 增强了模型的注意力焦点能力. 最后经过全连接层得到肺炎

状态预测. 通过这种融合, 不仅继承了卷积神经网络的优点 (如高效的局部特征提取和更少的数据增强敏感性), 还结合了 Transformer 的优势 (如自适应加权和全局信息处理能力), 使得 CTFNet 模型在肺炎图像分类任务中表现出色.

CTFNet 模型的架构如图 1 所示. 带有↓2 标记的模块表示进行下采样操作, 使特征图尺寸缩小为原来的 1/2. 虚线表示残差连接 (residual connection), 即将模块的输入与输出相加, 以缓解网络退化问题, 促进梯度传播.

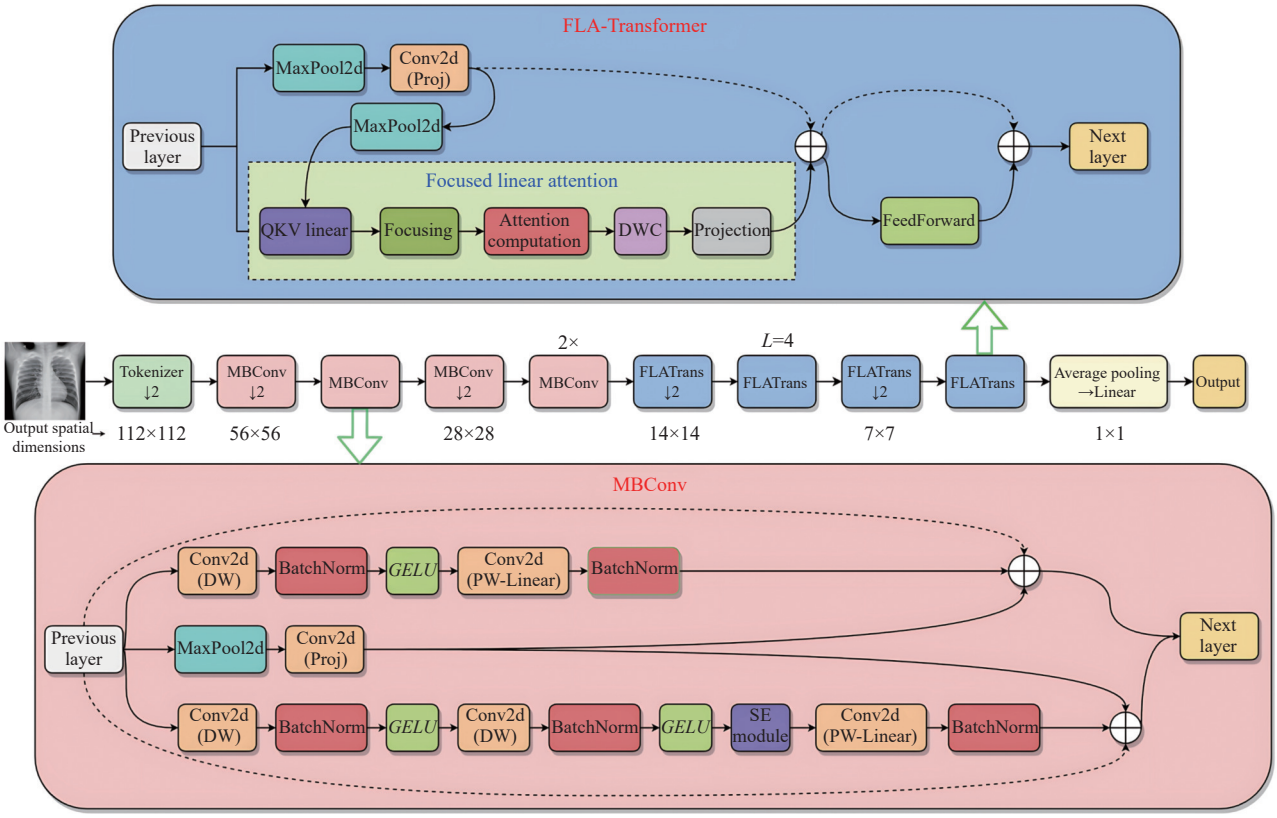


图 1 CTFNet 模型的架构示意图

1.1 卷积分词器

在肺炎图像分类任务中, 准确提取并保留图像中的局部特征至关重要. 然而 Transformer 模型通常需要对输入数据使用 Patch Embedding 进行分块[22], 这种方法在处理医学图像时存在显著缺陷, 这不仅增加了计算复杂度, 还可能导致关键病变区域的信息丢失, 影响分类准确性[23].

为了解决这些问题, 本研究引入了卷积分词器, 卷积分词器使用 3×3 的卷积核对输入图像进行卷积操作, 这种小型卷积核能够捕捉到肺炎病变区域的细微特征, 避免信息丢失. 卷积操作后的特征图再经过 *GELU* 激活函数, *GELU* 相比于传统的 ReLU 激活函数提供了更平滑的非线性处理, 有助于提高模型的稳定性和表示能力. 最后通过最大池化层进一步提取关键特征, 减少特征图的维度, 使得模型更加紧凑和高效. 如图 2 所示.

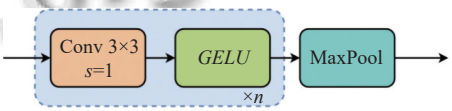


图 2 卷积分词器结构图

具体地, 记输入图像为X, $X \in \mathbb{R}^{H\times W\times C}$. H和W分别是图像的高度和宽度, C表示通道数. 图像通过一系列卷积层进行处理, 每个卷积层用函数$Conv_i$表示, 其中i表示层的索引. 对于第i个卷积层, 公式如下:

$$F_i = Conv_i(X) = GELU(W_i * X + b_i) \tag{1}$$

其中, W_i和b_i分别是第i个卷积层的权重和偏置; $*$表示卷积运算. 经过所有卷积层后, 我们得到最终特征表示F, 其中$F \in \mathbb{R}^{H'\times W'\times C'}$:

$$F = Conv_n(Conv_{n-1}(\cdots Conv_1(X)\cdots)) \tag{2}$$

1.2 MBConv 模块

为了高效提取肺炎图像中丰富的局部信息, 本研

究引入了 MBConv 模块[21], 该模块结合了深度可分离卷积和压缩激发模块 (squeeze-and-excitation, SE)[24], 有效提升了模型的计算效率及性能. 使用 1×1 的卷积核对输入特征进行通道数扩展, 这一步骤为后续深度处理创造了更丰富的特征空间; 采用 3×3 的深度可分离卷积对每个通道进行独立处理, 显著降低了模型的参数量和计算需求; SE 模块通过对每个通道的重要性进行动态调整, 增强了模型对关键特征的捕捉能力, 同时抑制了不重要的信息; 1×1 的线性卷积将处理后的特征映射回原始的输出通道数. 这种结构不仅提高了在资源受限设备上的适用性, 而且通过强化关键特征的表示, 提升了肺炎图像分类的准确性.

1.3 FLA-Transformer 模块

Transformer 编码器由多头注意力机制 (multi-head attention, MSA) 和前馈神经网络 (feedforward neural network, FFNN) 组成. Vision Transformer[14]、PVT[25,26]、Swin Transformer (SwinT)[27]和 CSwin Transformer[28]等模型都采用了 Softmax 注意力机制, 然而这种机制的计算复杂度为二次方O(N^2), 在处理肺炎图像分类任务时, 需要大量的计算资源, 从而限制了模型的实际应用. 因此, 本研究引入了计算复杂度为O(Nd)的聚焦线性注意力机制[29].

$$O_{\text{att}} = \phi_f(Q)\phi_f(K)^{\mathrm{T}}V + DWC(V) \tag{3}$$

其中, Q、K和V分别为查询 (query)、键 (key) 和值 (value), DWC代表深度可分离卷积, R 代表 ReLU 激活函数, $\|\cdot\|$代表欧几里得范数, 聚焦因子f是一个超参数, 通过实验验证来确定其最优值. 在具体实现中, 首先对输入特征应用线性变换以生成Q、K和V的表示. 为了增强模型的焦点性, 通过映射函数ϕ来调整Q和K, 使得相似的查询-键对更靠近, 不相似对更远离, 从而实现更有效的聚焦. 调整后的Q和K表示为: $Q' = (Q/scale)^f$ 和 $K' = (K/scale)^f$. 其中$scale$是通过 Softplus 函数得到的可学习的参数. 调整后的Q'和K'计算注意力权重, 并与V加权求和得到注意力输出. 深度可分离卷积处理注意力输出, 以进一步改善特征表示, 确保捕获局部特征信息, 为模型提供更丰富的表示. FLA 注意力机的制框架图如图 3 所示.

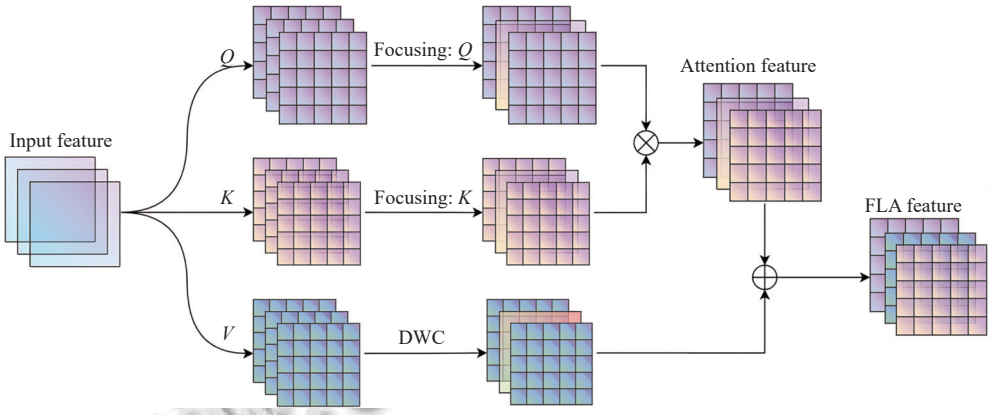


图 3 聚焦线性注意力机制

通过引入聚焦线性注意力机制, 不仅降低了模型的计算复杂度, 同时增强了模型的注意力焦点能力, 从而提升了肺炎图像分类任务的整体性能.

2 实验配置

2.1 数据集

在本研究中, 我们使用了广受认可的 CXR 图像数据集 Chest X-ray Images[30]来进行测试和实验, 该数据集在推进肺炎诊断方面发挥了重要作用. 该数据集源自广州妇女儿童医疗中心 1–5 岁儿童患者的 CXR 图像. 包括两个主要分类: 正常和肺炎, 两个类别中共有 5 856 张 CXR 图像. 这些 JPEG 格式的图像已由医学专家精心标记和分类. 图 4(a) 和图 4(b) 显示了代表性的 CXR 扫描图像.

为了验证 CTFNet 模型的稳健性和泛化能力, 我们还使用了 Kaggle 上的 COVID-19 Radiography Database 数据集[31], 获取了 5 000 张正常 CXR 图像、1 345 张非 COVID-19 病毒性肺炎图像和 3 616 张 COVID-19 病毒性肺炎图像. 图 4(c)、(d) 和 (e) 展示了该数据集中特定的 CXR 扫描图像.

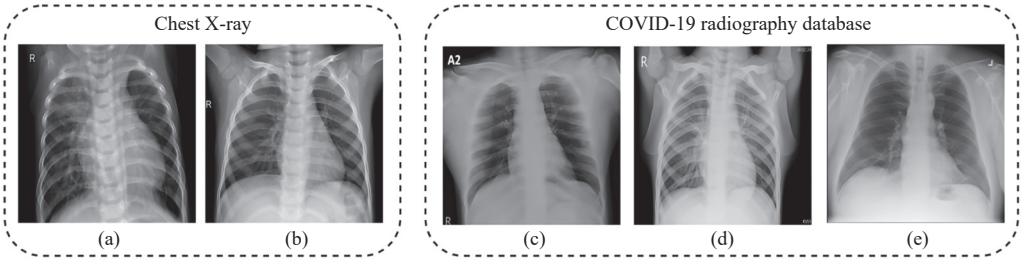


图 4 CXR 图像示例图

2.2 数据预处理

为了增强模型验证的稳健性, 并在训练、验证和测试阶段保持数据分布的一致性, 我们将两个数据集的数据按 8:1:1 的比例划分为训练集、验证集和测试集. 在 Chest X-ray Images 数据集中, 这种分布产生了 4 684 个训练样本、585 个验证样本和 587 个测试样本. 考虑到 COVID-19 radiography database 数据集中非 COVID-19 病毒性肺炎图像相对于正常 CXR 图像数量的差异, 我们选择使用 4 000 张正常 CXR 图像与非 COVID-19 病毒性肺炎数据进行对比. 同样在比较 COVID-19 病毒性肺炎与正常图像时, 我们选择了相同数量的 5 000 张正常 CXR 图像. 在将所有 CXR 图像输入模型之前, 我们对其进行了归一化处理, 并将其调整为统一的 224×224 像素尺寸, 以促进训练稳定性并加快模型收敛速度.

2.3 实验环境和网络配置

本研究中的模型是使用 PyTorch 深度学习框架开发的. 所有实验过程都在 Linux 系统环境中进行, 配备了 AMD EPYC 7551P 处理器、NVIDIA GeForce RTX 3090 显卡和 64 GB 内存.

在模型的训练阶段, 我们选择了双稳态逻辑损失函数 (binary cross-entropy with logits loss, *BCEWithLogitsLoss*)[32], 这种损失函数通过缩小模型预测的概率分布与实际标签之间的差距, 以此引导模型在参数空间中寻找能够减小预测误差的优化解. *BCEWithLogitsLoss* 的优点在于它将 Sigmoid 激活函数与二元交叉熵损失函数结合在一起, 提升模型的数值稳定性和计算效率, 从而在肺炎图像分类任务中获得更好的模型表现. 其具体公式为:

$$BCEWithLogitsLoss(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}[y_i\log(\sigma(\hat{y}_i))+(1-y_i)\log(1-\sigma(\hat{y}_i))] \quad (4)$$

对于模型参数的优化, 我们采用了 AdamW 优化器[33]. AdamW 是一种自适应学习率优化算法, 在参数更新过程中结合了 L2 正则化. 在我们的实验中, 我们将初始学习率 (α) 设置为 0.001, 权重衰减系数 (λ) 设置为 1×10^{-4}. 这有助于在优化过程中对模型复杂度施加惩罚, 以防过拟合.

为了避免模型过早收敛陷入局部最优, 并提高模型在新的未见数据上的泛化能力, 我们实施了基于验证损失的学习率调整策略, 即 ReduceLROnPlateau[34]. 当模型的验证损失在连续 5 个训练周期 (PATIENCE=5) 内没有显著降低时, 该方法会自动降低学习率. 这种动态学习率调整的目的是让模型在训练过程中更细致地探索参数空间, 使模型在接近局部最优解时, 通过减小学习步长更精确地定位到一个更优的解, 从而在一定程度上提升模型对未见数据的预测性能.

2.4 评估指标

在本研究中, 我们采用了几个广泛认可的医学图像分类评估指标, 即准确率 (*Accuracy*)、精确率 (*Precision*)、召回率 (*Recall*)、*F*1 值 (*F*1-*score*) 和混淆矩阵 (confusion matrix). 准确率量化了模型的整体正确分类率. 精确率衡量在所有预测为肺炎的样本中, 真正肺炎病例的比例. 召回率表示模型正确识别的真实肺炎病例的比例. *F*1 值提供了精确率和召回率的调和平均值, 在两者之间取得平衡. 混淆矩阵通过呈现真阳性 (*TP*)、真阴性 (*TN*)、假阳性 (*FP*) 和假阴性 (*FN*), 阐明了模型在肺炎和非肺炎类别上的分类性能, 进一步揭示了模型在不同类别上的分类倾向和潜在缺陷. 这些指标的计算公式如下:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1\text{-}score = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (8)$$

3 实验分析

3.1 消融实验

为了解析每个组件对模型分类效能的个体贡献, 我们进行了一系列消融研究. 通过系统地移除关键组件, 我们评估了它们对模型性能的影响. 具体而言, 我们检验了 CTFNet 及其衍生变体模型.

表 1 展示了结果, 其中本研究提出的 CTFNet 模型在准确率、精确率、召回率和 $F1$ 分数方面均超过其变体, 分别达到 99.32%、99.55%、99.55% 和 99.55%. 这些数据证实了所提出的 CTFNet 模型的优越性和实用性. 通过引入卷积分词器, 模型使用较小的卷积核和 GELU 激活函数有效地处理局部图像特征丰富了特征表示. 通过集成聚焦线性注意力机制, 模型的性能得到进一步提升. 与传统的 Softmax 注意力相比, 聚焦线性注意力机制降低了计算复杂度, 增强了注意力和特征多样性. 这种多样性对于在长距离依赖关系中保持重要的局部特征至关重要. CTFNet 模型则很好地将上述关键组件所展现的优势聚集起来, 展现出卓越的分类性能.

表 1 CTFNet 模型及其变体性能比较 (%)

Model	*Accuracy*	*Precision*	*Recall*	*F1-score*
w/o CT, FLA	96.93	96.41	99.54	97.95
w/o FLA	97.27	98.41	97.97	98.19
w/o CT	97.79	99.54	97.52	98.52
CTFNet	99.32	99.55	99.55	99.55

图 5 展示了 CTFNet 模型及其变体的混淆矩阵. 对角线上的值代表正确分类的样本. 值得注意的是, CTFNet 模型的错误分类数量最少为 4 个, 而移除卷积分词器的模型为 13 个, 移除聚焦线性注意力机的模型为 16 个, 两个关键组件都移出的模型为 18 个, 表明其在正负分类方面都具有稳健的性能. 图 6 描绘了各模型的 ROC 曲线, CTFNet 模型的 AUC (曲线下面积) 接近 1, 证实了其在区分肺炎和非肺炎实例方面的效率和可靠性.

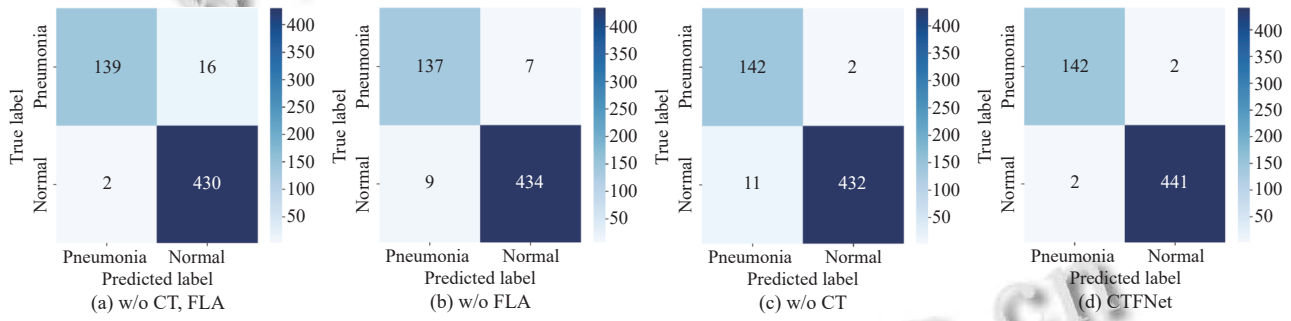


图 5 CTFNet 模型及其变体混淆矩阵分析

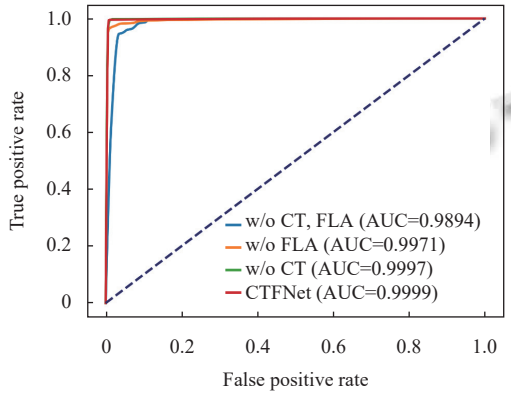


图 6 CTFNet 模型及其变体 ROC 曲线

3.2 模型分类性能研究

为了进一步验证 CTFNet 模型的有效性, 我们将其与已有的方法进行了比较. 表 2 显示, CTFNet 显著优于这些方法, 准确率达到 99.32%, 精确率和召回率超过 99.5%. CTFNet 在特征提取方面的优化减少了对大量预训练数据的依赖, 增强了模型对细节的捕捉及对稀疏数据的处理能力. 确保了高度可靠的临床决策支持, 并减少了假阳性和假阴性的发生, 这对于肺炎的准确诊断以便于及时的治疗至关重要. 图 7 展示了 CTFNet 模型与 ResNet[35]、SwinT[27]、DeiT[36]模型的训练迭代过程对比, 整体上训练过程中的损失值变化结果优于其他网络.

为了评估 CTFNet 模型算法的泛化能力, 我们在 COVID-19 Radiography Database 数据集上进行实验验证. 我们设置了两个分类任务: 区分非 COVID-19 病毒性肺炎和正常图像, 以及区分 COVID-19 病毒性肺炎和正常图像. 在此实验中我们采用了两种不同的训练策略: 一种是加载 Chest X-ray Images 数据集保存的预训练权重 (with pre-training weights), 另一种则是完全从头开始训练 (from scratch). 由表 3 中的结果可知,

CTFNet 在各种数据集上保持出色的性能, 表现出强大的泛化能力. 值得注意的是无论是否使用预训练权重, CTFNet 在 COVID-19 Radiography Database 数据集上的表现都相似, 突显了该模型的稳健性, 并表明其对初始权重配置的依赖性较低, 具有较强的学习和适应能力.

表 2 CFTNet 与现有研究方法对比 (%)

文献	数据集	方法	*Accuracy*	*Precision*	*Recall*
[9]	Chest X-ray Images	改进的VGG-16	97.20	97.30	97.40
[10]	Chest X-ray Images	Dense CNN-160, ResNet-121和VGG-16集成	98.88	98.08	98.08
[11]	COVID-19 Radiography Database	加载ImageNet预训练的VGG-16和MobileNetV2	96.48	96.59	96.57
[17]	COVID-19 Radiography Database	基于ResNet50和多头注意力机制	98.91	97.80	98.76
[18]	COVID-19 Radiography Database	融合新的特征提取框架和ViT模型	97.84	96.80	97.76
[35]	Chest X-ray Images	ResNet	96.25	94.80	96.09
	COVID-19 Radiography Database		97.11	96.76	95.83
[27]	Chest X-ray Images	Swin Transformer	97.79	97.01	97.53
	COVID-19 Radiography Database		97.32	96.91	96.22
[36]	Chest X-ray Images	DeiT	96.93	95.89	96.57
	COVID-19 Radiography Database		96.49	96.07	94.92
Ours	Chest X-ray Images	CTFNet	99.32	99.55	99.55
	COVID-19 Radiography Database		98.97	97.71	98.46

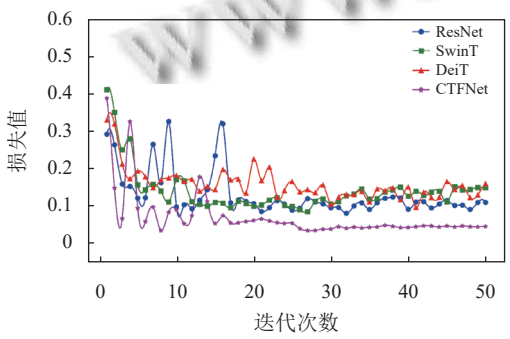
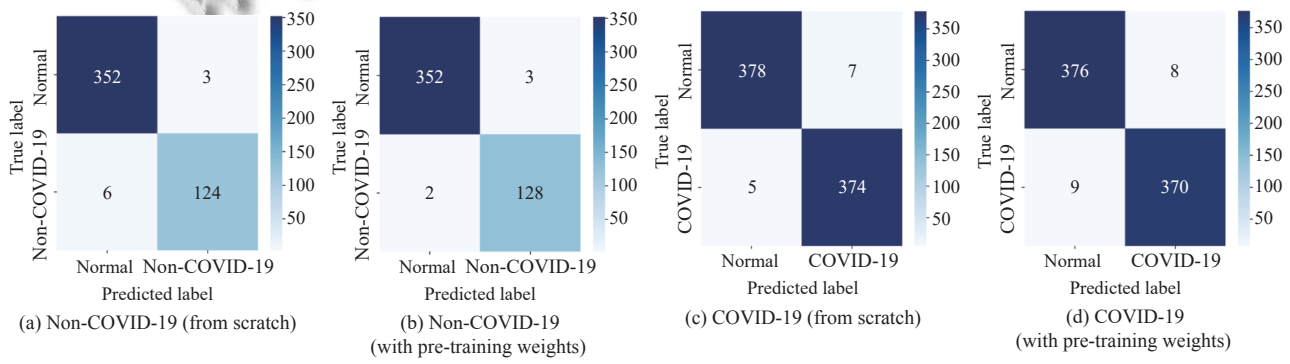


图 7 迭代次数和损失值关系

图 8 和图 9 展示了 CTFNet 在 COVID-19 Radiography Database 数据集上的混淆矩阵和 ROC 曲线, 显示了上述两种训练策略在两个分类任务中的结果. 混淆矩阵揭示了假阳性和假阴性的数量较少, 表明在对两个类别进行分类时表现出色. 在不同情况下, 分类错误的数量始终保持在较低水平: 非 COVID-19 类别在从头开始训练和加载预训练权重的情况下分别为 9 个和 5 个; COVID-19 类别在从头开始训练和加载预训练权重的情况下分别为 12 个和 17 个. ROC 曲线的 AUC 值接近 1, 进一步验证了该模型出色的分类能力, 以及准确区分肺炎阳性和阴性的能力.

表 3 CTFNet 模型泛化能力验证实验 (%)

类别	数据集	方法	*Accuracy*	*Precision*	*Recall*	*F1-score*
Pneumonia and healthy	Chest X-ray Images	From scratch	99.32	99.55	99.55	99.55
Non-COVID-19 and healthy	COVID-19 Radiography Database	From scratch	98.14	97.64	95.38	96.50
		With pre-training weights	98.97	97.71	98.46	98.08
COVID-19 and healthy	COVID-19 Radiography Database	From scratch	98.43	98.16	98.68	98.42
		With pre-training weights	97.77	97.88	97.63	97.75



(a) Non-COVID-19 (from scratch) (b) Non-COVID-19 (with pre-training weights) (c) COVID-19 (from scratch) (d) COVID-19 (with pre-training weights)

图 8 CTFNet 模型泛化能力验证实验混淆矩阵分析

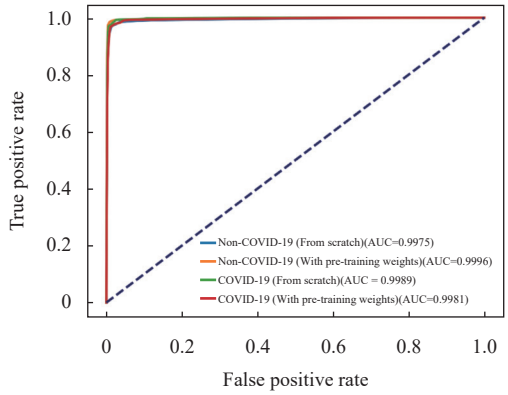


图 9 CTFNet 模型泛化能力验证实验 ROC 曲线

4 结论与展望

本研究提出了 CTFNet 模型, 旨在提高肺炎辅助诊断的准确性. 通过将卷积神经网络特征与 Transformer 架构相结合, 并引入卷积分词器和聚焦线性注意力机制, CTFNet 实现了关键图像特征的高效提取和优化, 这显著提高了模型的计算效率和分类性能. 实验结果证实, CTFNet 在肺炎分类任务中表现出卓越的效能和出色的泛化能力. 本研究所提出的模型不仅能帮助医生准确识别和分类肺炎, 而且显著减少诊断中的漏诊与误诊, 从而为患者提供更加及时和精准的治疗方案.

参考文献

1 Biemba G, Chiluba B, Yeboah-Antwi K, *et al*. Impact of mobile health-enhanced supportive supervision and supply chain management on appropriate integrated community case management of malaria, diarrhoea, and pneumonia in children 2–59 months: A cluster randomised trial in Eastern Province, Zambia. Journal of Global Health, 2020, 10(1): 010425. [doi: 10.7189/jogh.10.010425]

2 Li Q. Convolutional neural networks for pneumonia diagnosis based on chest X-ray images. Proceedings of the 2022 International Conference on Big Data, Information and Computer Network (BDICN). Sanya: IEEE, 2022. 717–720.

3 Wei XX, Niu XK, Zhang XS, *et al*. Deep pneumonia: Attention-based contrastive learning for class-imbalanced pneumonia lesion recognition in chest X-rays. Proceedings of the 2022 IEEE International Conference on Big Data (Big Data). Osaka: IEEE, 2022. 5361–5369.

4 Qi X, Foran DJ, Nosher JL, *et al*. Multi-feature semi-supervised learning for COVID-19 diagnosis from chest X-ray images. Proceedings of the 12th International Workshop on Machine Learning in Medical Imaging. Strasbourg: Springer International Publishing, 2021. 151–160.

5 Suganyadevi S, Seethalakshmi V. CVD-HNet: Classifying pneumonia and COVID-19 in chest X-ray images using deep network. Wireless Personal Communications, 2022, 126(4): 3279–3303. [doi: 10.1007/s11277-022-09864-y]

6 Kundu R, Das R, Geem ZW, *et al*. Pneumonia detection in chest X-ray images using an ensemble of deep learning models. PLoS One, 2021, 16(9): e0256630. [doi: 10.1371/journal.pone.0256630]

7 Salahuddin Z, Woodruff HC, Chatterjee A, *et al*. Transparency of deep neural networks for medical image analysis: A review of interpretability methods. Computers in Biology and Medicine, 2022, 140: 105111. [doi: 10.1016/j.compbiomed.2021.105111]

8 Litjens G, Kooi T, Bejnordi BE, *et al*. A survey on deep learning in medical image analysis. Medical Image Analysis, 2017, 42: 60–88. [doi: 10.1016/j.media.2017.07.005]

9 Szepesi P, Szilágyi L. Detection of pneumonia using convolutional neural networks and deep learning. Biocybernetics and Biomedical Engineering, 2022, 42(3): 1012–1022. [doi: 10.1016/j.bbe.2022.08.001]

10 Abubeker KM, Baskar S. B2-Net: An artificial intelligence powered machine learning framework for the classification of pneumonia in chest X-ray images. Machine Learning: Science and Technology, 2023, 4(1): 015036. [doi: 10.1088/2632-2153/acc30f]

11 Sharma A, Singh K, Koundal D. A novel fusion based convolutional neural network approach for classification of COVID-19 from chest X-ray images. Biomedical Signal Processing and Control, 2022, 77: 103778. [doi: 10.1016/j.bspc.2022.103778]

12 Reiter W. Domain generalization improves end-to-end object detection for real-time surgical tool detection. International Journal of Computer Assisted Radiology and Surgery, 2023, 18(5): 939–944.

13 Ghojogh B, Ghodsi A. Attention mechanism, Transformers, BERT, and GPT: Tutorial and survey. Open Science Framework, 2020. [doi: 10.31219/osf.io/mru2x]

14 Dosovitskiy A, Beyer L, Kolesnikov A, *et al*. An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. OpenReview.net, 2021.

15 Zhang ZH, Gong ZJ, Hong QQ. A survey on: Application of Transformer in computer vision. Proceedings of the 8th International Conference on Intelligent Systems and Image

Processing 2021. The Institute of Industrial Applications Engineers, 2021. 21–28.
16 Viteritti LL, Rende R, Becca F. Transformer variational wave functions for frustrated quantum spin systems. Physical Review Letters, 2023, 130(23): 236401. [doi: 10.1103/PhysRevLett.130.236401]
17 Chen H, Zhang T, Chen RB, *et al*. A novel COVID-19 image classification method based on the improved residual network. Electronics, 2023, 12(1): 80.
18 Ukwuoma CC, Qin ZG, Heyat MBB, *et al*. Automated lung-related pneumonia and COVID-19 detection based on novel feature extraction framework and vision Transformer approaches using chest X-ray images. Bioengineering, 2022, 9(11): 709. [doi: 10.3390/bioengineering9110709]
19 Gandhi D, Shah V, Chawan PM. A vision Transformer approach for classification an a small-sized medical image dataset. Proceedings of the 5th International Conference on Advances in Science and Technology (ICAST). Mumbai: IEEE, 2022. 519–524.
20 Dai ZH, Liu HX, Le QV, *et al*. CoAtNet: Marrying convolution and attention for all data sizes. Proceedings of the 35th International Conference on Neural Information Processing Systems. Curran Associates Inc., 2021. 303.
21 Sandler M, Howard A, Zhu ML, *et al*. MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4510–4520.
22 Clark JH, Garrette D, Turc I, *et al*. Canine: Pre-training an efficient tokenization-free encoder for language representation. Transactions of the Association for Computational Linguistics, 2022, 10: 73–91. [doi: 10.1162/tacl_a_00448]
23 Viriyasaranon T, Woo SM, Choi JH. Unsupervised visual representation learning based on segmentation of geometric pseudo-shapes for Transformer-based medical tasks. IEEE Journal of Biomedical and Health Informatics, 2023, 27(4): 2003–2014. [doi: 10.1109/JBHI.2023.3237596]
24 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.
25 Wang WH, Xie EZ, Li X, *et al*. Pyramid vision Transformer: A versatile backbone for dense prediction without convolutions. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 548–558.
26 Wang WH, Xie EZ, Li X, *et al*. PVT v2: Improved baselines with pyramid vision Transformer. Computational Visual Media, 2022, 8(3): 415–424. [doi: 10.1007/s41095-022-0274-8]
27 Liu Z, Lin YT, Cao Y, *et al*. Swin Transformer: Hierarchical vision Transformer using shifted windows. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 9992–10002.
28 Dong XY, Bao JM, Chen DD, *et al*. CSWin Transformer: A general vision Transformer backbone with cross-shaped windows. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022. 12114–12124.
29 Han DC, Pan XR, Han YZ, *et al*. FLatten Transformer: Vision Transformer using focused linear attention. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023. 5938–5948.
30 Kermany DS, Goldbaum M, Cai WJ, *et al*. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 2018, 172(5): 1122–1131. e9.
31 Chowdhury MEH, Rahman T, Khandakar A, *et al*. Can AI help in screening viral and COVID-19 pneumonia? IEEE Access, 2020, 8: 132665–132676.
32 Amid E, Warmuth MK, Anil R, *et al*. Robust bi-tempered logistic loss based on bregman divergences. Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 1344.
33 Loshchilov I, Hutter F. Fixing weight decay regularization in adam. arxiv:1711.05101, 2017.
34 Al-Kababji A, Bensaali F, Dakua SP. Scheduling techniques for liver segmentation: ReduceLROnPlateau vs OneCycleLR. Proceedings of the 2nd International Conference on Intelligent Systems and Pattern Recognition. Hammamet: Springer, 2022. 204–212.
35 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 770–778.
36 Touvron H, Cord M, Douze M, *et al*. Training data-efficient image Transformers & distillation through attention. Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021. 10347–10357.

(校对责编: 张重毅)