



融合空间域与多轴频域分析的儿童肺炎分类模型

吕佳, 龚泉, 何玲

引用本文:

吕佳, 龚泉, 何玲. 融合空间域与多轴频域分析的儿童肺炎分类模型[J]. 智能系统学报, 2026, 21(4): 963-978.

LÜ Jia, GONG Xiao, HE Ling. Classification model for pediatric pneumonia built by integrating spatial and multi-axis frequency domain analysis[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(4): 963-978.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202510034>

您可能感兴趣的其他文章

多视图主动学习的多样性样本选择方法研究

Diversity sample selection method of multiview active learning classification

智能系统学报. 2021, 16(6): 1007-1014 <https://dx.doi.org/10.11992/tis.202007037>

新冠肺炎疫情趋势预测模型

Epidemics trend prediction model of COVID-19

智能系统学报. 2021, 16(3): 528-536 <https://dx.doi.org/10.11992/tis.202008037>

自步稀疏最优均值主成分分析

Sparse optimal mean principal component analysis based on self-paced learning

智能系统学报. 2021, 16(3): 416-424 <https://dx.doi.org/10.11992/tis.201911028>

样本仿真结合迁移学习的声呐图像水雷检测

Detection of underwater mine target in sidescan sonar image based on sample simulation and transfer learning

智能系统学报. 2021, 16(2): 385-392 <https://dx.doi.org/10.11992/tis.202101030>

基于孪生变分自编码器的小样本图像分类方法

A small-sample image classification method based on a Siamese variational auto-encoder

智能系统学报. 2021, 16(2): 254-262 <https://dx.doi.org/10.11992/tis.201906022>

海底声呐图像智能底质分类技术研究综述

Survey of the intelligent seabed sediment classification technology based on sonar images

智能系统学报. 2020, 15(3): 587-600 <https://dx.doi.org/10.11992/tis.202004026>

DOI: 10.11992/tis.202510034

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20260408.1331.002>

融合空间域与多轴频域分析的儿童肺炎分类模型

吕佳^{1,2}, 龚泉¹, 何玲³

(1. 重庆师范大学 计算机与信息科学学院, 重庆 401331; 2. 重庆师范大学 重庆国家应用数学中心, 重庆 401331; 3. 重庆医科大学附属儿童医院 国家儿童健康与疾病临床医学研究中心儿童发育与疾病教育部重点实验室, 重庆 400014)

摘要: 儿童肺炎是一种严重威胁全球儿童健康的疾病, 其精准分类是有效治疗的关键, 但其 X 射线图像兼具细微病理纹理与大范围弥漫性病变, 单一空间域模型难以高效协同表征。为此, 本文提出融合空间域与多轴频域分析的儿童肺炎分类模型。模型前两阶段构建位移感知多尺度单元, 捕捉细粒度病理纹理; 后两阶段设计多轴傅里叶单元, 在频域建模长距离依赖, 刻画大范围病变特征。此外, 设计多核卷积门控线性单元增强各阶段特征表示。实验表明, 模型在 ChestXRay2017 数据集上, 准确率和 F_1 分数较最优对比模型分别提升 2.72% 和 2.75%, 并在 VinDr-PCXR 和 Curated X-Ray 数据集上均取得具有竞争力的性能。本文结果验证了模型的有效性与泛化性, 为儿童肺炎诊断提供了新方案。

关键词: 儿童肺炎; 图像分类; 深度学习; X 射线; 空频协同; 多轴频域分析; 多尺度; 动态门控

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2026)04-0963-16

中文引用格式: 吕佳, 龚泉, 何玲. 融合空间域与多轴频域分析的儿童肺炎分类模型 [J]. 智能系统学报, 2026, 21(4): 963-978.

英文引用格式: LÜ Jia, GONG Xiao, HE Ling. Classification model for pediatric pneumonia built by integrating spatial and multi-axis frequency domain analysis[J]. CAAI transactions on intelligent systems, 2026, 21(4): 963-978.

Classification model for pediatric pneumonia built by integrating spatial and multi-axis frequency domain analysis

LÜ Jia^{1,2}, GONG Xiao¹, HE Ling³

(1. College of Computer and Information Sciences, Chongqing Normal University, Chongqing 401331, China; 2. National Center for Applied Mathematics in Chongqing, Chongqing Normal University, Chongqing 401331, China; 3. Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders, Children's Hospital of Chongqing Medical University, Chongqing 400014, China)

Abstract: Pediatric pneumonia poses a serious threat to children's health worldwide, and its precise classification is crucial for effective treatment. Chest X-ray images used for pneumonia classification exhibit subtle pathological textures and large-scale diffuse lesions, making it difficult for a single spatial domain model to efficiently and effectively represent both characteristics. To address this issue, this paper proposes a classification model that fuses spatial domain and multi-axis frequency domain analysis. In the first two stages of the model, a shift-aware multi-scale unit is constructed to capture fine-grained pathological textures. In the last two stages, a multi-axis Fourier unit is constructed to model long-range dependencies in the frequency domain and characterize large-scale lesion features. Additionally, a multi-kernel convolutional gated linear unit is developed to enhance feature representation at each stage. Experimental results showed that on the ChestXRay2017 dataset, the model achieved improvements of 2.72% and 2.75% in accuracy and F_1 -score, respectively, compared to the state-of-the-art model. It also achieved competitive performance on the VinDr-PCXR and Curated X-Ray datasets, validating the model's effectiveness and generalization. Thus, the proposed model provides a new solution for the auxiliary diagnosis of pediatric pneumonia.

Keywords: pediatric pneumonia; image classification; deep learning; X-Ray; spatial-frequency domain collaboration; multi-axis frequency domain analysis; multi-scale; dynamic gating

收稿日期: 2025-10-28. 网络出版日期: 2026-04-09.

基金项目: 国家自然科学基金重大项目(11991024); 重庆市自然科学基金创新发展联合基金重点项目(CSTB2025-NSCQ-LZX0074, CSTB2024NSCQ-LZX0090).

通信作者: 吕佳. E-mail: lvjia@cqu.edu.cn.

©《智能系统学报》编辑部版权所有

儿童肺炎是全球范围内威胁儿童健康的主要疾病之一, 研究显示, 它不仅发病率高、传播广, 而且常导致重症与并发症, 对医疗资源和公共卫生系统造成巨大压力^[1]. 该疾病通常是由细菌或

病毒等病原体感染引起,使肺泡充满脓液或填充液,进而导致患儿出现呼吸困难、咳嗽和发热等典型症状^[2]。在临床实践中,准确区分细菌性肺炎与病毒性肺炎对于指导抗生素使用、优化治疗方案和减少耐药性风险至关重要^[3]。胸部X射线(chest X-ray, CXR)因其低成本、无创性和广泛可及性,成为儿童肺炎诊断的首选影像学检查手段^[4]。然而,与成人肺炎常表现为边界清晰的肺叶实变不同,儿童肺部发育尚未成熟,肺实质纹理更为细腻复杂,且病毒性感染常导致大范围的弥漫性间质改变而非局灶性病灶。这种极细微的解剖纹理特征与广泛的弥散性病理特征在儿童CXR图像上的高度交织,导致不同类型肺炎的影像表现重叠度极高^[5]。传统的人工判读在面对这种既需要关注微观纹理又需要把握宏观分布的复杂情况时,往往难以兼顾,导致主观性强、误诊率高^[6]。因此,实现儿童CXR图像肺炎自动化分类可提升医生的诊断效率,并有助于应对医疗资源紧缺这一现实挑战。

在早期的儿童肺炎自动化分类研究中,多采用支持向量机、 K 近邻和随机森林等传统机器学习算法,对手工提取的纹理、形状等低层次特征进行分类^[7]。这类方法在特定数据集上取得了一定效果,但它们难以捕捉区分儿童肺炎所需的高层语义信息,易导致诊断效果不佳^[8]。

近年来,以卷积神经网络(convolutional neural network, CNN)为代表的深度学习算法,具有强大的特征学习和特征表达能力,在医学图像领域得到广泛应用,并在肺部疾病的临床辅助诊断中取得了显著成效^[9]。早期,多数研究以VGG(visual geometry group)、ResNet和DenseNet等主流CNN为基础,并结合迁移学习,在儿童肺炎二分类和多分类任务中取得了较高的精度^[10-11]。然而,受限于卷积核的有限感受野,这类方法难以充分捕捉跨区域的全局依赖关系;同时,在面对形态多样的复杂病灶时,其对细粒度特征的捕捉也相对欠缺。针对这些问题,研究者提出了多种改进策略以增强CNN的全局特征建模与多尺度信息交互能力。例如, Du等^[12]提出的PF-SAA(parameter-free similarity-aware attention)模块通过无参数相似度建模的注意力机制,在不增加额外可训练参数的情况下实现全局上下文捕捉和跨尺度特征融合,显著提升了儿童肺炎影像分类的准确率与泛化性能。MADenseNet(multi-resolution attention dense network)^[13]结合多分辨率空间注意与坐标频率注意机制,通过深浅层特征融合与语义交互增

强,实现多尺度病灶与细粒度特征的联合提取,有效提升了肺炎分类的准确性与鲁棒性。在局部与全局结构协同建模上, SAGCN(self-adaptive graph convolutional network)^[14]引入自适应图卷积机制,自动学习像素块之间的结构关系,同时捕捉局部与全局依赖,有效提升复杂病灶识别结果的稳定性。

尽管上述方法通过改进CNN增强了模型的全局特征建模和多尺度信息交互能力,但基于CNN的分类模型固有的局部归纳偏置仍是其性能瓶颈,导致其捕捉肺炎病灶全局分布的能力始终有限。为克服这种局限性,源自自然语言处理领域的Transformer被引入视觉任务。ViT(vision transformer)^[15]作为其中的开创性工作,其核心的自注意力机制能够计算图像中所有块之间的关系,展现出良好的全局建模优势,这对于识别如肺炎等具有弥散性病灶的疾病具有一定优势。但仅基于ViT的方法缺乏CNN固有的平移不变性和层次化特征提取能力,在处理图像中病灶边缘、纹理等局部细节特征时表现不够精细,同时其高计算复杂性和数据依赖仍限制其在资源受限设备上的部署。因此,各类CNN与ViT的混合结构随之出现。IEViT(input enhanced ViT)^[16]在标准ViT的基础上引入卷积增强的patch embedding和位置感知机制,有助于模型识别肺泡密度变化与支气管周围模糊影。尽管如此,该方法在应对小病灶与模糊边界时存在不足,对局部细粒度纹理的捕捉能力有限。为弥补该短板, Ukwuoma等^[17]引入并行卷积分支与Transformer分支的混合结构,并结合注意力可视化机制,在儿童肺炎分类中不仅提高了对典型病灶区域的敏感性,还增强了可解释性。但其在跨数据集时的鲁棒性有限,泛化性能有待加强。为此, MedViT^[18]引入对抗攻击设计,通过特征均值和方差置换增强模型在处理CXR图像中的噪声和细微病灶差异的鲁棒性。然而, MedViT侧重于提升模型的稳健性,容易忽视跨区域的全局依赖。为解决这一问题, Laskar等^[19]提出紧密耦合的CNN-Transformer融合机制,通过多尺度特征混合与全局上下文协同建模,有效提升了对弥漫性病灶与复杂纹理的识别能力,从而在儿童肺炎分类任务中取得了更均衡的性能表现。尽管这些混合架构在一定程度上提升了模型性能,但它们本质上仍局限于空间域内的信息整合。这导致模型在应对复杂的儿童肺炎影像时,难以高效兼顾细微纹理特征与全局病理模式两种尺度信息。

在医学图像分析中,空间域是指直接在像素空间进行特征提取和分析的方法,其核心是利用像素的空间位置关系、邻域结构和局部纹理模式来表征图像特征。具体地,深度神经网络通过直接作用于图像像素或局部区域的计算操作,能够捕捉不同尺度下的局部形态特征、边缘纹理和结构信息。相对而言,频域则将图像转换到频率空间,以频谱的幅度和相位来表征图像的全局周期性模式和长距离依赖关系。近年来,在医学图像处理领域,融合频域分析的深度学习模型表现出更好的性能。例如,SFC-Net(space-frequency co-ordination cnn-transformer encoder-decoder network)^[20]通过引入基于离散傅里叶变换的多视图频域提取器,从不同视角捕获医学图像的全局频谱特征,并结合频域空间协同注意机制强化多频谱信息建模,有效提升了多器官分割的结构识别能力;SFIPNet(spatial frequency domain interactive perception network)^[21]通过跨层傅里叶差异建模与分组频谱感知模块,从幅度与相位差异中提取多频段特征,有效提升医学图像分割的边界精度与细节表现;Ruan等^[22]则系统性地分析了医学图像中类间与类内的频率差异,并设计了频谱融合模块,通过整合不同频段的信息,显著提升了模型对边界结构的敏感度和类别判别能力;此外,FreMIM方法^[23]将傅里叶变换引入掩码图像建模框架,通过在频域执行重建任务,引导模型捕捉图像中隐式的周期性结构,从而有效增强了模型在医学图像分割任务中的性能。这些研究共同表明,频域分析能够提供与空间域互补的关键信息,有助于模型构建更全面、更鲁棒的特征表示。然而,上述模型主要针对医学图像分割任务设计,侧重于像素级边界细节的恢复与重建。若将其直接应用于儿童肺炎分类任务,不仅存在编码器-解码器架构上的冗余,且难以针对性地提取分类所需的图像级全局语义特征。鉴于分割与分类在特征抽象层级上的本质差异,现有方法难以直接适配儿童肺炎中细微纹理与弥漫性病灶并存的复杂场景。

鉴于上述分析,本文将空间域的局部细节捕捉能力与频域的全局结构建模优势相结合,提出了一种结合空间域和频域特征的儿童肺炎分类模型(spatial and multi-axis frequency domain fusion former, SMFDFormer),旨在通过挖掘 CXR 图像中的多尺度空间特征与多频率全局信息,有效捕捉区分不同肺炎所需的细微病理纹理特征和大范围弥漫性病变特征,提升儿童肺炎类型的自动识

别精度,增强模型的泛化能力。本研究的主要贡献如下:

1)设计了位移感知多尺度单元(shift-aware multi-scale unit, SAMU)。该单元通过零参数的通道位移操作与多核卷积并行结构相结合,专注于捕捉儿童肺炎 CXR 图像中丰富的细粒度病理纹理。

2)设计了多轴傅里叶单元(multi-axis Fourier unit, MAFU)。该单元通过在多个特征平面进行频域变换与动态门控处理,捕获肺部病变的全局上下文关系和长距离依赖关系,适配儿童肺炎中大范围分布特征。

3)设计了多核卷积门控线性单元(multi-kernel convolution gate linear unit, MK-CGLU)。该单元通过多核卷积门控机制,自适应地筛选并融合多尺度特征,从而提升模型对复杂病理模式的识别能力。

1 融合空间域与多轴频域的儿童肺炎分类方法

1.1 模型整体结构

SMFDFormer 由卷积主干(stem)和4个串联处理阶段构成,整体结构如图1所示。stem负责对输入 CXR 图像进行初步特征提取与下采样,通过多步卷积操作将原始像素信息转换为低阶特征图,为后续阶段提供稳定的特征输入。4个处理阶段通过逐级下采样构建特征金字塔,实现从局部细节到全局结构的多尺度特征感知。各阶段空间分辨率依次减半,通道维度依次倍增。在每个阶段均采用 MetaFormer 的统一块结构设计,这些块具体实现为 SIBlock 和 MAFBlock。每个块包含令牌混合器和通道混合器两个核心组件。其中,令牌混合器根据不同阶段进行切换。其具体的模块嵌入策略遵循视觉感知中的层级化特征表示原则,以适配儿童肺炎的病理特性。考虑到细菌性肺炎的局部实变特征主要体现为高频的细粒度纹理,且主要保留在高分辨率的浅层特征中,因此,Stage1~2采用 SAMU 作为令牌混合器。该模块利用位移操作在保持高分辨率空间信息的同时,强化局部纹理的捕捉能力。由病毒性肺炎引起的大范围弥漫性病变,属于低频的全局分布特征,更适合在模型的深层进行抽象建模。鉴于传统卷积在深层仍受限于局部感受野,Stage3~4设计 MAFU 作为令牌混合器。该模块利用傅里叶变换的全局属性,在深层特征空间中直接捕获长距离依赖关系,以弥补空间域卷积对弥散性病灶

建模的不足。通道混合器采用 MK-CGLU 结构进行通道维度的特征融合。最终,模型通过全局平

均池化聚合高阶特征,并经分类头输出肺炎分类结果。

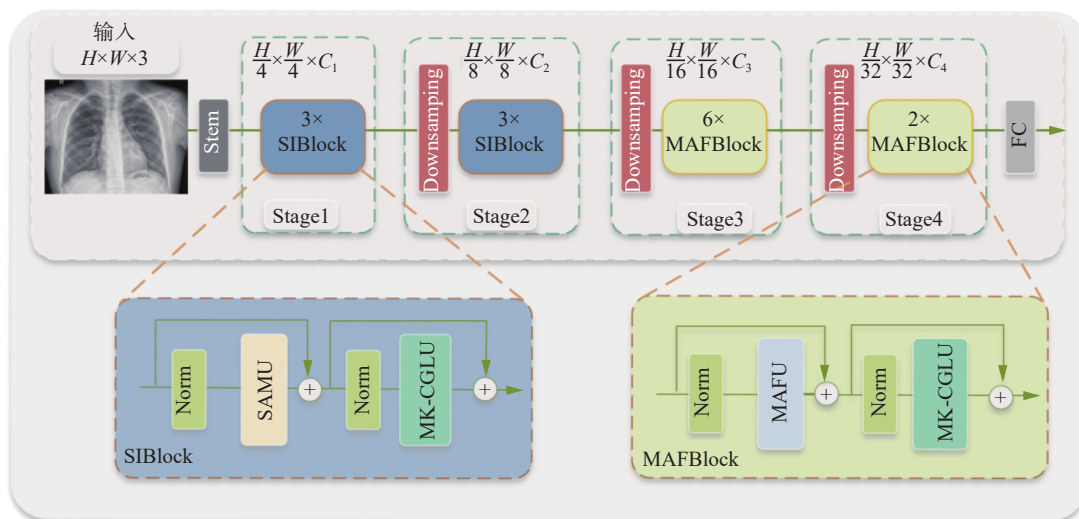


图 1 SMFDFormer 整体结构

Fig. 1 Overall structure of SMFDFormer

1.2 SAMU

为高效捕获 CXR 图像中的多尺度病灶和细

粒度特征,本文在模型的 Stage1~2 采用 SAMU 作为空间域特征提取的核心模块,结构如图 2 所示。

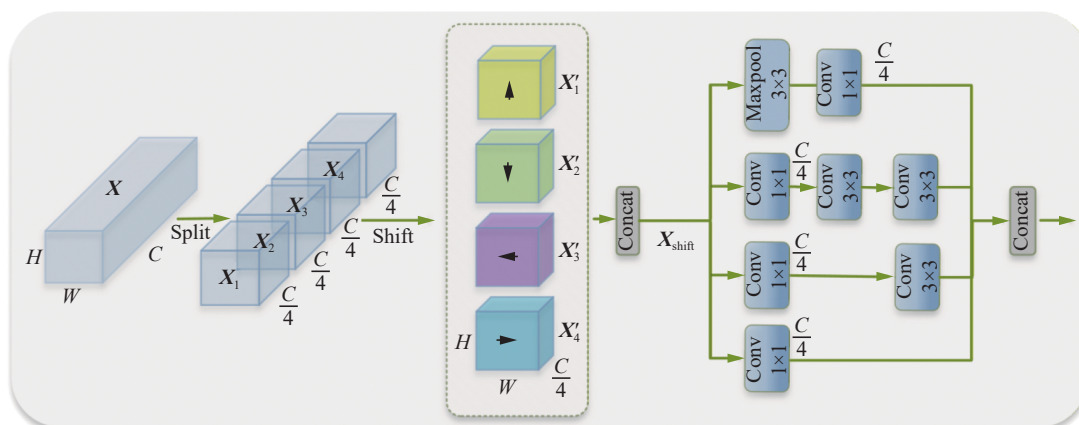


图 2 SAMU 结构

Fig. 2 Structure of SAMU

该模块直接在像素空间对图像进行操作,主要由通道位移操作和多分支特征提取结构两部分组成。输入特征图 $X \in \mathbf{R}^{C \times H \times W}$ 首先会执行通道位移混合^[24]操作,位移细节如图 3 所示。以图 3 左上中心的白色像素为例,其空间邻域包含上、左、下、右等不同特征。在执行空间位移操作时,特征图按通道被均分为 4 组,分别为 X_1 、 X_2 、 X_3 、 X_4 ,如图 3 右下所示。然后各组分别执行特定方向的循环移位,原本位于周边的绿色、蓝色、黄色和紫色像素被对齐至中心像素。由于该操作是基于整个特征张量并行执行的,因此特征图上的每一个像素点均遵循完全相同的聚合逻辑,将其对应方位的邻域特征对齐至自身位置。

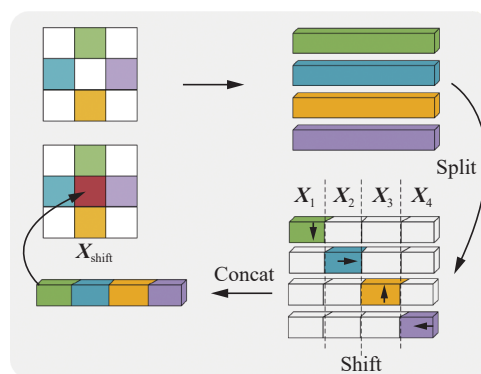


图 3 位移操作细节

Fig. 3 Details of shift operations

具体过程可表示为

$$\begin{cases} X'_1 = \text{Shift}(X_1, s, 0) \\ X'_2 = \text{Shift}(X_2, -s, 0) \\ X'_3 = \text{Shift}(X_3, 0, s) \\ X'_4 = \text{Shift}(X_4, 0, -s) \end{cases}$$

式中 s 为位移步长。将这 4 个经过信息重组的子特征图沿通道维度重新拼接, 得到一个经过空间信息增强的中间特征图 X_{shift} 。这一步骤在不引入任何额外参数和计算量的情况下, 对空间信息进行了预混合, 使得后续的卷积核在执行计算时, 能够感知更大范围的上下文信息。

随后, X_{shift} 被送入一个多分支特征提取结构^[25]中。该结构并行部署了 1×1 卷积用于通道维度的信息整合与降维, 3×3 卷积专注于捕捉微小的局部病理纹理, 串联 3×3 卷积在减少参数量的同时获得了等效于 5×5 的感受野, 旨在引入更宽阔的局部邻域以识别具有一定空间跨度的纹理模式, 最大池化分支则用于保留局部区域内最显著的特征响应。这种四路并行的结构在不同尺度上同步提取特征, 以适应肺炎病灶形态各异的特点。具体过程表示为

$$\begin{cases} Y_1 = \text{Conv}_{1 \times 1}(\text{MaxPool}_{3 \times 3}(X_{\text{shift}})) \\ Y_2 = \text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(X_{\text{shift}})) \\ Y_3 = \text{Conv}_{3 \times 3}(\text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(X_{\text{shift}}))) \\ Y_4 = \text{Conv}_{1 \times 1}(X_{\text{shift}}) \end{cases}$$

最终, 所有分支的输出在通道维度上进行拼接, 以聚合所有尺度的特征。拼接后的特征图再

经过批归一化和 GELU 激活函数处理, 以稳定训练并增强模型的非线性表达能力。其计算过程为

$$X_{\text{out}} = \text{GELU}(\text{BN}(\text{Concat}(Y_1, Y_2, Y_3, Y_4)))$$

这种通过空间信息重组来增强多尺度特征提取的一体化设计, 增强了模型对空间局部特征的捕捉能力, 有助于更精准地识别形态各异的肺部病灶结构。

1.3 MAFU

传统卷积操作受限于感受野, 难以高效捕捉 CXR 图像中弥漫性病变等具有大尺度分布特征的全局模式。傅里叶变换虽具有全局建模优势, 但频域中的特征差异并非分布在单一维度, 而是分散于通道与空间等多个维度组合中^[22]。为了弥补空间域操作的不足并从不同维度组合中全面捕捉这些分散的频域特征, 本文设计了 MAFU, 主要分为多轴频域分解和傅里叶频域处理两个步骤, 结构如图 4 所示。

1.3.1 多轴频域分解

为实现多轴的频域分析, MAFU 首先将输入特征图 $X \in \mathbf{R}^{C \times H \times W}$ 按通道均匀分为 3 组。

$$X = [X_s \ X_{ch} \ X_{cw}], C_s, C_{ch}, C_{cw} = \frac{C}{3}$$

式中 X_s 、 X_{ch} 和 X_{cw} 分别用于空间平面 ($H \times W$)、通道高度平面 ($C_{ch} \times W$) 和通道宽度平面 ($C_{cw} \times W$) 的频域分析, 如图 4(a)。此后, 对 3 个平面做频域转换以进行频域处理, 变换步骤如下。

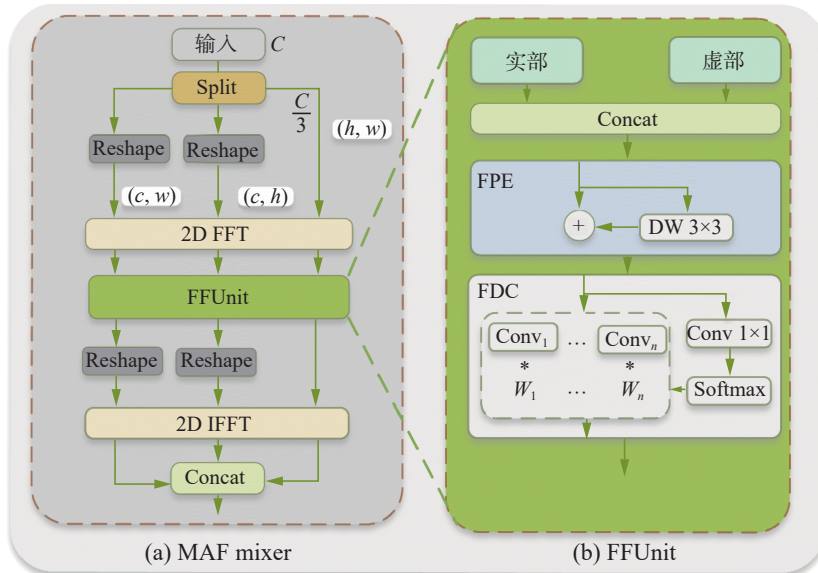


图 4 MAFU 结构

Fig. 4 Structure of MAFU

首先, 在空间平面直接对 X_s 进行二维傅里叶变换:

$$F_s = \mathcal{F}_{2D}(X_s) \in \mathbb{C}^{C_s \times H \times W'}$$

式中: $\mathcal{F}_{2D}(\cdot)$ 为对最后两个维度的二维傅里叶变换, $W' = W/2 + 1$, $\mathbb{C}$ 为复数域。

其次, 为了在通道高度平面进行频域分析, 需

要对 X_{ch} 做维度重排, 以方便后续的频域变换。在实现上, 将 X_{ch} 通过 P_{ch} 进行转换并对其最后两个维度执行傅里叶变换。

$$X_{ch}^p = P_{ch}(X_{ch}) \in \mathbf{R}^{W \times C_{ch} \times H'}$$

$$F_{ch} = \mathcal{F}_{2D}(X_{ch}^p) \in \mathbb{C}^{W \times C_{ch} \times H'}$$

式中 $H' = H/2 + 1$ 。同理, X_{cw} 通过 P_{cw} 转换并对其最后两个维度执行傅里叶变换。

$$X_{cw}^p = P_{cw}(X_{cw}) \in \mathbf{R}^{H \times C_{cw} \times W'}$$

$$F_{cw} = \mathcal{F}_{2D}(X_{cw}^p) \in \mathbb{C}^{H \times C_{cw} \times W'}$$

经过上述操作, 3 个平面分别被转换到频域内进行处理。

1.3.2 傅里叶频域处理单元

每个轴的频域特征由傅里叶频域处理单元 (fourier frequency unit, FFUnit) 处理, 结构如图 4(b) 所示。由于频域中的各个分量对应于空间域中的特定像素集, 需要对每个频率分量进行不同的处理。因此, 为了在各正交平面有效地提取唯一的频率特征, FFUnit 包含了频率位置编码 (frequency position encoding, FPE) 和频率动态卷积 (frequency dynamic convolution, FDC) 两个模块。FFUnit 的输入为频域复数特征的实部与虚部拼接形成的实数张量。

$$F_{real} = \text{Concat}(\text{Re}(F), \text{Im}(F))$$

FPE 采用 3×3 深度卷积作为可学习位置编码器, 以残差形式增强频率位置感知。

$$F_{fpe} = F_{real} + \text{DWConv}(F_{real})$$

FDC 由特征变换与动态权重生成两部分协同构成: 前者通过 1×1 分组卷积对频域特征进行深度变换, 输出 F_{fdc} ; 后者并行通过 1×1 卷积与 Softmax 生成数据驱动的门控权重 W , 实现对不同频率分量的自适应聚焦。两者融合, 完成频谱的动态选择与增强。

$$F_{gate} = \sum_{g=1}^G F_{fdc}^{(g)} \cdot W^{(g)}$$

式中: G 为分组数, $F_{fdc}^{(g)}$ 和 $W^{(g)}$ 分别为第 g 组的特征与权重。

最后, 经过 FFUnit 处理的特征被重塑为复数形式, 通过逆快速傅里叶变换转回空间域。3 个轴的输出特征在通道维度拼接, 形成 MAFU 的最终输出。

1.4 MK-CGLU

在网络的特征处理流程中, 通道混合器扮演着通道特征融合的关键角色。卷积门控线性单元 (convolutional gated linear unit, CGLU)^[26] 作为一种高效的通道混合器实现, 它能够通过门控机制

高效地进行通道间的融合交互, 并利用卷积来建模空间上下文。然而, CGLU 一个局限性在于, 其卷积层采用了单一的、固定大小的卷积核。这一设计迫使模型只能在某一个特定尺度上学习空间关系, 从而束缚了所学得的特征表示的丰富性。为了克服传统 CGLU 的局限性, MK-CGLU 用一个更具动态性和功能性的多核深度卷积模块, 替代了标准的单核卷积, 具体实现如图 5 所示。

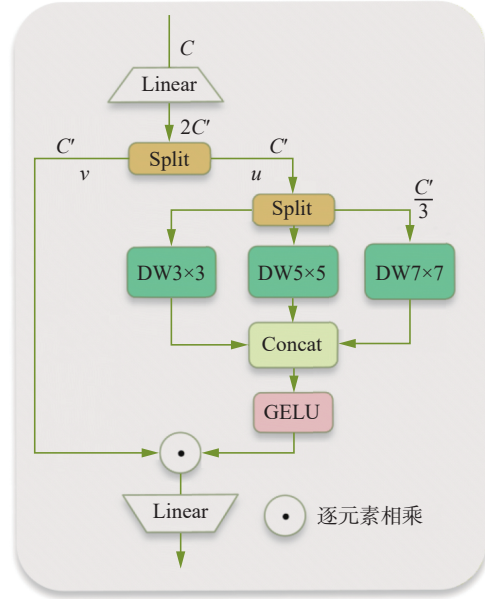


图 5 MK-CGLU 结构

Fig. 5 Structure of MK-CGLU

对于输入的特征张量 $X \in \mathbf{R}^{C \times H \times W}$, 经线性层投影至 $2C'$ 维度 (C' 为隐藏层通道数), 沿通道维度均分为两个并行分支 u 和 v 。其中, u 负责特征变换, v 作为后续通道门控融合分支。对于 u 分支, 通过一组并行的深度可分离卷积路径对其进行处理, 每个路径使用不同尺寸的卷积核。此后, 将从不同卷积核得到的输出特征进行拼接, 经激活后与 v 分支逐元素相乘并投影回 C 维度得到最终输出。

1.5 算法实现流程

为了清晰阐明 SMFDFormer 从图像输入到分类输出的完整过程以及模型参数的优化策略, 本文设计了具体的算法流程, 具体参数设置如表 1 所示。模型整体训练流程包含前向特征传播与反向参数更新两个核心阶段。在前向传播中, 输入 CXR 图像 Z 首先经过 stem 模块变换为初始特征图 X 。随后, 模型进入 4 个阶段的特征提取循环。根据儿童肺炎病理特征的层级分布, 算法执行关键的模块切换逻辑: 当处于浅层阶段 ($n \leq 2$) 时, 调用 SAMU 模块以提取高频的细微纹理特征; 当进入深层阶段 ($n > 2$) 时, 切换至 MAFU 模块

以捕获低频的全局弥漫性特征。每个阶段内的特征图经令牌混合器处理后,通过残差连接与MK-CGLU通道混合器进行融合,并在阶段结束前(除最后一层外)执行下采样操作。最终,高阶特征经全局平均池化和MLP分类头映射为预测结果 Y_{pred} 。

表1 模型参数设置

Table 1 Model parameter setting

参数名称	设定
训练轮数	100
批次大小	24
优化器	AdamW
初始学习率	1×10^{-4}
权重衰减系数	0.05
学习率调度策略	Cosine
最小学习率	1×10^{-6}
各阶段通道数 C	[96,192,384,768]
各阶段模块数 L	[3,3,6,2]
位移 s	2

在反向传播阶段,模型计算预测值 Y_{pred} 与真实标签 Y_{true} 的交叉熵损失,并计算梯度 g_{θ} 。随后,利用梯度信息更新模型参数 θ ,得到优化后的参数集合 θ^* 。参数优化采用AdamW算法,并配合余弦退火策略在训练过程中动态衰减学习率,以保证模型在收敛速度与泛化能力之间的平衡。如下是模型训练流程的伪代码。

输入 CXR图像输入样本 Z ;输入样本对应的真实标签 Y_{true} ;模型初始参数集合 θ ;每个阶段 n 的堆叠块数量 L ;训练轮数 E ;学习率 η ;

输出 输出样本的分类预测结果 Y_{pred} ;经反向传播优化后的模型参数集合 θ^* 。

1)初始化参数

2)for epoch=1 to E do

3)for each batch(Z, Y_{true}) in DataLoader do

4) $X \leftarrow \text{stem}(Z)$ //初始特征提取:卷积层将图像变换为初始特征图

5)for $n=1$ to 4 do//4阶段特征提取与增强

6)for $l=1$ to $L[n]$ do

7) $X_{mix} \leftarrow (n \leq 2) ? \text{SAMU}(X) : \text{MAFU}(X)$ //浅层Stage1~2空间局部特征提取或深层Stage3~4频域全局特征提取

8) $X \leftarrow X + X_{mix}$

9) $X_{fused} \leftarrow X + \text{MK-CGLU}(X)$

10)end for

11)if $n < 4$ then

12) $X \leftarrow \text{DownSample}(X)$ //下采样压缩特征图尺寸并扩展通道

13)end if

14)end for

15) $Y_{pred} \leftarrow \text{MLP}_{head}(\text{GlobalAvgPool}(X))$ //全局平均池化生成特征向量分类头输出预测结果

16) $\mathcal{L} \leftarrow \text{CrossEntropy}(Y_{pred}, Y_{true})$ //计算预测值与真实标签的损失

17) $g_{\theta} \leftarrow \text{gradient}(\mathcal{L}, \theta)$ //计算梯度

18) $\theta^* \leftarrow \text{AdamW}(\theta, g_{\theta}, \eta)$ //反向传播更新模型参数

19)end for

20)使用余弦退火策略更新学习率

21)end for

22)return θ^*, Y_{pred}

2 实验设置与结果分析

2.1 数据集

本文使用ChestXRay2017^[27]、VinDr-PCXR^[28]和Curated X-Ray^[29]3个公开数据集。

ChestXRay2017数据集基于广州妇女儿童医学中心1~5岁儿科患者的X射线扫描数据库制作,共包含5856张标记为正常和肺炎两个类别的CXR图像,图像格式为JPEG。根据数据集中的名称标注再将其分为正常肺、细菌性肺炎和病毒性肺炎。因该数据集被广泛用于儿童肺炎分类任务^[5],故将其作为基准数据集来验证模型的基础性能和各模块的有效性。

VinDr-PCXR和Curated X-Ray数据集分别在诊断任务的类别和数据来源上有所不同,能够更真实地评估模型在复杂应用场景下的稳健性与适应性,用于进一步验证模型的泛化能力。其中,VinDr-PCXR是一个开放的、大规模的儿科CXR数据集,用于儿童常见胸部疾病的辅助诊断。该数据集包含9125张CXR图像,于2020—2021年从越南一家儿科医院收集。该数据集被标记为存在36项关键发现和15种疾病。为进行儿童肺炎的诊断,本文从中选取正常、肺炎和支气管肺炎3种类别构建新的数据集,各包含528、450和629张CXR图像。Curated X-Ray是通过整理15个公开可用的数据集所得到的联合数据集,混合有儿童和成人肺炎。该数据集包含1281张2019冠状病毒病CXR图像、3270张正常CXR图像、1656张病毒性肺炎CXR图像和3001张细菌性肺炎CXR图像。本文选取其中正常、病毒性肺

炎和细菌性肺炎构建新数据集。

2.2 实验环境

模型的训练在配备了 NVIDIA RTX4060Ti GPU 和 Intel 12600KF 处理器的个人计算机上进行,运行环境为 Windows 11 操作系统,使用 Python 语言编程,基于 PyTorch 框架实现。本文以 8:1:1 的比例划分数据集进行训练、验证和测试。另外本文将输入图像统一缩放为 224 像素×224 像素,并使用随机翻转、中心裁剪和亮度对比度调整操作对其进行数据增强。

2.3 评价指标

本文使用准确率(accuracy, Acc) A_{accuracy} 、精确率(precision, Pre) P_{recision} 、召回率(recall, Rec) R_{ecall} 、F1 分数和 AUC(area under the curve) 值对模型进行量化评价。AUC 表示受试者工作特征曲线(receiver operating characteristic, ROC)下的面积,是刻画模型阈值无关性分类能力的关键指标,其值越接近 1,模型区分样本的性能越出色。各评价指标的计算公式如表 2 所示。其中, N_{TP} 是真正例的数量, N_{FP} 是假正例的数量, N_{TN} 是真负例的数量, N_{FN} 是假负例的数量。

表 2 评价指标
Table 2 Evaluation metric

指标	描述
A_{accuracy}	$\frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FP}} + N_{\text{FN}}}$
P_{recision}	$\frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}$
R_{ecall}	$\frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}$
F_1	$\frac{2 \times P_{\text{recision}} \times R_{\text{ecall}}}{P_{\text{recision}} + R_{\text{ecall}}}$

表 3 在 ChestXRay2017 数据集上不同令牌混合器的组合效果
Table 3 Combination effects of different TokenMixers on the ChestXRay2017 dataset

令牌混合器组合	各阶段使用的令牌混合器		Acc(↑)	F_1 (↑)
	Stage1~2	Stage3~4		
SePConv+SA	SePConv	SA	84.29	81.70
SAMU+SA	SAMU	SA	84.46	81.77
SePConv+MAFU	SePConv	MAFU	85.26	82.36
4×SAMU	SAMU	SAMU	86.54	84.28
4×MAFU	MAFU	MAFU	83.65	79.83
SAMU+MAFU	SAMU	MAFU	88.94	86.77

注:加粗表示最优结果,下划线表示次优结果。

为进一步探究模型挖掘细微纹理与全局病理模式的机理,本文对比了 SA 与 MAFU 的跨阶段

2.4 消融实验

为评估本文模型各组件的有效性,本文在 ChestXRay2017 数据集上进行了一系列消融研究。在这些实验中,主要采用 Acc 和 F_1 分数作为核心评价指标。Acc 能够直观反映模型的整体分类正确性,而 F_1 分数作为精确率和召回率的调和平均数,能够更均衡地评估模型在类别可能分布不均情况下的综合性能,从而对模型的有效性提供更全面的衡量。

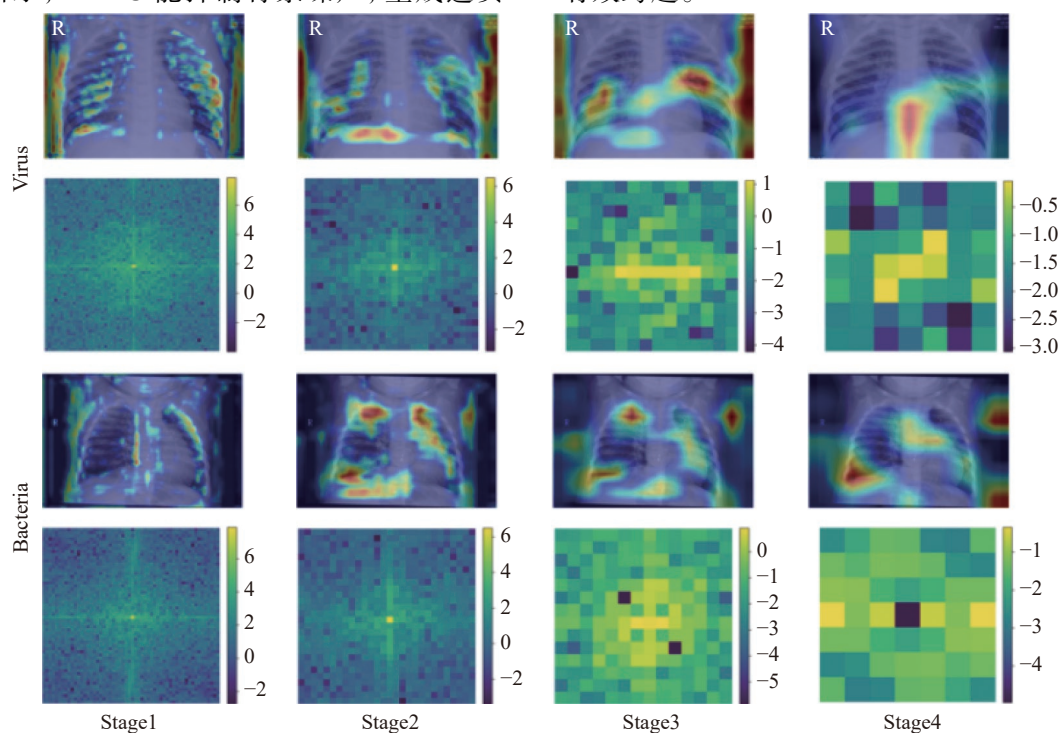
2.4.1 令牌混合器有效性分析

为验证不同令牌混合器对模型性能的影响,本文以可分离卷积(separable convolution, SeP-Conv)与标准自注意力(self attention, SA)构成的混合模型作为基准,设计了 5 组消融实验,结果如表 3 所示。实验结果显示,当将 SepConv 替换为本文的 SAMU 后,模型的性能实现了初步提升;进一步将 SA 替换为 SAMU 后,模型得到了较大的性能增益,在 Acc 和 F_1 上分别增长了 2.25% 和 2.58%。这充分证明了 SAMU 在捕获空间特征方面的强大能力。在 SePConv 加 MAFU 的组合中,Acc 和 F_1 分数较基准分别提升了 0.97% 和 0.66%,这不仅证明了 MAFU 的有效性,同时表明频域特征建模的引入能增强模型对不同类型肺炎的判别能力。在此基础上,将 Stage1~2 的 SePConv 替换为 SAMU,模型的性能得到了进一步提升,在所有组合中达到最佳效果,说明 SAMU 与 MAFU 在空间局部依赖与频域全局特征建模方面具有互补性。值得注意的是,仅使用 MAFU 的模型性能表现不佳,本文分析其可能原因在于 MAFU 作为全局建模单元,在模型中过早使用会破坏对后续识别任务重要的局部空间信息,从而限制了模型的表征能力。

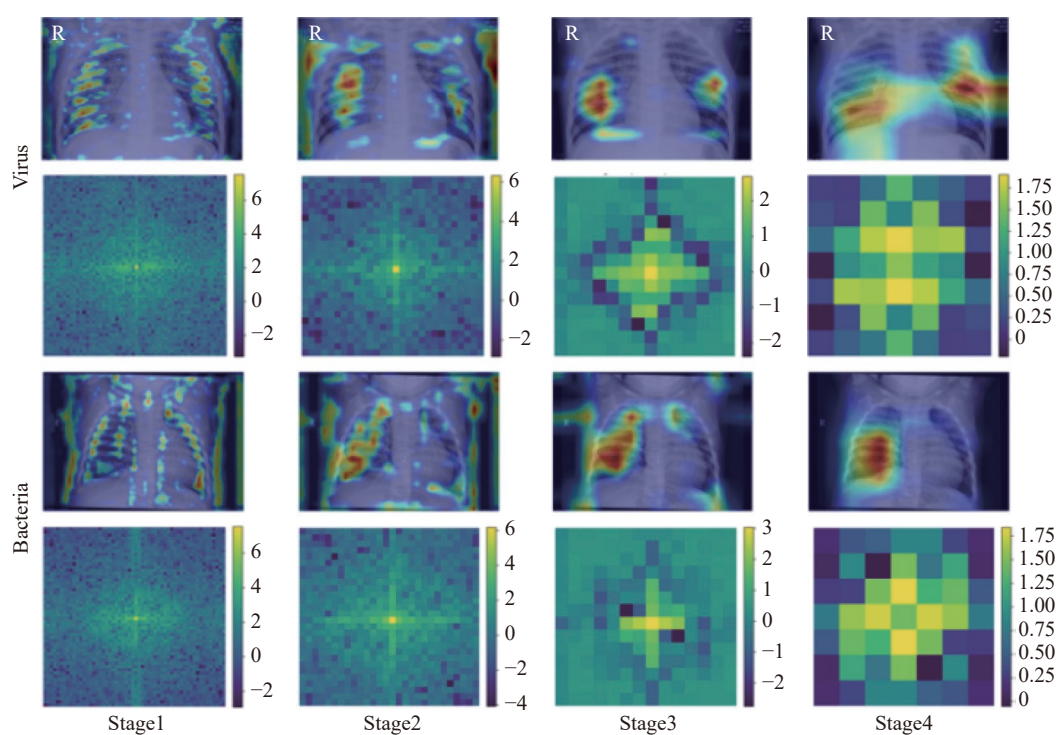
特征可视化,如图 6 所示。在浅层阶段 Stage1~2,在都使用 SAMU 的情况下两者表现一致,激活区

域沿病灶及纹理边缘呈离散碎片状。这证实了SAMU模块能有效捕捉细菌性实变边界与病毒性早期纹理等细粒度特征。进入深层阶段Stage3~4,表现有所分化。如图6(a)所示,SA受限于对高对比度特征的依赖,易受骨骼结构干扰,难以覆盖病毒性肺炎的弥漫性磨玻璃影。相比之下,如图6(b)所示,MAFU能抑制背景噪声,生成连贯

的团块状激活,精准定位细菌性实变与病毒性弥漫病灶。频谱分析进一步揭示,MAFU输出的能量高度集中于代表主体语义的低中频区域,有效抑制了SA中的高频噪声。这证明了空间-频域协同架构比单纯空间注意力具有更强的语义筛选与抗干扰能力,实现了从局部细节到全局模式的有效跨越。



(a) SA跨阶段特征可视化



(b) MAUFU跨阶段特征可视化

图6 SA与MAFU的跨阶段特征可视化

Fig. 6 Cross-stage feature visualisation of SA and MAUFU

2.4.2 通道混合器有效性分析

为评估 MK-CGLU 的有效性, 本文将其与不同通道混合器进行了对比分析。如表 4 所示, 本文以基础的 MLP 作为基线。将 MLP 替换为 CGLU 后, 模型的 Acc 和 F_1 分数均获得了明显提升, 验证了门控机制对特征筛选的有效性。进一步采用改进的 MK-CGLU 后, 性能达到最佳, 相较于 CGLU 其 Acc 和 F_1 分数分别提升了 1.12% 和 1.54%。实验结果表明多核卷积门控结构在捕获复杂特征关系上更具优势, 能够显著提升模型的整体性能。

表 4 不同通道混合器的效果
Table 4 The effectiveness of different ChannelMixers %

通道混合器	Acc(↑)	F_1 (↑)
MLP	87.02	84.51
CGLU	<u>87.82</u>	<u>85.23</u>
MK-CGLU	88.94	86.77

注: 加粗表示最优结果, 下划线表示次优结果。

2.4.3 多轴特征建模的影响评估

为验证多轴特征建模中不同轴向组合的有效性, 本文进行了消融实验, 其结果如表 5 所示。实验以仅包含空间交互的 HW 轴作为基线模型, 在此基础上, 当引入 CW 轴后, 模型性能得到较大增强, Acc 和 F_1 分数的增幅分别达 1.44% 和 1.42%。同样, 引入 CH 轴的组合也带来了性能增长, 但其提升幅度略低于前者, 这表明通道与宽度维度的特征交互对本任务的贡献更为关键。最终, 当 HW、CW、CH 三轴协同建模时, 模型性能达到最优, 较基线模型, Acc 和 F_1 分数实现了 3.20% 和 3.95% 的显著增长。实验结果证明, 三轴联合的建模方式能够最全面地捕获 CXR 图像中空间与通道维度的多维依赖关系, 通过有效的

特征互补, 从而最大化地提升了儿童肺炎分类的准确性。

表 5 多轴特征建模的性能评估
Table 5 Performance evaluation of multi-axis feature modeling %

HW	CW	CH	Acc(↑)	F_1 (↑)
√			85.74	82.82
√	√		<u>87.18</u>	<u>84.24</u>
√		√	86.54	83.89
√	√	√	88.94	86.77

注: 加粗表示最优结果, 下划线表示次优结果。

2.5 对比实验结果分析

为验证 SMFDFormer 的基础性能, 本文在 ChestXRay2017 数据集上进行了实验, 并与多个主流分类网络进行了对比分析。这些对比网络覆盖了不同的发展阶段与结构范式, 包括经典 CNN 架构及其变体 ResNet50^[30]、DenseNet121^[31]、EfficientNetv2^[32]、MobileNetv3^[33]、Auto COVID19^[34] 与 SupEx^[35]; 现代 CNN 与特征增强架构 ConvNext^[36]、LSNet^[37] 和 TSDFS^[38]; 基于 Transformer 或 CNN 与 Transformer 的混合架构 Swin Transformer^[39]、Mobile-ViT^[40]、Caformer^[41] 和 TransXNet^[42]。

2.5.1 ChestXRay2017 数据集上的对比实验

在 ChestXRay2017 数据集上, 本文首先进行了是否有肺炎的二分类任务对比。如表 6 所示, SMFDFormer 在 Acc 和 F_1 分数两项指标中均取得最优值, 另外 3 个指标也取得次优结果, 整体性能优于其他对比模型。这表明模型在整体上对儿童肺炎的判断最为准确, 并且在精确率和召回率之间取得了更好的平衡。

表 6 ChestXRay2017 数据集上的性能 (正常 vs 肺炎)
Table 6 Performance on the ChestXRay2017 dataset (normal vs. pneumonia) %

方法	Acc(↑)	Pre(↑)	Rec(↑)	F_1 (↑)	AUC(↑)
ResNet50 ^[30]	95.03	99.51	87.18	92.94	99.68
DenseNet121 ^[31]	96.63	99.53	91.45	95.32	99.75
EfficientNetv2 ^[32]	93.75	98.51	84.62	91.03	99.33
MobileNetv3 ^[33]	92.15	99.47	79.49	88.36	99.37
Auto COVID19 ^[34]	96.79	98.19	93.16	95.61	99.68
SupEx ^[35]	97.28	96.97	95.73	96.34	99.70
ConvNext ^[36]	93.27	94.16	95.13	94.64	99.03
LSNet ^[37]	93.59	94.91	87.61	91.11	98.79
TSDFS ^[38]	<u>97.60</u>	97.82	95.73	<u>96.76</u>	99.26
Pfree ^[12]	96.96	97.36	<u>94.44</u>	95.88	99.81

续表 6

方法	Acc(↑)	Pre(↑)	Rec(↑)	F_1 (↑)	AUC(↑)
SAGCN ^[13]	94.23	94.19	90.17	92.14	97.79
SwinTransformer ^[39]	91.19	97.35	78.63	86.99	98.76
Mobile-ViT ^[40]	92.63	99.47	80.77	89.15	99.65
Caformer ^[41]	92.47	98.45	81.20	88.99	99.14
TransXNet ^[42]	94.71	97.63	88.03	92.58	99.56
Medvit ^[18]	97.59	100.00	93.59	96.69	99.93
LungConvT ^[19]	96.63	97.76	93.16	95.40	99.82
SMFDFormer	97.76	99.55	94.44	96.93	99.86

注:加粗表示最优结果,下划线表示次优结果。

其次,在正常、细菌性肺炎和病毒性肺炎三分类任务中,表7结果显示,SMFDFormer的性能超越了包括Medvit、TransXNet以及LungConvT在内的所有对比模型。其Acc比表现次优的LungConvT和Pfree分别高出2.72%和2.73%,同时 F_1 分数也分别高出3.03%和2.75%,这主要得益于SMFDFormer通过浅层的SAMU有效捕捉细菌性肺炎的局部细节,同时通过深层的MAFU

在频域中高效建模病毒性肺炎的全局依赖关系。这种针对性的设计,使其在区分这两种病理模式时,比单纯依赖空间域卷积或注意力机制的模型更具优势。此外,SMFDFormer的AUC虽然略低于Mobilenetv3,但Mobilenetv3的 F_1 分数较低,表明SMFDFormer的整体平衡更好,验证了其在区分细菌性肺炎和病毒性肺炎方面的突出能力。

表7 ChestXRay2017数据集上的性能(正常vs细菌性肺炎vs病毒性肺炎)
Table 7 Performance on the ChestXRay2017 dataset (normal vs. bacterial pneumonia vs. viral pneumonia)

方法	Acc(↑)	Pre(↑)	Rec(↑)	F_1 (↑)	AUC(↑)	%
ResNet50 ^[30]	83.49	80.26	80.13	80.19	94.10	
DenseNet121 ^[31]	85.42	82.49	81.97	82.19	94.86	
EfficientNetv2 ^[32]	84.94	81.94	81.57	81.75	93.53	
MobileNetv3 ^[33]	84.46	81.78	80.03	80.65	96.12	
Auto COVID19 ^[34]	84.78	82.74	81.21	81.82	94.75	
SupEx ^[35]	84.78	81.73	81.26	81.47	93.90	
ConvNext ^[36]	84.94	81.94	81.46	81.66	94.27	
LSNet ^[37]	83.97	81.33	80.72	80.96	94.21	
TSDFS ^[38]	85.42	82.94	83.09	83.01	94.69	
Pfree ^[12]	86.21	<u>83.76</u>	84.38	<u>84.02</u>	95.01	
SAGCN ^[13]	82.37	79.31	77.48	78.01	90.28	
SwinTransformer ^[39]	82.53	79.51	78.20	78.69	93.52	
Mobile-ViT ^[40]	84.29	81.09	79.90	80.32	94.34	
Caformer ^[41]	84.46	81.39	81.16	81.27	94.46	
TransXNet ^[42]	86.06	83.42	<u>84.50</u>	83.81	<u>95.85</u>	
Medvit ^[18]	84.94	81.70	81.07	81.33	95.29	
LungConvT ^[19]	<u>86.22</u>	83.53	84.02	83.74	95.56	
SMFDFormer	88.94	86.77	86.77	86.77	94.93	

注:加粗表示最优结果,下划线表示次优结果。

为了更细致地探究模型对不同病理特征的区分能力,本文绘制了SMFDFormer与次优对比模型的混淆矩阵,如图7所示。针对以细微纹理特征为核心的细菌性肺炎,SMFDFormer实现了最

高的识别精度,优于侧重全局建模但忽略局部细节的TransXNet以及LungConvT。这验证了浅层SAMU模块通过空间位移和多尺度卷积操作,有效保留了诊断所需的微小病灶纹理。针对以全局

弥漫性特征为核心的病毒性肺炎,传统的 CNN 模型,如 DenseNet121 等因感受野受限,识别率较低。而 SMFDFormer 得益于深层 MAFU 模块的频域全局感知能力,成功识别了 96 例病毒性肺炎样

本,缓解了对弥散性病变的漏诊。混淆矩阵的结果有力证明了 SMFDFormer 通过空频协同设计,成功兼顾了局部纹理与全局依赖的捕捉,缓解了单一架构模型在特征识别上存在的局限性。

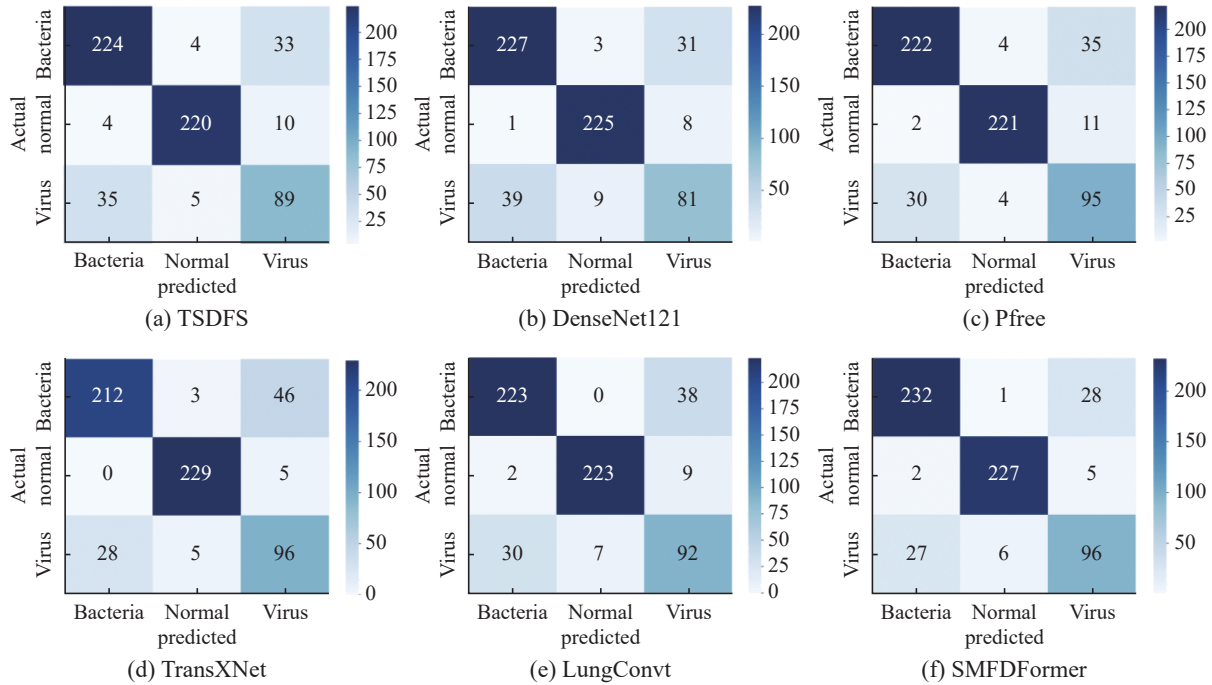


图 7 混淆矩阵对比

Fig. 7 Confusion Matrix Comparison

2.5.2 VinDr-PCXR 数据集上的对比实验

在 VinDr-PCXR 数据集上的实验结果如表 8 所示。SMFDFormer 的平均 AUC(mAUC)在所有对比模型中表现优异,仅次于 Medvit。值得注意的是,在肺炎类别诊断上,本模型的 AUC 值取得了所有模型中的最佳性能,这证明了本文提出的频域分析方法对于识别具有大范围病变特征的肺

炎的敏感性。对于正常类别 AUC 相对较低,这可能是因为模型在学习过程中更加关注于捕捉与病理相关的显著频域和空间特征,从而对正常肺部中存在的个体性细微变异的表征能力稍弱。这在一定程度上反映了模型设计的侧重点,即提升对关键病灶的特异性识别能力,这在临床辅助诊断中具有重要价值。

表 8 VinDr-PCXR 数据集上的性能
Table 8 Performance on the VinDr-PCXR dataset

%

方法	mAUC	Nofinding	Pneumonia	Broncho_pneumonia
ResNet50 ^[30]	67.01	78.12	63.94	58.96
DenseNet121 ^[31]	65.04	73.11	59.53	62.09
EfficientNetv2 ^[32]	67.83	76.02	65.23	62.23
MobileNetv3 ^[33]	67.20	74.86	63.63	63.12
Auto COVID19 ^[34]	68.11	77.09	64.23	63.01
SupEx ^[35]	65.80	73.98	62.88	60.54
ConvNext ^[36]	64.90	71.91	60.24	62.48
LSNet ^[37]	63.15	69.18	59.09	61.18
TSDFS ^[38]	64.62	72.81	64.83	56.22
Pfree ^[12]	68.79	77.92	67.87	60.59
SAGCN ^[13]	66.12	74.12	65.63	58.62
SwinTransformer ^[39]	60.16	66.79	51.06	62.64
Mobile-ViT ^[40]	67.01	78.12	63.94	58.96

续表 8

方法	mAUC	Nofinding	Pneumonia	Broncho_pneumonia
Caformer ^[41]	68.55	<u>78.52</u>	65.61	61.52
TransXNet ^[42]	68.46	80.19	66.78	58.41
Medvit ^[18]	71.44	74.18	<u>68.63</u>	71.51
LungConvT ^[19]	68.79	78.45	64.89	63.03
SMFDFormer	69.05	74.51	69.05	<u>63.58</u>

注: 加粗表示最优结果, 下划线表示次优结果。

2.5.3 Curated X-Ray 数据集上的对比实验

在 Curated X-Ray 这一更具挑战性的多源异构数据集上的实验结果如表 9 所示。SMFDFormer 的 Acc、Pre 和 F_1 这 3 项核心指标均在所有对比模型中位列第一, 全面超越了所有的对比模型。相较于表现同样优异的 Medvit, SMFDFormer 在 F_1 分数上领先了 0.85%, 显示出更均衡的分类性能。对于 AUC 指标略低于 Auto COV-

ID19 和 Medvit 的情况, 本文分析这可能是由于 Curated X-Ray 数据集整合自 15 个不同的来源, 图像异质性引入的噪声可能对频域变换的稳定性造成了轻微干扰。尽管如此, SMFDFormer 在分类准确性和平衡性上的全面领先, 依然证实了空间-频域协同建模策略在应对复杂数据时具有强大的泛化能力和稳健性。

表 9 Curated X-Ray 数据集上的性能
Table 9 Performance on the Curated X-Ray dataset

%

方法	Acc(↑)	Pre(↑)	Rec(↑)	F_1 (↑)	AUC(↑)
ResNet50 ^[30]	87.75	85.63	84.02	84.61	96.21
DenseNet121 ^[31]	88.89	86.59	86.28	86.42	96.37
EfficientNetv2 ^[32]	88.51	86.33	85.13	85.61	96.53
MobileNetv3 ^[33]	87.12	84.43	83.90	84.14	96.07
Auto COVID19 ^[34]	89.39	87.58	86.72	87.09	96.87
SupEx ^[35]	84.85	81.66	80.64	81.03	94.64
ConvNext ^[36]	83.08	79.88	78.04	78.55	93.72
LSNet ^[37]	84.47	81.94	79.63	80.29	93.99
TSDFS ^[38]	90.91	89.05	88.20	88.57	96.24
Pfree ^[12]	89.39	87.15	86.82	86.96	96.58
SAGCN ^[13]	83.71	80.95	77.96	78.58	92.84
SwinTransformer ^[39]	85.23	82.37	80.26	80.86	94.60
Mobile-ViT ^[40]	86.87	84.23	82.77	83.30	95.67
Caformer ^[41]	84.22	80.88	79.34	79.82	94.73
TransXNet ^[42]	89.65	88.02	86.50	87.10	96.02
Medvit ^[18]	<u>91.04</u>	<u>89.09</u>	<u>88.97</u>	<u>89.03</u>	<u>96.81</u>
LungConvT ^[19]	89.27	87.03	86.14	86.52	96.77
SMFDFormer	91.79	90.62	89.72	89.88	96.74

注: 加粗表示最优结果, 下划线表示次优结果。

2.5.4 可视化分析

为了直观体现 SMFDFormer 模型在特征提取上的优势, 本文将其与 10 个模型在各个数据集上的分类结果进行了可视化对比, 结果如图 8 所示。在 ChestXRay2017 数据集中, 针对代表细微纹理特征的细菌性肺炎, 其病理常表现为局部的肺叶实变或支气管充气征。如图 8 第 1 行所示, SMFDFormer 的激活区域呈现出高度的局部聚焦特性,

精准锁定了右肺的实变核心, 且边缘轮廓清晰。相比之下, ResNet50 等 CNN 模型的激活区域过于弥散, 未能准确勾勒病灶边界。这一结果直观验证了浅层 SAMU 模块通过空间位移与多尺度卷积, 成功捕捉到了诊断所需的细粒度纹理信息。针对代表全局病理特征的病毒性肺炎, 其病理常表现为双肺弥漫性的磨玻璃影或间质性改变。如图 8 第 2 行所示, SMFDFormer 的热图呈现出显著的全

局覆盖特性,在双肺形成了均匀且连贯的激活分布,完整覆盖了弥散性病变区域。反观 LungConvT 等混合架构,其热图多呈团块状或零散分布,难以

体现病变的整体性。这一结果有力证明了深层 MAFU 模块利用多轴频域分析,有效突破了局部感受野的限制,实现了对全局病理依赖关系的精准建模。

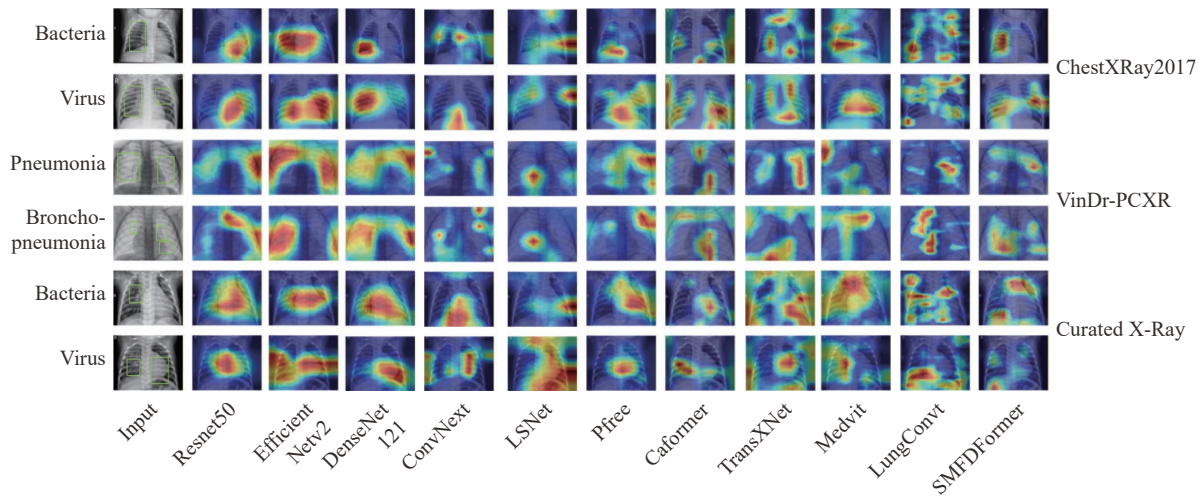


图 8 在不同数据集上的可视化

Fig. 8 Visualization on the different datasets

在 Vindr-PCXR 数据集上的热图对比中,可以看到支气管肺炎的特点为沿支气管分布的小片状炎症。本文模型生成的激活热图呈现出符合其病理特征的、遍布双肺的散在小片状高亮区域。反观对比模型, EfficientNetV2 等传统 CNN 模型的激活过于集中, ConvNext、LSNet 和 LungConvT 的热图激活区域零散且缺乏规律性。而 Medvit 的激活区域呈现出不合常理的“T”字形,其注意力被胸骨、锁骨等非相关的骨骼结构严重干扰。

在 Curated X-Ray 数据集上不同肺炎样本的热图对比,无论是对于细菌性肺炎的局部实变,还是病毒性肺炎的弥漫性改变, SMFDFormer 始终能生成最贴合临床病理表现的热图。其对局部病灶的定位精准度,以及对全局病变分布的覆盖完整性,均稳定地优于其他基线模型,后者则重复表现出注意力弥散、定位不准或全局感知缺失等问题。这进一步验证了 SMFDFormer 的鲁棒性和泛化能力。

3 结束语

本文针对儿童肺炎 CXR 图像复杂多变、局限于空间域特征提取、难以高效兼顾细微纹理和全局病理特征等问题,提出了一种名为 SMFDFormer 的分类模型。该模型通过分阶段混合范式,系统性地融合了空间域与频域信息,旨在提升诊断的准确性与鲁棒性。其中, SAMU 模块通过零参数的通道位移与多尺度卷积相结合,专注于高效捕捉区分不同肺炎类型所需的细粒度病理

纹理; MAFU 模块通过在多个特征平面上进行傅里叶变换和动态门控处理,有效地捕捉了肺部病变的全局上下文和长距离依赖关系;同时, MK-CGLU 作为一个高效的通道混合器,通过多核卷积门控机制自适应地筛选和增强关键病理区域的特征,克服了传统门控单元在感受野尺度上的局限性,提升了模型对复杂病理模式的识别能力。

SMFDFormer 展现了良好的性能,但仍存在可探索的改进方向。未来的研究可集中于两个方面:一方面,进一步优化模型的轻量化设计,在保持高性能的同时降低计算资源消耗,以更好地适应资源受限的基层医疗环境;另一方面,探索将模型的可解释性方法与频域分析相结合,以可视化方式揭示模型决策的关键频率成分,从而增强临床医生对模型诊断结果的信任度。

参考文献:

- [1] ZHENG Jingyuan, LIU Dingming, YANG Zhifu, et al. SEACC: self-evolving and adaptive contrastive learning for classification of pediatric pneumonia and anteroposterior/posteroanterior chest radiographs[J]. *Displays*, 2024, 81: 102607.
- [2] 吕佳, 于明楷, 陈欣, 等. 差异感知引导的边界 Transformer 儿童肺炎 CT 图像分割网络[J]. *光学学报*, 2025, 45(15): 112-136.
LÜ Jia, YU Mingkai, CHEN Xin, et al. Difference aware guided boundary Transformer network for childhood-Pneumonia CT image segmentation[J]. *Acta optica sinica*, 2025, 45(15): 112-136.
- [3] KAYA M. Feature fusion-based ensemble CNN learning

- optimization for automated detection of pediatric pneumonia[J]. *Biomedical signal processing and control*, 2024, 87: 105472.
- [4] 负恺, 贾荣浩, 魏国辉, 等. 基于 CNN 与 Transformer 混合模型的肺炎辅助诊断[J]. *计算机系统应用*, 2025, 34(2): 216–224.
- YUN Kai, JIA Ronghao, WEI Guowei, et al. Pneumonia assisted diagnosis based on hybrid model of CNN and Transformer[J]. *Computer systems and applications*, 2025, 34(2): 216–224.
- [5] RICKARD D, KABIR M A, HOMAIRA N. Machine learning-based approaches for distinguishing viral and bacterial pneumonia in paediatrics: a scoping review[J]. *Computer methods and programs in biomedicine*, 2025, 268: 108802.
- [6] 程文娟, 于国庆. 基于 XDense-RC-NET 的 CXR 图像分类算法[J]. *计算机应用研究*, 2022, 39(12): 3803–3807.
- CHENG Wenjun, YU Guoqing. Improved CXR image classification algorithm based on XDense-RC-NET[J]. *Application research of computers*, 2022, 39(12): 3803–3807.
- [7] KHAN W, ZAKI N, ALI L. Intelligent pneumonia identification from chest X-rays: a systematic literature review [J]. *IEEE access*, 2021, 9: 51747–51771.
- [8] 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.
- [9] RAJPURKAR P, IRVIN J, ZHU K, et al. Chexnet: radiologist-level pneumonia detection on chest X-rays with deep learning[EB/OL]. (2017–12–25)[2025–08–21]. <https://arxiv.org/abs/1711.05225>.
- [10] 潘伟, 马志庆, 魏国辉, 等. 卷积神经网络在肺炎图像分类中的研究进展与展望[J]. *中国医疗设备*, 2025, 40(7): 147–154.
- PAN Wei, MA Zhiqing, WEI Guohui, et al. Research progress and prospect of convolutional neural networks in pneumonia image classification[J]. *China medical devices*, 2025, 40(7): 147–154.
- [11] HAMMOUDI K, BENHABILES H, MELKEMI M, et al. Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19[J]. *Journal of medical systems*, 2021, 45(7): 75.
- [12] DU Jie, GUAN Kai, ZHOU Yanhong, et al. Parameter-free similarity-aware attention module for medical image classification and segmentation[J]. *IEEE transactions on emerging topics in computational intelligence*, 2023, 7(3): 845–857.
- [13] 周涛, 叶鑫宇, 陆惠玲, 等. 基于多分辨率注意密集网络的肺炎分类识别方法[J]. *模式识别与人工智能*, 2023, 36(6): 544–555.
- ZHOU Tao, YE Xinyu, LU Huiling, et al. Pneumonia classification and recognition method based on multi-resolution attention dense network[J]. *Pattern recognition and artificial intelligence*, 2023, 36(6): 544–555.
- [14] SUN Junding, XUE Jianxiang, XU Zhaozhao, et al. SAGCN: self-adaptive graph convolutional network for pneumonia detection[J]. *Biomedical signal processing and control*, 2025, 106: 107634.
- [15] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[C]//International Conference on Learning Representations. Online: ICLR Foundation, 2021: 1–22.
- [16] OKOLO G I, KATSIGIANNIS S, RAMZAN N. IEViT: an enhanced vision transformer architecture for chest X-ray image classification[J]. *Computer methods and programs in biomedicine*, 2022, 226: 107141.
- [17] UKWOMA C C, QIN Zhiguang, BELAL BIN HEYAT M, et al. A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images[J]. *Journal of advanced research*, 2023, 48: 191–211.
- [18] MANZARI O N, AHMADABADI H, KASHIANI H, et al. MedViT: a robust vision transformer for generalized medical image classification[J]. *Computers in biology and medicine*, 2023, 157: 106791.
- [19] LASKER A, GHOSH M, SK M O, et al. LungConVT-Net: a visual transformer network with blended features for Pneumonia detection[J]. *Pattern recognition*, 2026, 171: 112150.
- [20] 王梦溪, 雷涛, 姜由涛, 等. 基于空频协同的 CNN-Transformer 多器官分割网络[J]. *智能系统学报*, 2025, 20(5): 1266–1280.
- WANG Mengxi, LEI Tao, JIANG Youtao, et al. CNN-transformer multiorgan segmentation network based on space-frequency collaboration[J]. *CAAI transactions on intelligent systems*, 2025, 20(5): 1266–1280.
- [21] 葛斌, 卞文强, 夏晨星, 等. 基于空间频域交互感知的医学图像分割[J/OL]. *光电子·激光* (2025–09–26) [2025–11–26]. <https://link.cnki.net/urlid/12.1182.O4.20250926.1622.022>.
- GE Bin, BIAN Wenqiang, XIA Chenxing, et al. Medical image segmentation based on spatial frequency domain interactive perception[J/OL]. *Journal of optoelectronics·laser* (2025–09–26) [2025–11–26]. <https://link.cnki.net/urlid/12.1182.O4.20250926.1622.022>.
- [22] RUAN Jiacheng, GAO Jingsheng, XIE Mingye, et al. Learning multi-axis representation in frequency domain for medical image segmentation[J]. *Machine learning*, 2025, 114(1): 10.
- [23] WANG Wenxuan, WANG Jing, CHEN Chen, et al. Fre-MIM: fourier transform meets masked image modeling for medical image segmentation[C]//2024 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2024.
- [24] GAO Tian, ZHANG Yu, ZHANG Zhiyuan, et al. BHViT: binarized hybrid vision transformer[C]//2025 IEEE/CVF

- Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2025.
- [25] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016.
- [26] SHI Dai. TransNeXt: robust foveal visual perception for vision transformers[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024.
- [27] KERMANY D S, GOLDBAUM M, CAI Wenjia, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning[J]. *Cell*, 2018, 172(5): 1122-1131. e9.
- [28] PHAM H H, NGUYEN N H, TRAN T T, et al. PediCXR: an open, large-scale chest radiograph dataset for interpretation of common thoracic diseases in children[J]. *Scientific data*, 2023, 10: 240.
- [29] SAIT U, LAL K, PRAJAPATI S, et al. Curated dataset for covid-19 posterior-anterior chest radiography images (X-rays)[EB/OL]. (2020-09-15)[2025-08-21]. <https://data.mendeley.com/datasets/9xkhgts2s6/1>.
- [30] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016.
- [31] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017.
- [32] TAN M, LE Q. Efficientnetv2: smaller models and faster training[C]//Proceedings of the 38th International Conference on Machine Learning. Online: IMLS, 2021.
- [33] HOWARD A, SANDLER M, CHEN Bo, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019.
- [34] KUMAR V, GUPTA A, HAZARIKA B B, et al. Automatic diagnosis of COVID-19 from chest X-ray images using transfer learning-based deep features and machine learning models[J]. *China communications*, 2025, 22(7): 274-289.
- [35] KILIÇARSLAN S, KÖZKURT C, BAŞ S, et al. Detection and classification of pneumonia using novel superior exponential(SupEx) activation function in convolutional neural networks[J]. *Expert systems with applications*, 2023, 217: 119503.
- [36] LIU Zhuang, MAO Hanzhi, WU Chaoyuan, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022.
- [37] WANG Ao, CHEN Hui, LIN Zijia, et al. LSNet: see large, focus small[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2025.
- [38] OUYANG Haibin, LIU Dongmei, LI S, et al. Two-stage deep feature selection method using voting differential evolution algorithm for pneumonia detection from chest X-ray images[J]. *IEEE transactions on emerging topics in computational intelligence*, 2025, 9(1): 918-932.
- [39] LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021.
- [40] MEHTA S, RASTEGARI M. MobileViT: light-weight, general-purpose, and mobile-friendly uivision transformer[EB/OL]. (2021-10-05)[2025-10-17]. <https://arxiv.org/abs/2110.02178>.
- [41] YU Weihao, LUO Mi, ZHOU Pan, et al. MetaFormer is actually what you need for vision[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022.
- [42] LOU Meng, ZHANG Shu, ZHOU Hongyu, et al. TransXNet: learning both global and local dynamics with a dual dynamic token mixer for visual recognition[J]. *IEEE transactions on neural networks and learning systems*, 2025, 36(6): 11534-11547.

作者简介:



吕佳, 教授, 博士, 主要研究方向为机器学习、数据挖掘及其在医学图像处理中的应用。主持或参与国家、省部级科研项目 20 余项, 发表学术论文 70 余篇。E-mail: lvjia@cqnu.edu.cn。



龚泉, 硕士研究生, 主要研究方向为深度学习及其在医学图像处理中的应用。E-mail: 2024110516036@stu.cqnu.edu.cn。



何玲, 主任医师, 博士, 主要研究方向为儿科影像、人工智能和影像诊断。E-mail: heling508@sina.com。