

DOI:10.19651/j.cnki.emt.2519242

注意力与多尺度特征融合的胸部疾病分类*

胡欣茹¹ 吕晓琪^{1,2} 谷宇^{1,3}(1. 内蒙古科技大学数智产业学院 包头 014010; 2. 内蒙古工业大学信息工程学院 呼和浩特 010051;
3. 河北建筑工程学院信息工程学院 张家口 075000)

摘要: 胸部疾病在早期诊断中具有重要意义,现有 X 光图像分类方法因特征提取中信息交互不足以及微小病灶辨识困难等问题导致分类效果不佳。为此提出一种基于注意力机制与多尺度特征融合的胸部 X 光图像疾病分类网络 FFA-Net。首先,该网络通过任务交叉注意模块在水平和垂直方向上有效捕获全局上下文信息,增强特征间的交互;其次,通过构建多分支提取模块,在不同尺度下融合特征信息,使其深层特征能够聚焦于浅层特征中识别出的细微病理区域;最后,设计了一个能够提取不同频率特征并抑制噪声干扰的多频率语义注意力模块。在 ChestX-ray14 数据集上对所提出的方法进行了综合实验,结果显示,平均 AUC 值为 0.856 4,对疝气的 AUC 值达到 0.973 4;并通过消融实验以及在 CheXpert 和 COVID-19 Radiography Database 两个数据集进行泛化实验,数据显示,在 CheXpert 数据集上的平均 AUC 值为 0.811;在 COVID-19 Radiography Database 数据集上的平均 Accuracy 为 0.956 0。相较于当前流行的分类网络,FFA-Net 具备更好的特征提取能力和分类效果。

关键词: 医学图像处理;胸部 X 光片;多标签图像分类;注意力机制;特征融合

中图分类号: TP391;TN911.73 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Attention and multiscale feature fusion for chest disease classification

Hu Xinru¹ Lyu Xiaoqi^{1,2} Gu Yu^{1,3}(1. School of Digital Intelligence Industry, Inner Mongolia University of Science and Technology, Baotou 014010, China;
2. College of Information Engineering, Inner Mongolia University of Technology, Hohhot 010051, China;
3. Information Engineering College, Hebei University of Architecture, Zhangjiakou 075000, China)

Abstract: Chest diseases are important in early diagnosis, and the existing X-ray image classification methods have poor classification results due to the problems of insufficient information interaction in feature extraction and difficulty in recognizing small lesions. To this end, a chest X-ray image disease classification network FFA-Net based on attention mechanism and multi-scale feature fusion is proposed. First, the network effectively captures the global context information in horizontal and vertical directions through task crossing attention module to enhance the interaction between features; second, the network fuses the feature information at different scales by constructing a multi-branch extraction module so that its deeper features can focus on the subtle pathology regions identified in the shallow features; finally, a multi-frequency semantic attention module. Comprehensive experiments on the proposed method were performed on the CheX-ray14 dataset, which showed a mean AUC value of 0.856 4 and an AUC value of 0.973 4 for hernias; and generalization experiments were performed by ablation experiments as well as on the two datasets, CheXpert and COVID-19 Radiography Database. The data show that the average AUC value on the CheXpert dataset is 0.811; the average Accuracy on the COVID-19 Radiography Database dataset is 0.956 0. Compared with the current popular classification networks, FFA-Net has better feature extraction ability and classification effect.

Keywords: medical image processing; chest X-ray; multi-label image classification; attention mechanism; feature fusion

0 引言

胸部 X 光片 (chestX-ray, CXR) 检查作为全球应用最

广泛的放射学筛查手段,因其低成本的优势已在世界范围内普及,报告指出^[1],每年全球进行的胸部 X 光检查超过 20 亿次,凸显了 CXR 在临床诊疗中的关键作用。然而,由

收稿日期:2025-07-01

* 基金项目:国家自然科学基金(62001255,61771266)、中央引导地方科技发展资金项目(2021ZY0004)、内蒙古自治区自然科学基金(2024MS06008)、内蒙古自治区高等学校青年科技英才计划(NJYT23057)、内蒙古自治区直属高校基本科研业务费项目优秀青年基金(042)项目资助

于胸部病变类型多样且病理特征复杂, CXR 的临床解读仍面临重大挑战, 目前主要依赖具有丰富经验的放射科医师进行专业判断。这一现状推动了 CXR 自动分析技术的快速发展, 使其逐渐成为临床诊断领域的重要研究方向。

近年来, 深度学习的兴起和迅猛发展极大地推动了计算机辅助诊断技术 (computer-aided diagnosis, CAD) 在医学领域的广泛应用。这一技术已被成功应用于多种医学场景, 包括皮肤病变的识别^[2]、乳腺癌的筛查^[3]、心率失常诊断^[4]以及脑电图的分类^[5]等。随着众多大规模 CXR 图像数据集的公开, 基于卷积神经网络 (convolution neural network, CNN) 的胸部 X 光图像疾病分类方法也随之出现, 并催生了大量相关研究。Wang 等^[6]发布了 ChestX-ray8 数据集, 并利用 4 种经典的卷积神经网络架构并对其性能进行了评估。Rajpurkar 等^[7]提出了 CheXNet 网络, 该网络在肺炎检测任务中表现优异。CheXNet 使用预训练的权重进行初始化, 并在 DenseNet121 网络的基础上进行微调实现。Wang 等^[8]提出了三重注意力学习 (A³Net) 模型, 该模型以预训练的 DenseNet121 作为特征提取的骨干网络, 并在统一框架中集成了 3 个注意力模块, 分别用于通道、元素和尺度的注意力学习。以上方法训练时使用经典网络独立处理特征图, 虽然能够提取特征, 但是没有考虑上下文信息之间的交互, 而 CheXNet 网络仅在 DenseNet121 网络基础上对全连接层进行微调, 忽略了网络其他层特征提供的信息。

此外, 许多研究人员将注意力机制也应用与胸部疾病检测领域。Chen 等^[9]提出了一种病变位置注意引导网络 (LLAGnet), 引入了两个注意力模块, 将区域级注意力和通道级注意力整合到一个统一的框架中, 用于 CXR 中的多标签胸部疾病分类, 平均 AUC 值达到 0.824。Zhu 等^[10]提出一种像素分类与注意力网络 (PCAN), 同时执行疾病分类与弱监督定位, 提供疾病分类的可解释性, 平均 AUC 值达到 0.824。邵凌云等^[11]提出了一种基于高效通道注意力的分类算法, 通过使用非对称卷积来增强网络的特征提取能力, 平均 AUC 值达到 0.824。以上研究聚焦于构建更复杂的注意力机制以优化模型性能, 虽提升了分类精度, 但依然存在不足, 如由于疾病病理特征复杂, 如不同疾病的尺寸差异使得模型难以有效捕捉不同尺度的特征, 尤其在复杂背景中检测微小异常时无法准确识别不同尺度大小的病变区域。

此外, 上述现有模型的较高识别率通常依赖于深度网络层次结构或相对复杂的网络设计, 并且它们在资源受限的医疗临床环境中通常表现不佳。由此, EfficientNetV2^[12]网络成为本研究的自然选择, EfficientNetV2 是基于 EfficientNet^[13]的改进网络。利用其神经网络搜索 (neural architecture search, NAS) 技术, 优化并确定了网络宽度、深度与分辨率之间的最佳复合缩放系数。该策略不仅使模型能够学习到更丰富、更复杂的特征表示, 还显著提升了模型的泛化能力

与训练稳定性。

目前, 许多基于 EfficientNetV2 改进的算法由于其高效和准确性被广泛应用于医学成像分析领域, 并且已经显示出显著结果。Pacal 等^[14]提出一种 EfficientNetV2 网络的变体网络, 集成两种不同的注意力机制, 以提高脑肿瘤分类任务的准确性, 为脑肿瘤分类设定了新的基准。Huang 等^[15]提出 LightEfficientNetV2 网络, 能高效准确地识别 COVID-19 疾病, 表现出较好性能。Zeng 等^[16]针对肺炎图像识别中病变区域形态复杂、识别模型部署硬件资源有限等问题, 提出了一种基于多尺度注意力的改进 EfficientNetV2 肺炎识别模型, 在公共数据集 ChestX-ray 的三重分类任务上表现较好, 然而, 这些模型普遍仅针对 3~4 种特定疾病进行优化与验证, 所解决的仍是一个相对受限的分类问题。相比之下, 真实的胸部疾病诊断场景涵盖的病症类型更为多样, 病理表现也错综复杂, 现有模型在此环境下的泛化能力尚未得到有效验证。

针对上述研究中存在的问题, 提出了一种基于注意力机制与多尺度特征融合的胸部 X 射线图像分类网络 (feature fusion attention network, FFA-Net), 旨在实现对 14 种胸部疾病的鉴别。首先, 在特征提取阶段引入任务交叉注意力模块 (task cross attention module, TCAM), 通过双轴注意力机制同步构建全局上下文感知, 在水平和垂直维度上建立像素级依赖关系, 实现跨维度特征交互。其次, 针对无法准确识别不同尺度大小的病变区域设计了一个多分支提取模块 (multi-branch extraction module, MBEM), 通过多分支并行架构实现跨尺度特征融合, 重点增强对微小病理区域的聚焦能力。最后, 构建了一个多频率语义注意力模块 (multi-frequency semantic attention module, MFSM), 基于二维离散余弦变换 (2-Dimensional discrete cosine transform, 2D DCT) 的频域分解理论实现高频边缘特征增强与低频语义特征的跨频段交互, 在抑制背景噪声干扰的同时强化特征表达, 以捕捉更复杂的边界信息。

1 FFA-Net 网络结构

所提的 FFA-Net 网络架构如图 1 所示, 先对 CXR 图像进行预处理, 输入网络进行特征提取与尺度变换, FFA-Net 以 EfficientNetV2 作为主干网络, 该网络包含 6 个层级 (stage), 前 3 个 stage 为 Fused-MBConv, 后 3 个 stage 为 MBConv, 各阶段空间分辨率逐渐下降。Fused-MBConv 的详细描述如图 1(a) 所示, Fused-MBConv1 与 Fused-MBConv4 的不同在于扩展比率 (expansion ratio) 的不同。同样地, MBConv 的详细描述如图 1(b) 所示, MBConv4 与 MBConv6 模块的不同也表示扩展比率的不同, SE 表示使用了 squeeze-and-excitation 模块。

之后网络经过任务交叉注意力模块, 多分支提取模块, 多频率语义注意力模块进一步优化特征: 具体来说, EfficientNetV2 经过多轮卷积操作后, 产生了不同尺度的

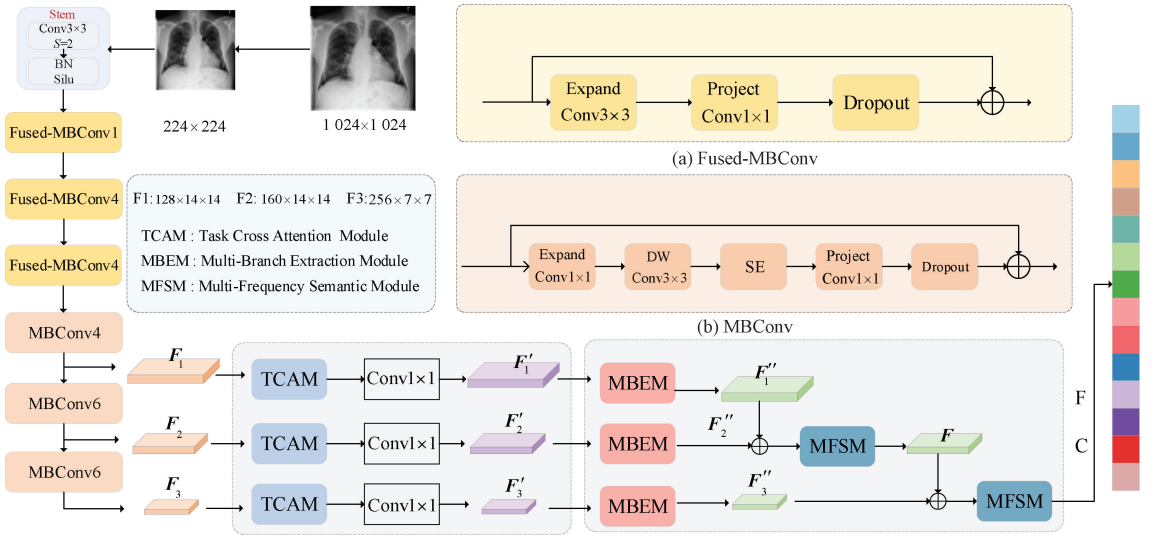


图1 FFA-Net 整体框架

Fig. 1 Overall framework of FFA-Net

特征图, 深层特征更能捕捉图像的全局信息。因此将网络最后 3 个 stage 产生的特征图提取出来, 得到 F_1, F_2, F_3 。其次经过 TCAM 模块, 捕获上下文信息, 得到 F'_1, F'_2, F'_3 ; 再通过 MBEM 模块提取不同尺度的特征得到 F''_1, F''_2, F''_3 , 将提取到的 F''_1 经过下采样与 F''_2 进行特征融合; 得到的特征通过 MFSM 模块得到 F , MFSM 能够提取不同频率和语义的特征, 再将 F 进行下采样与 F''_3 进行特征融合; 将得到的特征再次经过 MFSM 模块, 由此将得到的特征有效地融合。

最后, 在全连接 (fully connected, FC) 层中, 特征图进行 7×7 池化操作, 输出 14 维特征向量, 并输出相应的疾病分类结果, 从而完成整个训练过程。

1.1 任务交叉注意力模块

传统的 CNN 主要依赖局部卷积操作, 通过堆叠多层来扩大感受野, 但效率低且难以建模全局上下文。Wang 等^[17]针对这一问题提出了非局部 (non-local block) 模块, 计算特征图中所有位置之间的关系来捕获全局信息。受此启发, 本研究通过深入研究交叉注意力, 提出一种将交叉路径上每个像素的上下文信息聚合在一起的 TCAM 模块, 仅计算特征图中每个像素与其同一行和同一列上所有像素的关系, 逐步覆盖全局上下文, 然后将融合上下文信息的特征图与局部特征进行拼接, 在减少计算量的同时, 保持对全局上下文的有效建模。

如图 2 所示, 给定一个局部特征图 $H \in R^{H \times W \times C}$, 该模块先在 H 上应用 2 个 1×1 的卷积进行降维, 再分别经过 3×3 和 1×1 的深度可分离卷积提取特征, 得到两个特征图 Q 和 K , 其中 $\{Q, K\} \in R^{H \times W \times C'}$, C' 是降维之后的通道数。在得到 Q 和 K 之后, 进一步通过亲和力操作 (affinity operation) 来计算位置间的相关性权重生成注意力图 A , 然后在通道维度上应用 softmax 来计算注意力图 A 。其

中 $A \in R^{(H+W-1) \times (W \times H)}$, affinity 的定义如式(1)所示。

$$\rho_{i,u} = Q_u \Omega_{i,u}^T \quad (1)$$

式中: 向量 $Q_u \in R^{C'}$ 表示 Q 空间的每个位置 u ; 集合 $\Omega_u \in R^{(H+W-1) \times C'}$ 表示从 K 中提取与位置 u 处于同一行或同一列的特征向量; $\Omega_{i,u} \in R^{C'}$ 表示 Ω_u 中的第 i 个元素; $\rho_{i,u}$ 表示 Q_u 与 $\Omega_{i,u}$ 之间的相似度, 其中 $i = [1, \dots, H+W-1]$ 。

在 H 上应用另一个 1×1 卷积, 生成特征图 V , 其中 $V \in R^{H \times W \times C}$ 。上下文信息通过聚合操作 (aggregation operation) 聚合特征图中相关位置的信息生成 H'_u , Aggregation 定义如式(2)所示。

$$H'_u = \sum_{i=0}^{H+W-1} A_{i,u} \cdot \varepsilon_{i,u} \quad (2)$$

式中: H'_u 表示在特征图 H' 中位置 u 处的特征向量, $A_{i,u}$ 表示在位置 u 处对第 i 个通道或位置的关注程度, $\varepsilon_{i,u} \in R^{(H+W-1) \times C}$ 表示 V 中与 u 在同一行或同一列的特征向量的集合, $\varepsilon_{i,u} \in R^C$ 表示从位置 i 采样得到的特征向量。

然后将加权后的特征与输入特征图相加得到 H' , 如式(3)所示。

$$H' = \lambda \cdot (H'_u) + H \quad (3)$$

式中: λ 表示可学习参数, 初始值为 0。

在局部特征 H 中加入上下文信息增强了像素级表示, 因此它具有广泛的上下文视图, 并根据空间注意力图选择性地聚合上下文。为了获得更丰富、更密集的上下文信息, 将此模块进行串联。因此 H 中的每个位置实际上收集了来自所有像素的信息, 前后两个交叉注意力模块参数共享, 避免额外参数, 这些特征实现了相互增益, 并且对于分类任务更具鲁棒性。

1.2 多分支提取模块

由于某些胸部疾病的病理特征在尺度上存在差异, 而浅层网络对此类特征提取能力有限, 导致特征提取过程难

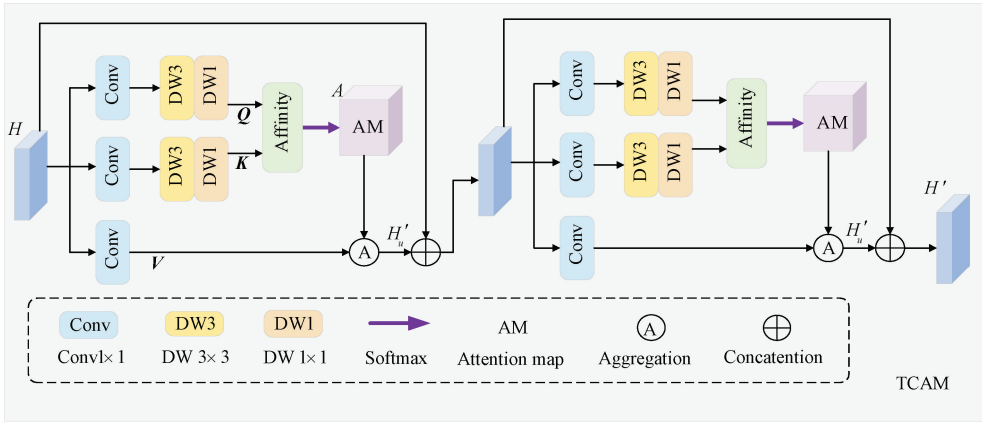


图 2 任务交叉注意力模块结构

Fig. 2 Task cross attention module structure

度较高。此外,在特征提取过程中,细小的病理特征容易丢失,从而影响模型对较小疾病的识别能力。由此,提出多分支提取 MBEM 模块,MBEM 模块是由局部感知提取

模块(patch-aware extraction modular, PAEM)和图通道增强模块(graph channel enhancement module, GCEM)两部分组成。其结构如图 3 所示。

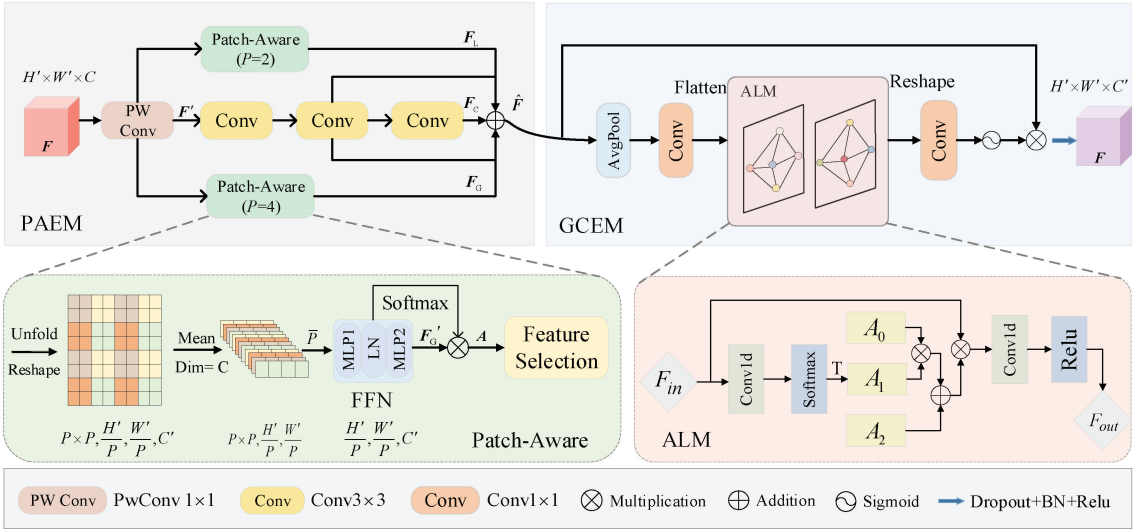


图 3 多分支提取模块结构

Fig. 3 Multi-branch extraction module structure

第 1 部分为 PAEM 模块,此模块采用并行结构,每个分支提取不同尺度和不同层次的特征,这种设计能够有效捕捉目标的多尺度特征,实现局部细节增强与全局语义校准的动态平衡,最终通过特征融合提升医学影像中微小病变的感知能力。该模块涉及 3 个并行分支:局部(local)、全局(global)和卷积(Conv)分支。

其一是局部与全局分支:给出输入特征张量 F , 通过逐点卷积(PWConv)进行调整,得到 F' ; 经过逐点卷积之后,分别采用不同的补丁大小参数 $P=2$ 和 $P=4$ 来捕捉局部和全局特征。在空间维度上对非重叠补丁进行聚合和位移,计算补丁之间的注意力矩阵,从而实现局部与全局特征的提取和交互。对特征 F' 进行展开(unfold)和重塑(reshape)操作,将输入特征拆分为多个不重叠的补丁;

然后在通道维度上进行求均值(mean)操作,得到补丁的平均值,如式(4)所示。

$$\bar{P} = \frac{1}{C} \sum_{c=1}^C \text{Reshape}(\text{Unfold}(F', P))^{(c)} \quad (4)$$

之后通过 FFN^[18]中的全连接层 MLP1 和 MLP2 对补丁进行特征变换,归一化得到更加稳定的特征 F'_c ;

使用 softmax 激活函数生成空间维度上的概率分布,生成相应权重,再将权重进行加权融合得到特征 A , 如式(5)、(6)所示。

$$F'_c = \text{MLP2}(\text{LN}(\text{MLP1}(\bar{P}))) \quad (5)$$

$$A = F'_c \cdot \zeta(F'_c, \text{dim} = -1) \quad (6)$$

式中: ζ 表示 softmax 函数。

在加权结果中,使用特征选择算法^[19],来选择与任务

相关的特征进一步增强。过程如下:

先计算特征 \mathbf{A} 与可学习参数 λ 之间的余弦相似度,将余弦相似度限制在 $[0,1]$ 之间,生成注意力掩码 \mathbf{M} ,然后应用掩码操作进行加权,这一过程可以表示为式(7)、(8),即:

$$\psi(\mathbf{F}', \lambda) = \frac{\mathbf{A} \cdot \lambda}{\|\mathbf{A}\| \cdot \|\lambda\|} \quad (7)$$

$$\mathbf{M} = \text{Clamp}(\psi, 0, 1) \quad (8)$$

式中: ψ 表示余弦相似度, clamp 为逐元素截断操作,表示把张量中的每个数值限制在指定的区间内, clamp 操作在数学上如式(9)所示。

$$\text{Clamp}(x, \min, \max) = \begin{cases} \min, & x < \min \\ x, & \min \leq x \leq \max \\ \max, & x > \max \end{cases} \quad (9)$$

其二是中间分支:包含了多个卷积,用于捕捉更细致的特征。将3个传统的 $7 \times 7, 5 \times 5$ 和 3×3 卷积替换为3个 3×3 卷积,生成了3个不同的卷积输出,相加得到 \mathbf{F}_c 。

最后,将局部,全局和卷积分支生成的特征 $\mathbf{F}_L, \mathbf{F}_G, \mathbf{F}_c$

通过元素相加 (addition) 的方式进行特征融合得到 $\hat{\mathbf{F}}$, 如式(10)所示。

$$\hat{\mathbf{F}} = \mathbf{F}_L + \mathbf{F}_G + \mathbf{F}_c \quad (10)$$

第2部分为 GCEM 模块,在局部感知提取模块的基础上,为进一步增强对重要特征的关注,采用图通道^[20]增强模块进行特征增强。将图卷积理论引入通道注意力中,每个通道作为一个特征顶点,它们之间的关系用邻接矩阵表示,不需要在局部和二维邻域结构中进行逐像素的运算,而是更加灵活地获得通道权值,并将提取的信息抽象为特征表示,将上一层的相关特征作为顶点,因此,顶点自然地形成了一个图结构。这个图能够强调信息特征,抑制无用特征,实现了通道的重新校准,避免了无关像素的干扰,该模块可以直接嵌入到任何骨干网络中,以更少的计算代价提高准确率。

首先将 PAEM 模块的输出 $\hat{\mathbf{F}}$ 输入 GCEM 中,经过最大池化 (avgpool) 和 1×1 卷积进行降维操作,再将特征图进行展平 (flatten) 得到 \mathbf{F}_{in} 。其次进入自适应学习模块 (ALM),在 ALM 中,将特征转换为一维卷积的形式,并利用 softmax 函数计算注意力权重 \mathbf{A}_1 ,通过 relu 激活函数对输出进行非线性变换,得到 \mathbf{F}_{out} 。最后,特征图会被重塑 (reshape),通过 1×1 卷积和 sigmoid 函数得到最终的输出特征 \mathbf{F} 。这一过程用公式表示为:

$$\mathbf{F}_{in} = \text{Flatten}(\text{Conv}(\text{Avg}(\hat{\mathbf{F}}))) \quad (11)$$

$$\mathbf{A}_1 = \text{T}(\zeta(\text{Conv1d}(\mathbf{F}_{in}))) \quad (12)$$

$$\mathbf{F}_{out} = \delta(\text{Conv1d}(\mathbf{A}_0 \odot \mathbf{A}_1) + \mathbf{A}_2) \quad (13)$$

$$\mathbf{F} = \hat{\mathbf{F}} \otimes \sigma(\text{Conv}(\text{Reshape}(\mathbf{F}_{out}))) \quad (14)$$

式中: \mathbf{A}_1 是 $N \times N$ 的对角矩阵, ζ 表示 softmax 函数, δ 表

示 relu 函数, σ 表示 sigmoid 激活函数, \mathbf{A}_0 是可学习的 $N \times N$ 单位矩阵, \mathbf{A}_2 是 $N \times N$ 的邻接矩阵。

1.3 多频率语义注意力模块

由于医学图像的某些特征表现出比尺度更高的频率方差,促进多频率信息对于提升医学图像分类模型至关重要^[21]。基于此提出了多频率语义注意力模块 MFSM,有效地集成多频率信息,在不同频段提取不同的特征,并抑制噪声的影响。

多频率语义注意力模块能够细化空间特征提取的过程,主要是离散余弦变换 (discrete cosine transform, DCT), DCT 表示为不同频率的余弦函数的加权和,能够将图像从空间域转化成频率域,将图像信息分解为不同频率的分量,这个分量可以对高频和低频成分进行分析。特别是在捕捉边界特征时,获得频域信息,同时生成通道注意力,有效地利用顺序注意力方法抑制噪声通道的影响,以捕捉复杂的细节和更广泛的结构信息,确保对图像内容的全面理解。结构如图4所示。

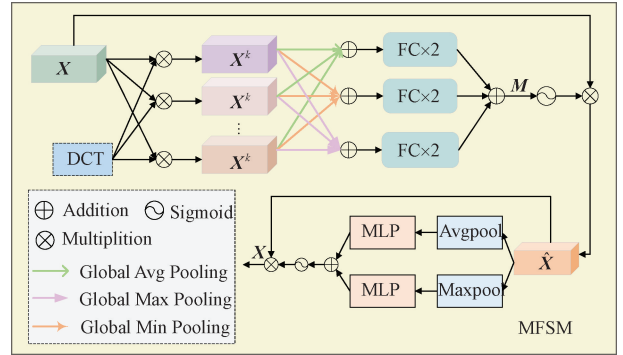


图4 多频率语义模块

Fig. 4 Multi-frequency semantic module

此模块每个频率分量 k 可以用二维 DCT 来表征,在数学上定义为式(15),即:

$$\mathbf{X}^k = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} (\mathbf{X})_{h,w} D_{h,k}^{u_k, v_k} \quad (15)$$

式中: k 是频率分量的索引,用于标识不同的频率分量,每个 k 对应一个特定的频率分量用 (u_k, v_k) 表示, u_k 和 v_k 是频率分量的横坐标和纵坐标。

2D DCT 定义为式(16),即:

$$D_{h,w}^{u_k, v_k} = \cos\left(\frac{\pi h}{H}\left(u_k + \frac{1}{2}\right)\right) \cos\left(\frac{\pi w}{W}\left(v_k + \frac{1}{2}\right)\right) \quad (16)$$

使用全局平均池化 (global average pooling)、全局最大池化 (global max pooling) 和全局最小池化 (global min pooling) 将每个 \mathbf{X}^k 压缩成 $Z_{avg}, Z_{max}, Z_{min}$; 经过全连接层对每个频率的统计信息进行聚合以生成通道注意力,重新校准得到 $\hat{\mathbf{X}}$ 。这一过程可以表示为:

$$Z_{avg} = \text{Avg}(\mathbf{X}^k) \quad (17)$$

$$Z_{max} = \text{Max}(\mathbf{X}^k) \quad (18)$$

$$Z_{\min} = \text{Min}(\mathbf{X}^k) \quad (19)$$

$$\hat{\mathbf{X}} = \mathbf{X} \cdot \sigma \left(\sum_{d \in \{\text{avg}, \text{max}, \text{min}\}} \tau_2(\delta(\tau_1 Z_d)) \right) \in R^C \quad (20)$$

式中: δ 表示 relu 激活函数, σ 表示 sigmoid 激活函数, τ 表示经过 FC。

最后将 $\hat{\mathbf{X}}$ 作为语义信息输入到平均池化和最大池化中,接着将特征分别馈送到同一个多层感知器(MLP)中,两者的输出相加后,通过 sigmoid 函数生成最终的特征图 \mathbf{X} 。计算语义特征的操作如式(21),即:

$$\mathbf{X}(\hat{\mathbf{X}}) = \hat{\mathbf{X}} \cdot \sigma(\text{MLP}(\text{Max}(\hat{\mathbf{X}})) + \text{MLP}(\text{Avg}(\hat{\mathbf{X}}))) \quad (21)$$

式中: $\sigma(\cdot)$ 是 sigmoid 函数。

2 实验和结果分析

2.1 实验环境与实验设置

实验在 Ubuntu18.04LTS 操作系统进行,采用 Pytorch1.10 深度学习框架,使用 Python 编程实现。硬件平台为 Intel (r) Xeon E5-2698 v4CPU 系列, Nvidia V100GPU,显存 32 GB。

实验中使用 AdamW 优化器优化网络参数,AdamW 优化器避免了 Adam 中权重衰减与梯度更新耦合的问题。训练 60 个 epoch。初始学习率设置为 0.000 1,使用余弦退火学习率调度器,此种方法能够在训练过程中自动调整学习率,避免学习率过高或过低,从而提高模型的收敛性和性能。权重衰减系数为 0.01,防止模型过拟合。训练时损失函数为二元交叉熵(BCELoss)。为提高实验效率将 batchsize 设置为 32。

2.2 实验数据

本实验采用 3 个公开的胸部 X 光片数据集,即 ChestX-ray14^[6]数据集、CheXpert^[22]数据集和 COVID-19 Radiography Database^[23]数据集。为了公平性评估,实验过程将 3 个数据集按照相同的划分比例:70%用于训练,10%用于验证,20%用于测试,其数据集具体设置如表 1 所示。在训练实验中,将输入图像大小调整到 256×256 ,之后随机裁剪并调整到 224×224 ,再以 50%的概率水平翻转,并基于 ImageNet 训练集中图像的均值和标准差进行归一化。在测试实验中,也进行同样的操作。下面是对各数据集的介绍。

表 1 数据集的构成

Table 1 Composition of the dataset

数据集	尺寸	训练	验证	测试
ChestX-ray14	224×224	78 484	11 212	22 424
CheXpert	224×224	156 553	22 364	44 731
COVID-19	224×224	14 818	2 116	4 231

ChestX-ray14 数据集: ChestX-ray14^[6]数据集是由美

国国立卫生研究院(NIH)在 2017 年发布的,该数据集收集了 30 805 名患者的 112 120 张正面胸部 X 射线图像。每张 X 光片标注了 14 种常见胸部疾病中的一种或多种类型。此外,还有 880 张 X 光片,984 个边界框评估疾病定位。

CheXpert 数据集: CheXpert^[22]数据集是由斯坦福大学发布的一个大规模公开胸部 X 射线数据集,包含来自 65 240 名不同患者的 224 316 张 CXR 图像。每张 CXR 图像在 CheXpert 数据集中都标注了 14 种观察结果。

COVID-19 Radiography Database 数据集: COVID-19 Radiography Database^[23]数据集由卡塔尔大学和达卡大学的研究人员合作建立。由多位研究员和医生一起制作的 CXR 图片,共有 4 个类别分别为: COVID-19 阳性(covid)、肺部混浊(lung opacity)、正常(normal)和病毒性肺炎(viral pneumonia)。

2.3 评价指标

针对 ChestX-ray14 和 CheXpert 两个数据集,采用受试者工作特征曲线下的面积(area under ROC curve, AUC)作为评价指标,值越大意味着分类性能越好;针对 COVID-19 Radiography Database 数据集,采用准确率(Accuracy),精确率(Precision),召回率(Recall),特异性(Specificity)和 F1-分数(F1-score)作为算法性能的评价指标。具体计算公式分别为:

$$TPR = \frac{TP}{TP + FN} \quad (22)$$

$$FPR = \frac{FP}{FP + TN} \quad (23)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (24)$$

$$Precision = \frac{TP}{TP + FP} \quad (25)$$

$$Recall = \frac{TP}{TP + FN} \quad (26)$$

$$Specificity = \frac{TN}{FP + TN} \quad (27)$$

$$F1\text{-Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (28)$$

2.4 对比实验分析

为了验证 FFA-Net 网络的有效性,在 ChestX-ray14 数据集中与其他现存的先进方法进行对比与消融实验。

表 2 为在 ChestX-ray14 数据集上测试的 AUC 值与其他算法的对比,其中,每种病理的最优值加粗显示,包括 PCSANet 网络^[24],可形变 Transformer 辅助诊断模型^[25], CheXNet 网络^[26], CheXNet-FPN 网络^[27], TransDD 网络^[28], DualAttentionNet 网络^[29]和 Hydravit 网络^[30]。其中, PCSANet 网络是以 ResNet50 作为特征提取,将标准卷积替换为金字塔卷积,能提取出更有鉴别力的病理异常特征,但仅仅替换聚卷缺乏信息之间的交互,导致分类效

果不佳;TransDD和CheXNet网络将局部特征与疾病标签进行关联,从而增强特征与标签间的交互能力,但易受到标签之间类别不平衡的影响,导致分类效果变差;Hydravit网络为混合深度学习网络,虽使用自注意机制自适应地聚焦于任务关键区域,但具有较高复杂度且对噪声较为敏感,在局部病灶的精细特征捕获上存在一定不足。可以发现,FFA-Net网络在14种疾病上的平均AUC值为0.8564,超过其7种算法且对14种疾病分类结果均有不同程度的提升,其中对Hernia的分类准确率最高,虽样本量最少,但网络的AUC值达到了0.9734。除Hernia外,AUC值最大的为Edema、Effusion、Mass等,分别为0.9006、0.8912、0.8810,表明网络较好的拟合了这几种疾病的特

征。此外,对于具有复杂病理的特征,如Infiltration,样本量虽多,但病理学表现为肺野内片状模糊阴影,边界不清,呈多发性或弥漫性分布,伴有网格状或结节状改变。但AUC值达到0.7251;如Pneumonia,其病理面积较大,且阴影密度不均匀,可能伴有支气管充气征或空洞识别难度大,但网络仍然达到0.7742。在部分疾病上略低的表现,可能是由于如Cardiomegaly、Nodule、Fibrosis等疾病在影像表现上与Effusion和Pneumonia等疾病存在一定病理相似性,且这些差距均处于较小范围内,在识别具有一定复杂性和相似性特征的疾病时,仍具备较好的判别能力。综上所述,FFA-Net网络对14种胸部疾病分类的平均AUC有进一步提升,总体分类能力有所增强。

表2 不同算法在ChestX-ray 14数据集上的对比实验

Table 2 Comparative experiments of different algorithms on the ChestX-ray 14 dataset

病症	AUC值							
	PCSA-Net ^[24]	文献[25]	CheXNet ^[26]	文献[27]	TransDD ^[28]	Dual ^[29]	Hydravit ^[30]	FFA-Net
Atelectasis	0.8018	0.8201	0.7424	0.803	0.791	0.794	0.810	0.8318
Cardiomegaly	0.9018	0.9115	0.9176	0.922	0.885	0.907	0.904	0.8981
Effusion	0.8815	0.8902	0.8196	0.857	0.842	0.853	0.878	0.8912
Infiltration	0.7084	0.7144	0.6841	0.718	0.715	0.712	0.712	0.7252
Mass	0.8378	0.8649	0.7629	0.875	0.837	0.824	0.874	0.8810
Nodular	0.7444	0.7725	0.6971	0.789	0.803	0.779	0.783	0.7998
Pneumonia	0.7495	0.7621	0.5961	0.752	0.745	0.716	0.759	0.7742
Pneumothrax	0.8740	0.9033	0.8196	0.914	0.885	0.847	0.894	0.9122
Consolidation	0.8047	0.8100	0.7169	0.788	0.753	0.727	0.822	0.8221
Edema	0.8903	0.8958	0.8376	0.877	0.859	0.877	0.882	0.9006
Emphysema	0.8997	0.9142	0.8109	0.935	0.944	0.925	0.908	0.9414
Fibrosis	0.7910	0.8082	0.7827	0.862	0.849	0.826	0.824	0.8265
Pleural Thickening	0.7829	0.8146	0.6739	0.814	0.803	0.757	0.797	0.8121
Hernia	0.8392	0.8757	0.8904	0.936	0.924	0.821	0.887	0.9734
平均AUC值	0.8219	0.8398	0.7680	0.846	0.831	0.819	0.838	0.8564

2.5 消融实验

为了评估FFA-Net网络的有效性以及网络中每个模块的贡献,在ChestX-ray14数据集上进行了4组消融实验,分别为:Baseline基准网络;移除TCAM模块;移除MBEM模块;移除MFSM模块。将消融实验测得每种疾病以及平均AUC值与FFA-Net进行比较,结果如表3所示。

由表3数据可知,通过在Baseline上增加模块,模型的AUC值均有一定程度的提升,从整体效果来看,当3个模块组合使用,平均AUC从0.8296提升至0.8564,提升约2.68%,提升最为显著,而且对每种疾病的识别率均有提升,说明FFA-Net网络在提升胸部疾病分类精度方面有显

著效果。针对各模块的消融实验而言,移除TCAM模块,模型的平均AUC值从0.8564下降至0.8522,下降约0.43%,9种病症的AUC值有所下降,表明TCAM模块通过构建全局上下文感知机制,在水平和垂直双维度上建立像素级依赖关系,促进跨通道信息的交互与融合,从而增强特征表达能力并提升分类精度;移除MBEM模块,平均AUC值从0.8564下降至0.8499,下降约0.66%,影响了11种病症的AUC值,表明MBEM模块采用多分支架构融合跨尺度特征,并结合图通道增强机制对关键特征进行重校准,有效增强网络对微小病理区域的聚焦能力,进而增强整体网络的特征提取能力;移除MFSM模块,平均AUC值从0.8564下降至0.8511,下降约0.53%,同样影响了

表 3 消融实验结果

Table 3 Results of ablation experiments

病症	Baseline	w/o TCAM	w/o MBEM	w/o MFSM	FFA-Net
Atelectasis	0.808 6	0.830 2	0.830 5	0.832 3	0.831 8
Cardiomegaly	0.892 2	0.902 1	0.902 0	0.898 1	0.898 1
Effusion	0.881 9	0.891 3	0.887 3	0.890 8	0.891 2
Infiltration	0.715 3	0.724 4	0.718 8	0.724 8	0.725 1
Mass	0.840 4	0.880 2	0.875 8	0.884 6	0.880 9
Nodual	0.756 1	0.795 8	0.783 9	0.792 4	0.799 8
Pneumonia	0.751 7	0.765 0	0.768 4	0.769 2	0.774 2
Pneumothorax	0.893 5	0.910 2	0.906 3	0.906 3	0.912 2
Consolidation	0.810 4	0.822 3	0.821 3	0.824 1	0.822 1
Edema	0.894 8	0.896 6	0.901 2	0.895 6	0.900 6
Emphysema	0.922 2	0.936 8	0.938 7	0.934 6	0.941 4
Fibrosis	0.798 0	0.836 5	0.833 1	0.831 2	0.826 5
Pleural Thickening	0.781 3	0.813 3	0.808 9	0.809 9	0.812 1
Hernia	0.867 5	0.925 7	0.921 7	0.922 1	0.973 4
AUC	0.829 6	0.852 2	0.849 9	0.851 1	0.856 4

9 种病症的 AUC 值,表明 MFSM 模块通过增强高频边缘特征并促进其与低频语义特征的跨频段交互,在有效抑制背景噪声的同时优化特征表达,从而提升对边界细节的解析能力。通过比较,表明 FFA-Net 网络采用的 3 个模块均能提升性能,而最终网络的优异表现正是源于各模块的协

同优化作用。

2.6 不同数据集下的泛化实验

为了评估 FFA-Net 网络的泛化能力,本实验采用 CheXpert 和 COVID-19 Radiography Database 数据集进行验证,实验结果如表 4、5 所示。

表 4 不同算法在 CheXpert 数据集上的泛化实验

Table 4 Experiments on generalization of different algorithms on CheXpert dataset

病症	AUC 值							
	0				1			
	GLMLL ^[31]	LCFL ^[32]	ClassNet ^[33]	本文	GLMLL ^[31]	LCFL ^[32]	ClassNet ^[33]	FFA-Net
No Finding	0.863	0.862	0.881	0.887	0.861	0.867	0.882	0.888
Cardiomegaly	0.835	0.850	0.846	0.862	0.838	0.839	0.838	0.850
Lung Opacity	0.712	0.716	0.732	0.744	0.713	0.715	0.728	0.740
Lung Lesion	0.731	0.718	0.770	0.796	0.733	0.735	0.777	0.799
Edema	0.834	0.829	0.849	0.856	0.837	0.835	0.857	0.864
Consolidation	0.680	0.713	0.730	0.743	0.678	0.680	0.695	0.701
Pneumonia	0.710	0.717	0.758	0.768	0.711	0.709	0.742	0.760
Atelectasis	0.698	0.667	0.701	0.711	0.695	0.700	0.717	0.729
Pneumothorax	0.794	0.810	0.870	0.895	0.793	0.793	0.840	0.864
Pleural Effusion	0.858	0.862	0.876	0.885	0.855	0.857	0.869	0.879
Pleural Other	0.772	0.769	0.802	0.821	0.775	0.785	0.808	0.820
Fracture	0.726	0.714	0.761	0.803	0.726	0.732	0.765	0.803
Support Devices	0.846	0.829	0.877	0.901	0.849	0.834	0.879	0.902
Enlarged Cardiomeastinum	0.629	0.644	0.656	0.677	0.628	0.627	0.626	0.654
AUC	0.763	0.764	0.793	0.811	0.764	0.765	0.787	0.804

表5 不同算法在 COVID-19 Radiography Database 数据集上的泛化实验

Table 5 Generalization experiments of different algorithms on the COVID-19 Radiography Database dataset

类型	方法	准确率	精确率	召回率	特异性	F1-分数
Classic Network	Transformer ^[18]	0.945 1	0.945 9	0.945 1	0.959 3	0.944 9
	MobileNetV2 ^[34]	0.946 5	0.947 0	0.946 5	0.963 4	0.946 2
	EfficientNet ^[13]	0.947 2	0.947 5	0.947 2	0.963 1	0.947 0
Popular Network	ChexNet ^[7]	0.952 0	0.952 3	0.952 0	0.966 0	0.951 8
	GLCM ^[35]	0.922 2	0.791 1	0.889 5	—	0.903 0
	PCSA Net ^[24]	0.946 5	0.946 7	0.946 5	0.964 2	0.946 3
	PPDL ^[36]	0.942 0	0.951 0	0.907 0	—	0.926 0
	ClassNet ^[33]	0.951 1	0.951 2	0.951 1	0.969 1	0.951 1
	FFA-Net	0.956 0	0.956 1	0.956 0	0.974 2	0.956 0

表4为在CheXpert数据集上的泛化实验结果与其他模型的比较,其中,每种指标的最高值以加粗显示。在此数据集中根据疾病存在情况标记为阴性(0)、阳性(1)或不确定(Uncertain)。本研究遵循GL-MLL^[31]的方法,采用两种不同的策略对不确定性标签进行实验:0策略将不确定性(Uncertain)标签视为负标签,记为0;1策略将不确定性(Uncertain)标签视为正标签,记为1。将FFA-Net网络分别与GL-MLL网络^[31]、LCFL网络^[32]和ClassificationNet网络^[33]进行对比。

从整体效果来看,对于GL-MLL、LCFL和ClassificationNet 3个网络,在两种策略下的AUC值均达到了最优,结果分别为0.811和0.804;再针对每种疾病来看:0策略下,在Cardiomegaly(0.846vs0.862)、Lung Lesion(0.770vs0.796)、Pneumothorax(0.870vs0.895)、Pleural Other(0.802vs0.821)、Support Devices(0.877vs0.901)和Enlarged Cardiomediastinum(0.656vs0.677)6种疾病检测方面,FFA-Net网络比其中最优化算法ClassificationNet的AUC值高约2%;而Fracture(0.761vs0.803)疾病,比其中最优化算法ClassificationNet的AUC值高约4%;在Lung Lesion、Pneumonia、Pleural Other、Fracture、Support Devices这5种疾病检测方面,FFA-Net网络比GL-MLL和LCFL网络的AUC值高约5%;1策略下,数据显示,在Lung Lesion(0.777vs0.799)、Pneumonia(0.742vs0.760)、Pneumothorax(0.840vs0.864)、Pleural Other(0.808vs0.820)和Support Devices(0.879vs0.902)5种疾病检测方面,FFA-Net网络比其中最优化算法ClassificationNet的AUC值高约2%;在Enlarged Cardiomediastinum(0.626vs0.654)疾病检测方面,比其中最优化算法ClassificationNet的AUC值高约3%,而Fracture(0.765vs0.803)疾病,比其中最优化算法

ClassificationNet的AUC值高约4%;在Lung Lesion、Pneumonia、Pleural Other、Fracture、Support Devices这5种疾病检测方面,所提网络比GL-MLL和LCFL网络的AUC值高约5%。综上所述,FFA-Net相比于GL-MLL、LCFL和ClassificationNet 3种网络均有不同程度的提升,表明该网络在CheXpert数据集上进行胸部疾病分类任务具有较好的泛化能力和鲁棒性。

表5为FFA-Net网络在COVID-19 Radiography Database数据集上泛化实验结果与其他模型的比较,其中,每种指标的最高值以加粗显示。首先从整体效果来看,FFA-Net网络在对比实验中的5项指标均为最优值,包括经典卷积模型Transformer^[18]、MobileNetV2^[34]、EfficientNet^[13]和GLCM网络^[35]、PCSA Net网络^[24]、PPDL网络^[36]、ChexNet网络^[7]和ClassificationNet^[33]改进网络。其次从对比网络来看,FFA-Net网络相较于其中最优化算法ChexNet在Accuracy、Precision、Recall、Specificity和F1-score上提升了0.4%、0.38%、0.4%、0.82%、0.42%,表明FFA-Net网络有更好的分类效果与泛化能力。

2.7 可视化分析

在实际临床诊断中,图像检测结果通常比数据更为直观。因此,将结果进行了可视化处理。通过采用加权梯度类激活映射(gradient-weighted class activation mapping, Grad-CAM)方法^[37],生成病灶定位热图;通过绘制边界框,直观评估了网络在定位病灶位置和学习疾病特征方面的能力。图5展示了8种胸部疾病的病灶区域可视化结果,将放射科医生对CXR图像的人工标注与各个模块作用后效果进行了对比。结果显示,所提网络的预测结果与医生的人工标注高度吻合,进一步证明了该网络的准确性和分类效果。这一成果对深度学习算法在辅助诊疗中的实际应用具有重要意义。

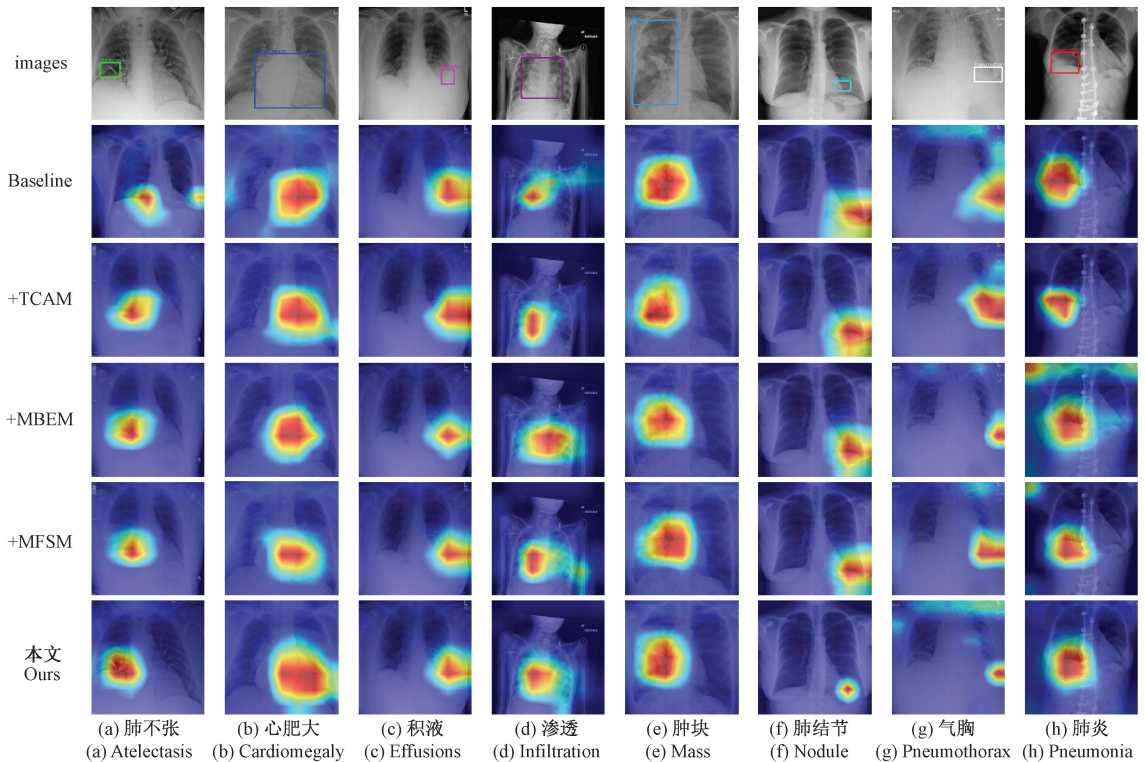


图 5 8 种胸部疾病的病灶区域可视化图

Fig. 5 Visualizations of lesion locations for 8 types of chest diseases

3 结 论

本研究提出了一种基于注意力与多尺度特征融合的 FFA-Net 网络架构,用于 CXR 中的胸部疾病分类。首先使用任务交叉注意力模块,在水平和垂直方向上聚合上下文信息,提高交互能力;其次,多分支提取模块,以获取不同尺度不同级别的特征,使其深层特征能够关注浅层特征中的细小病理区域;最后,利用多频率语义模块,有效提取不同频段的特征,并抑制噪声的影响。实验结果显示,FFA-Net 网络在 ChestX-ray14 数据集上 14 种病理的平均 AUC 值为 0.856 4,超过其他对比算法。并通过消融实验、在 CheXpert 和 COVID-19 Radiography Database 两个数据集上的泛化实验以及可视化分析等表明 FFA-Net 网络具有一定的优越性和较高的临床价值。本研究不仅提升了 CXR 影像的自动诊断能力,也为医学影像分析提供了一种高效的特征融合思路。然而,更复杂的技术仍有待探索,未来可进一步拓展其在不同医疗场景的适用性,推动其发展。

参考文献

[1] RAJPURKAR P, JOSHI A, PAREEK A, et al. CheXternal: Generalization of deep learning models for chest X-ray interpretation to photos of chest X-rays and external clinical settings [C]. Conference on Health, Inference, and Learning, 2021: 125-132.

[2] ESTEVA A, KURPREL B, NOVOA R A, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. Nature, 2017, 542(7639): 115-118.

[3] 司嘉龙,贾伟,赵雪芬,等.多放大倍率掩码自编码器的乳腺癌图像分类[J].电子测量技术,2025,48(10): 127-143.

SI J L, JIA W, ZHAO X F, et al. Breast cancer image classification based on multi-magnification mask autoencoders [J]. Electronic Measurement Technology, 2025, 48(10): 127-143.

[4] 梁国祥,韩亮.使用多特征融合的心律失常分类方法[J].电子测量与仪器学报,2024,38(7):109-115.

LIANG G X, HAN L. Method on arrhythmia classification utilizing multi-feature fusion [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(7): 109-115.

[5] 谢平,胡锦城,江国乾,等.基于多任务学习的快速序列视觉呈现脑电图分类[J].仪器仪表学报,2023, 44(11):215-223.

XIE P, HU J CH, JIANG G Q, et al. Classification of rapid serial visual presentation based EEG with multi-task learning [J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 215-223.

[6] WANG X S, PENG Y F, LU L, et al. ChestX-Ray8:

- Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE Press, 2017: 3462-3471.
- [7] RAJPURKAR P, IRVIN J, ZHU K L, et al. CheXNet: Radiologist-level pneumonia detection on chest x-rays with deep learning[J]. ArXiv preprint arXiv:1711.05225, 2017.
- [8] WANG H Y, WANG SH SH, QIN Z B, et al. Triple attention learning for classification of 14 thoracic diseases using chest radiography[J]. Medical Image Analysis, 2021, 67: 101846.
- [9] CHEN B ZH, LI J X, LU G M, et al. Lesion location attention guided network for multi-label thoracic disease classification in chest X-rays[J]. IEEE Journal of Biomedical and Health Informatics, 2019, 24(7): 2016-2027.
- [10] ZHU X F, PANG SH M, ZHANG X X, et al. PCAN: Pixel-wise classification and attention network for thoracic disease classification and weakly supervised localization [J]. Computerized Medical Imaging and Graphics, 2022, 102: 102137.
- [11] 邵凌云, 李镛, 关欣, 等. 基于高效通道注意力的胸部 X 光片疾病分类算法[J]. 激光与光电子学进展, 2023, 60(12): 1217001.
- SHAO L Y, LI Q, GUAN X, et al. Disease classification algorithm of Chest X-ray based on efficient channel attention[J]. Laser & Optoelectronics Progress, 2023, 60(12): 1217001.
- [12] TAN M X, LE Q V. EfficientNetV2: Smaller models and faster training [C]. International Conference on Machine Learning. PMLR, 2021: 10096-10106.
- [13] TAN M X, LE Q V. Efficientnet: Rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning. PMLR, 2019: 6105-6114.
- [14] PACAL I, CELIK O, BAYRAM B, et al. Enhancing EfficientNetV2 with global and efficient channel attention mechanisms for accurate MRI-based brain tumor classification [J]. Cluster Computing, 2024, 27(8): 11187-11212.
- [15] HUANG M L, LIAO Y C. A lightweight CNN-based network on COVID-19 detection using X-ray and CT images[J]. Computers in Biology and Medicine, 2022, 146: 105604.
- [16] ZENG ZH G, LIU J, ZHENG B, et al. A pneumonia recognition model based on multiscale attention improved EfficientNetV2 [J]. Computers, Materials & Continua, 2025, 84(1): 513-536.
- [17] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7794-7803.
- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. ArXiv preprint arXiv: 1706.03762, 2017.
- [19] SHI B F, GAI S, DARRELL T, et al. Toast: Transfer learning via attention steering [J]. ArXiv preprint arXiv:2305.15542, 2023.
- [20] XIANG X, WANG Z H, ZHANG J, et al. AGCA: An adaptive graph channel attention module for steel surface defect detection [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-12.
- [21] NAM J H, SYAZWANY N S, KIM S J, et al. Modality-agnostic domain generalizable medical image segmentation by multi-frequency in multi-scale attention [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 11480-11491.
- [22] IRVIN J, RAJPURKAR P, KO M, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison [C]. AAAI Conference on Artificial Intelligence, 2019, 33(1): 590-597.
- [23] CHOWDHUYY 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.
- [24] CHEN K, WANG X Q, ZHANG SH W. Thorax disease classification based on pyramidal convolution shuffle attention neural network [J]. IEEE Access, 2022, 10: 85571-85581.
- [25] 胡锦涛, 聂为之, 宋丹, 等. 可形变 Transformer 辅助的胸部 X 光影像疾病诊断模型 [J]. 浙江大学学报(工学版), 2023, 57(10): 1923-1932.
- HU J B, NIE W ZH, SONG D, et al. Chest X-ray imaging disease diagnosis model assisted by deformable Transformer [J]. Journal of Zhejiang University (Engineering Science), 2023, 57(10): 1923-1932.
- [26] WU X, FENG Y, XU H, et al. Chexnet: Combing transformer and CNN for thorax disease diagnosis from chest X-ray images [C]. Chinese Conference on Pattern Recognition and Computer Vision (PRCV), 2023: 73-84.
- [27] HASANAH U, AVIAN C, DARMAWAN J T, et al. CheXNet and feature pyramid network: A

- fusion deep learning architecture for multilabel chest X-ray clinical diagnoses classification [J]. The International Journal of Cardiovascular Imaging, 2024, 40(4): 709-722.
- [28] JIANG X B, ZHU Y, LIU Y T, et al. TransDD: A transformer-based dual-path decoder for improving the performance of thoracic diseases classification using chest X-ray [J]. Biomedical Signal Processing and Control, 2024, 91: 105937.
- [29] GONACLVES L A, JUNIOR G B, BESSA M L L, et al. DualAttentionNet: A convolutional neural network for thoracic disease classification in chest X-rays[J]. Procedia Computer Science, 2025, 256: 797-804.
- [30] ÖZTÜRK Ş, TURALI M Y, ÇUKUR T. Hydravit: Adaptive multi-branch transformer for multi-label disease classification from chest X-ray images [J]. Biomedical Signal Processing and Control, 2025, 100: 106959.
- [31] LI L T, CAO P, YANG J ZH, et al. Modeling global and local label correlation with graph convolutional networks for multi-label chest X-ray image classification[J]. Medical & Biological Engineering & Computing, 2022, 60(9): 2567-2588.
- [32] ZHANG K, LIANG W, CAO P, et al. Label correlation guided discriminative label feature learning for multi-label chest image classification [J]. Computer Methods and Programs in Biomedicine, 2024, 245: 108032.
- [33] TAHER M R H, HAGHIGHI F, GOTWAY M B, et al. Large-scale benchmarking and boosting transfer learning for medical image analysis[J]. Medical Image Analysis, 2025, 102: 103487.
- [34] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [35] GARG M, DHIMAN G. A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants [J]. Neural Computing and Applications, 2021, 33 (4): 1311-1328.
- [36] BOULILA W, AMMAR A, BENJDIRA B, et al. Securing the classification of covid-19 in chest X-ray images: A privacy-preserving deep learning approach[C]. 2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH). IEEE, 2022: 220-225.
- [37] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]. IEEE International Conference on Computer Vision, 2017: 618-626.

作者简介

胡欣茹, 硕士研究生, 主要研究方向为深度学习、智能图像处理、图像分类。

E-mail: 2987917956@qq.com

吕晓琪(通信作者), 博士, 教授, 主要研究方向为深度学习、智能信息处理、医学图像处理、数字化医疗图像处理。

E-mail: lxiaoqi@imut.edu.cn

谷宇, 博士, 教授, 主要研究方向为深度学习、智能图像处理、医学图像处理。

E-mail: guyu2010023@imust.edu.cn