

· 论 著 ·

基于全局分组坐标注意力的数字化放射 股骨头坏死分类模型

梁自晨^{1,2} 张子恒² 张竹岚^{2,3} 杨明翰² 杨家赵⁴

摘 要 目的 本研究旨在建立一种有效的术后股骨头坏死评估方案,特别关注股骨头数字化放射图像的复杂多样性。股骨头坏死是一种严重的骨科疾病,早期诊断对患者的预后至关重要。传统的影像学评估方法往往依赖于医生的专业判断,容易受到主观因素的影响,且对图像特征的提取和分析能力有限。因此,本文提出了一种基于深度学习的股骨头坏死分类模型,以提高模型对股骨头DR图像识别的准确度及计算效率。**方法** 本研究采用轻量级网络 MobileViT 作为主干网络,结合了卷积神经网络(convolutional neural network, CNN)和视觉变换器(Vision Transformer)的优点,以有效处理高维图像数据。为增强模型的特征提取能力,在MV2块前嵌入了 Squeeze-and-Excitation Networks(SeNet)注意力机制,通过自适应调整通道权重,帮助模型聚焦于重要特征。此外,课题组设计了全局分组坐标注意力(global grouped coordinate attention, GGCA)模块,使模型能够捕捉多维度的全局信息并进行特征增强。最后,采用基于 L1 非结构化剪枝策略来降低计算成本,通过去除冗余参数以提高模型的计算效率。**结果** 本文选取安徽省立医院总院及南区因股骨颈骨折接受闭合复位和空心螺钉复位的470例患者临床资料作为数据集,进行消融实验和对比实验,结果显示该模型的准确率达到94.57%,非零参数数量为 6.64×10^5 ,与改进前相比参数量降低30.14%,与常用CNN模型 AlexNet、MobileNet、ShuffleNet 相比,准确率分别提高10.24%,16.26%,21.98%。**结论** 本文提出的基于深度学习的股骨头坏死分类模型在术后股骨头坏死检测中具有较好的性能,是协助医生提高工作效率的一种有效尝试。通过结合轻量级网络架构、注意力机制和剪枝策略,该模型不仅提高了诊断准确性,还降低了计算成本。

关键词 深度学习;数字化放射图像;全局分组坐标注意力;模型压缩;特征提取

DOI: 10.3969/j.issn.1002-3208.2026.04.001.

中图分类号 R318 **文献标志码** A **文章编号** 1002-3208(2026)04-0331-09

本文著录格式 梁自晨,张子恒,张竹岚,等. 基于全局分组坐标注意力的数字化放射股骨头坏死分类模型[J]. 北京生物医学工程,2026,45(4):331-339. LIANG Zichen,ZHANG Ziheng,ZHANG Zhulan, et al. Digital radiography osteonecrosis classification model based on global grouped coordinate attention[J]. Beijing Biomedical Engineering,2026,45(4):331-339.

Digital radiography osteonecrosis classification model based on global grouped coordinate attention

LIANG Zichen^{1,2},ZHANG Ziheng²,ZHANG Zhulan^{2,3},YANG Minghan²,YANG Jiazhao⁴

1 School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei 230061;

2 Institute of Nuclear Energy Safety Technology, HFIPS, Chinese Academy of Sciences, Hefei 230031;

3 Huaibei Normal University, Huaibei, Anhui Province 235000;

4 Orthopedics Department of the First Affiliated Hospital of University of Science and Technology of China
(Anhui Provincial Hospital), Hefei 230001

Corresponding author:ZHANG Ziheng (E-mail: zhzhang@inest. cas. cn)

【Abstract】 Objective This study aims to establish an effective postoperative femoral head necrosis assessment protocol, with a particular focus on the complexity and diversity of digital radiography (DR) images of the femoral head. Femoral head necrosis is a severe orthopedic disease, and early diagnosis is crucial for patient prognosis. Traditional imaging assessment methods often rely on the professional judgment of doctors, are susceptible to subjective factors, and have limited ability to extract and analyze image features. Therefore, this paper proposes a deep learning-based classification model for femoral head necrosis, aiming to improve the accuracy and computational efficiency of the model in recognizing DR images of the femoral head. **Methods** This study uses the lightweight network MobileViT as the backbone network, combining the advantages of convolutional neural network (CNN) and vision transformers to effectively process high-dimensional image data. To enhance the feature extraction capability of the model, we embed squeeze and excitation networks (SeNet) attention mechanism before the MV2 block, which helps the model focus on important features by adaptively adjusting channel weights. In addition, we design a global grouped coordinate attention (GGCA) module to enable the model to capture multi-dimensional global information and perform feature enhancement. Finally, an unstructured pruning strategy based on L1 is adopted to reduce computational costs and improve the computational efficiency of the model by removing redundant parameters. **Results** This article selects clinical data of 470 patients who received closed reduction and hollow screw reduction for femoral neck fractures in the General Hospital and South District of Anhui Provincial Hospital as the dataset, and conducts ablation experiments and comparative experiments. The results show that the accuracy of the model reaches 94.57%, with a non-zero parameter quantity of 6.64×10^5 , which is 30.14% lower than that before the improvement. Compared with the commonly used CNN model AlexNet, MobileNet and ShuffleNet, this model improves accuracy by 10.24%, 16.26%, and 21.98% respectively. **Conclusions** The deep learning based classification model for femoral head necrosis proposed in this article has good performance in postoperative detection of femoral head necrosis and is an effective attempt to assist doctors in improving work efficiency. By combining lightweight network architecture, attention mechanism, and pruning strategy, this model not only improves diagnostic accuracy but also reduces computational costs.

【Keywords】 deep learning; digital radiography image; global grouped coordinate attention; model compression; feature extraction

0 引言

股骨头坏死 (avascular necrosis of the femoral head, AVN) 是一种严重的骨骼疾病, 早期诊断对改善预后至关重要。开发基于深度学习的分类模型可实现股骨头影像的自动化分析, 提高早期诊断的

准确性和效率。然而, 目前 DR 图像分类面临数据不平衡、标签不一致性、过拟合、特征提取复杂性等挑战。因此, 优化特征选择策略、平衡算法的泛化性与计算效率, 成为当前研究的核心问题。

近年来, 基于深度学习的医学图像分类/分级对于医学辅助诊断起到关键作用, 能够加快读片速度, 缩短患者的等待时间并减轻影像科医生的负担^[1]。其中 Transformer 模型凭借其出色的全局信息捕获能力和强大的表示能力, 在医学图像领域被广泛应用, 并取得了显著成效^[2]。宋宇等^[3]提出了一种 3DRes-ViT 多模态融合模型, 结合 ECA 模块增强通道相关性, 聚合局部与全局特征; 蒋舒颖等^[4]提出了一种融合注意力的主动迭代优化白细胞图像分类模型, 通过在 ResNet18 主干网络上附加用于主动学习的 LossNet 网络, 并加入 CBAM 注意力模块, 以

基金项目: 医学人工智能联合基金(MAI2023C009)、中科蚌埠技术转移专项项目(ZKBB202301)资助

作者单位: 1 安徽建筑大学机电学院(合肥 230061)

2 中国科学院合肥物质科学研究院核能安全技术研究所(合肥 230031)

3 淮北师范大学(安徽淮北 235000)

4 中国科学技术大学附属第一医院(安徽省立医院)骨科(合肥 230001)

通信作者: 张子恒。E-mail: zhzhang@inest. cas. cn

增强模型对白细胞特征区域的关注、提高模型的性能。平金如等^[5]提出一种特征融合的MV2-Transformer肺炎X线图像分类模型,设计跨层特征融合模块,进一步提高模型的语义特征提取能力。周杰^[6]提出了基于改进的MoBileNet V2的轻量级新冠CT图像分类模型,使得模块能够提取多尺度的特征,增强模型对全局特征的表达能力。Transformer的引入为医学图像领域注入了新的活力。

然而,在实际应用中,Transformer也面临一些挑战。Transformer模型的自注意力机制涉及大量的矩阵运算,这导致其在处理大规模图像数据时计算成本高昂和巨大的内存消耗,所以需要在保证模型性能的同时节约计算成本,魏榕剑等^[7]基于轻量级CNN模型MobileNetV2,提出一种基于MobileNetV2网络改进的MobileNetV2-SELN结构,使模型能更好地获取肺炎特征。赖裕等^[8]基于混合知识蒸馏(knowledge distillation, KD)的训练策略提出一种轻量级胸部疾病分类算法RMSNeT,在保持轻量化的同时分类精度更高,能有效降低算法运行时的硬件要求。郑洋等^[9]通过剪枝后蒸馏策略在Transformer模型宽度维度进行裁剪并恢复宽度子网络的精度,在保证模型性能的前提下缩减了模型的大小。以上研究都提供了解决模型计算资源复杂的问题,但是,DR图像形貌特征提取的研究将更加注重模型的自动化、智能化和高效性,所以需要研究更高效的深度学习模型结构和优化算法,以提高模型的性能和计算效率。

在医学影像分析中,尤其是移动端设备上,轻量化模型设计至关重要。本文对MobileViT进行改进,以满足DR图像形貌特征提取的自动化和高效性需求。引入SENet注意力机制自适应强调重要特征,提升模型对关键区域的关注。改进的全局分组坐标注意力(global grouped coordinate attention, GGCA)模块通过多尺度特征融合捕捉全局信息,增强特征提取的全面性与鲁棒性。这些改进确保模型在资源受限的设备上高效运行,并提高解析精度。

1 方法

1.1 总体流程

基于现有技术特点及DR图像的固有挑战,

MobileViT在DR图像分类中仍存在以下不足:全局注意力机制弱化了局部病灶特征提取,小样本下难以平衡多尺度信息;Transformer模块计算复杂度高,推理速度慢,移动端硬件难以满足实时性需求。针对以上问题,本文在MobileViT的基础上进行模型架构优化,构建了SG-MobileViT网络模型如图1所示。本研究拟进行模型架构优化和推理优化,具体步骤如下。

(1)在原模块局部特征表达中 3×3 标准卷积前引入SENet注意力机制,可以选择地强调输入特征图信息丰富的特征,同时抑制不太有用的特征,使模型能够专注于任务中最相关的特征。

(2)为进一步提升网络特征提取能力,在Layer5层后加入改进的全局分组坐标注意力模块,旨在利用特征图在空间维度(高度和宽度)上的全局信息来生成注意力图,并通过这些注意力图对输入特征图进行加权,增强特征表达能力。

(3)改进后的模型采用SiLU(sigmoid linear unit)激活函数,能够有效缓解神经元失活现象,从而提高模型的表达能力。

(4)通过剪枝策略对模型进行压缩,改进后的模型保持了轻量化特点且提高了复杂情况下对CT图像的识别能力,有助于实现在嵌入式端实现对复杂背景下的DR图像进行识别。

1.2 MobileViT网络

本文提出的轻量级的SG-MobileViT图像分类网络模型由MobileViT作为骨干网络结构。MobileViT结合了卷积神经网络(convolutional neural network, CNN)和视觉变换器(vision transformer, ViT)的优点^[10],采用轻量化结构,使其在计算资源有限的移动设备上仍能保持良好的性能。这种设计使得模型在处理高分辨率DR图像时,能够有效降低计算复杂度。由于其模块化设计,MobileViT可以根据具体应用需求进行调整和扩展,适应多种不同的任务和数据集。MobileViT网络结构如图2所示。

1.3 改进的全局分组坐标注意力模块

全局分组坐标注意力模块基于特征图在空间域(高度与宽度)的全局上下文信息构建注意力分布模型,通过该模型对原始特征实施动态权重调节,以提升特征表征的判别性能力。该方法将空间

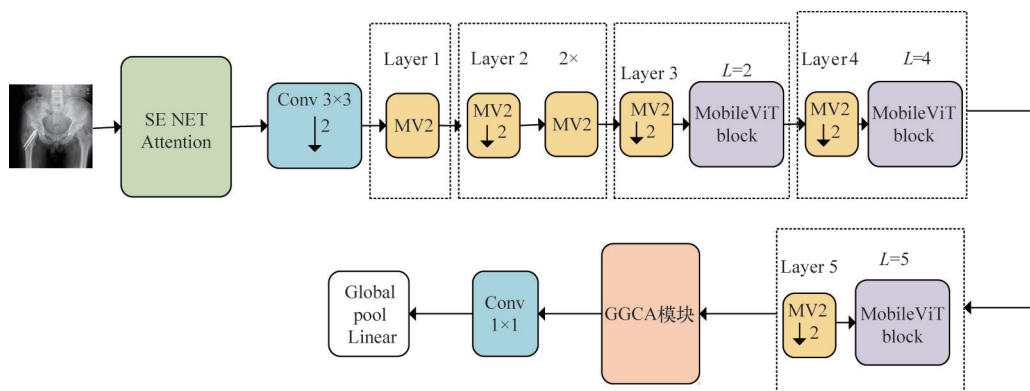


图 1 SG-MobileViT 模型结构

Figure 1 Structure of SG MobileViT model

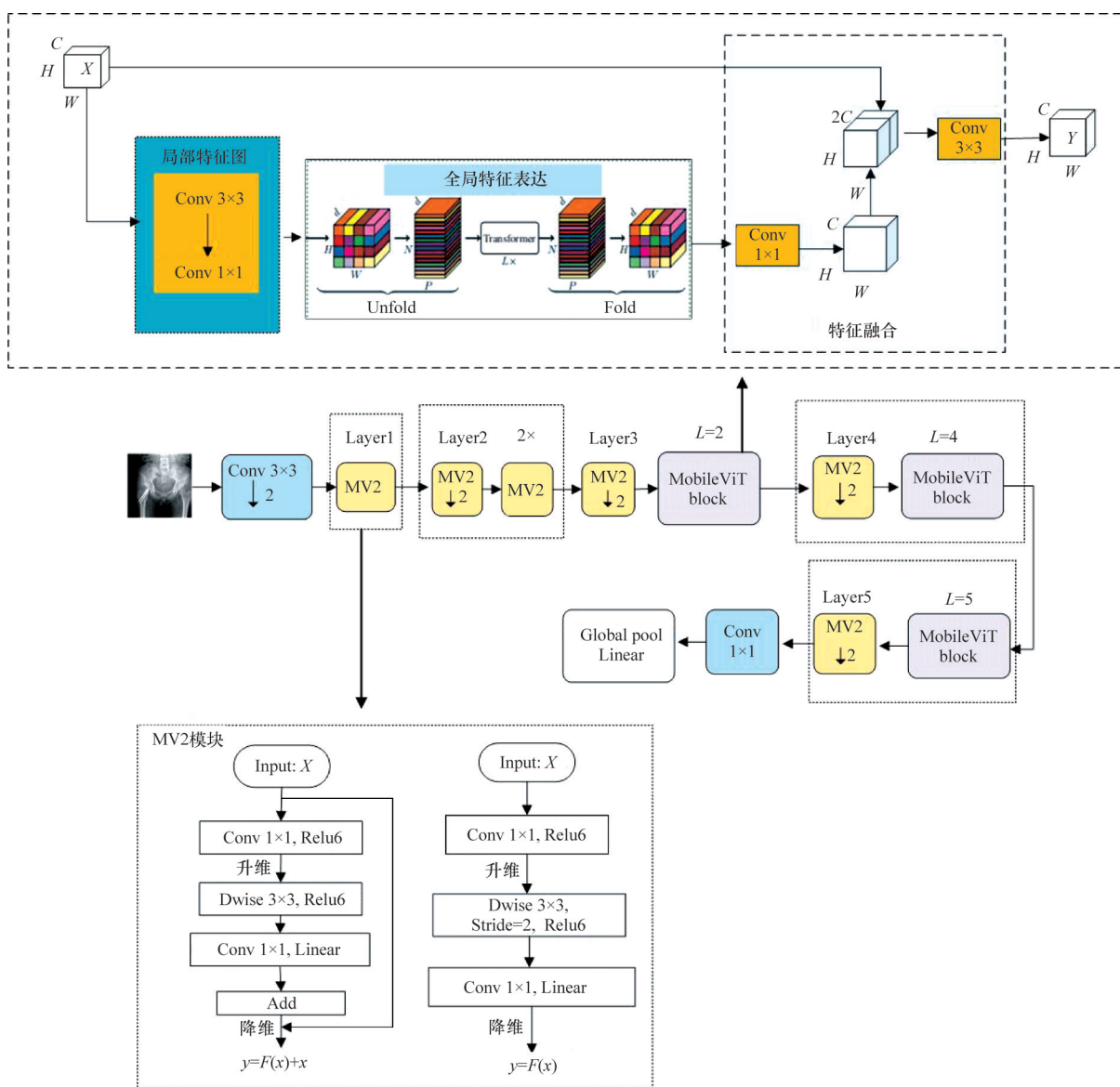


图 2 MobileViT 模型结构

Figure 2 Structure of MobileViT model

维度的整体关联信息转化为注意力权重系数,进而对输入特征进行自适应增强,优化了关键特征的区分度表达。该模块的具体结构如图3所示。

首先,对于输入特征图:

$$X \in R^{B \times C \times H \times W} \quad (1)$$

按照通道数将其分为 G 组,每组包含 C/G 个通道。这里, B 是批量大小, C 是通道数, H 和 W 分别是特征图的高度和宽度。分组后的特征图表示为:

$$X \in R^{C \times G \times \frac{C}{G} \times H \times W} \quad (2)$$

然后,对分组后的特征图在高度方向和宽度方向分别进行全局平均池化和全局最大池化操作:

$$\begin{aligned} X_{h, \text{avg}} &= \text{AvgPool}(X) \in R^{B \times G \times \frac{C}{G} \times H \times 1} \\ X_{h, \text{max}} &= \text{MaxPool}(X) \in R^{B \times G \times \frac{C}{G} \times H \times 1} \\ X_{w, \text{avg}} &= \text{AvgPool}(X) \in R^{B \times G \times \frac{C}{G} \times 1 \times W} \\ X_{w, \text{max}} &= \text{MaxPool}(X) \in R^{B \times G \times \frac{C}{G} \times 1 \times W} \end{aligned} \quad (3)$$

针对各特征图分组,采用参数共享的卷积结构实现特征变换。该结构由两次 1×1 卷积操作构成,其间嵌入批归一化处理与 ReLU 非线性激活,通过通道数压缩-恢复机制完成维度调整。具体表现为:首先缩减特征通道数降低计算复杂度,随后通过非线性映射恢复原始维度。

$$\begin{aligned} Y_{h, \text{avg}} &= \text{ConV}(X_{h, \text{avg}}), Y_{h, \text{max}} = \text{ConV}(X_{h, \text{max}}) \\ Y_{w, \text{avg}} &= \text{ConV}(X_{w, \text{avg}}), Y_{w, \text{max}} = \text{ConV}(X_{w, \text{max}}) \end{aligned} \quad (4)$$

通过将卷积层的输出相加,并应用 Sigmoid 激活函数,生成高度方向和宽度方向的注意力权重:

$$\begin{aligned} A_h &= \sigma(Y_{h, \text{avg}} + Y_{h, \text{max}}) \in R^{B \times G \times \frac{C}{G} \times H \times 1} \\ A_w &= \sigma(Y_{w, \text{avg}} + Y_{w, \text{max}}) \in R^{B \times G \times \frac{C}{G} \times 1 \times W} \end{aligned} \quad (5)$$

式中: σ 表示 Sigmoid 激活函数。

最后,将输入特征图按注意力权重进行加权,得到输出特征图:

$$O = X \times A_h \times A_w \in R^{B \times C \times H \times W} \quad (6)$$

注意力权重 A_h 和 A_w 会分别在高度和宽度方向上扩展以匹配输入特征图的尺寸。

GGCA 模块通过双重全局池化融合多维度全局特征,增强信息捕获能力。采用通道分组策略降低计算复杂度,同时保留特征多样性。该模块结合多维全局感知与注意力机制,利用特征强化优化模型性能,在复杂医学 DR 图像识别中表现出显著优势。

1.4 L1非结构化剪枝

L1非结构化剪枝是一种模型压缩技术,通过减少神经网络中不重要的权重来降低模型复杂度。具体来说,对于每个卷积层和全连接层,会计算各个权重的绝对值,并根据这些绝对值大小来判断其重要性。原理如图4所示。

该方法在批处理规范化(batch normalization, BN)层对缩放因子施加 L1 正则化,易于实现且不需更改现有 CNN 架构。通过 L1 正则化,BN 标度因子的值会被推向零,从而识别不重要的通道。在给定

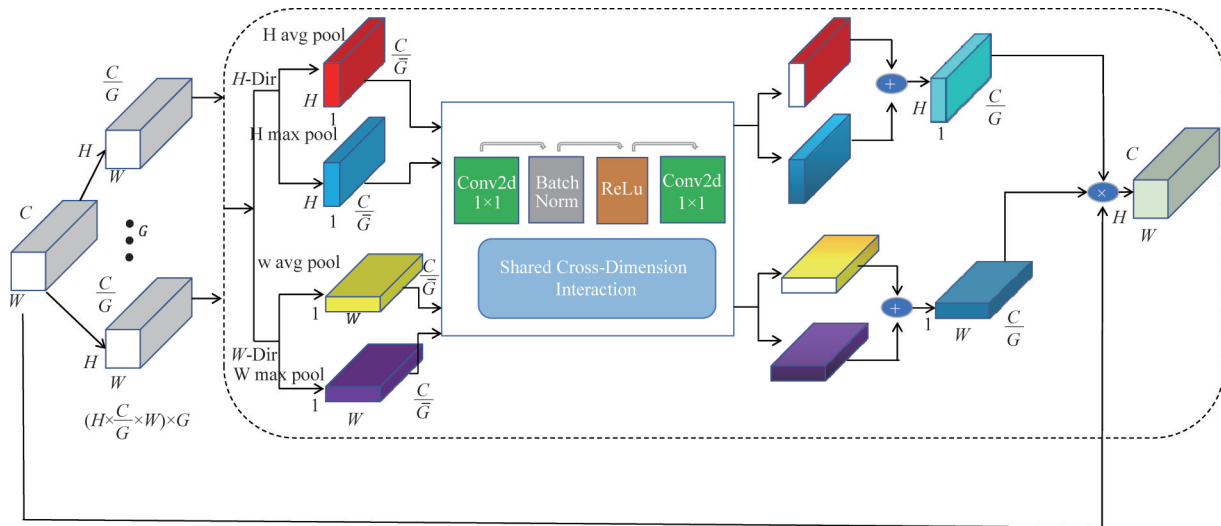


图3 GGCA模块图

Figure 3 GGCA module diagram

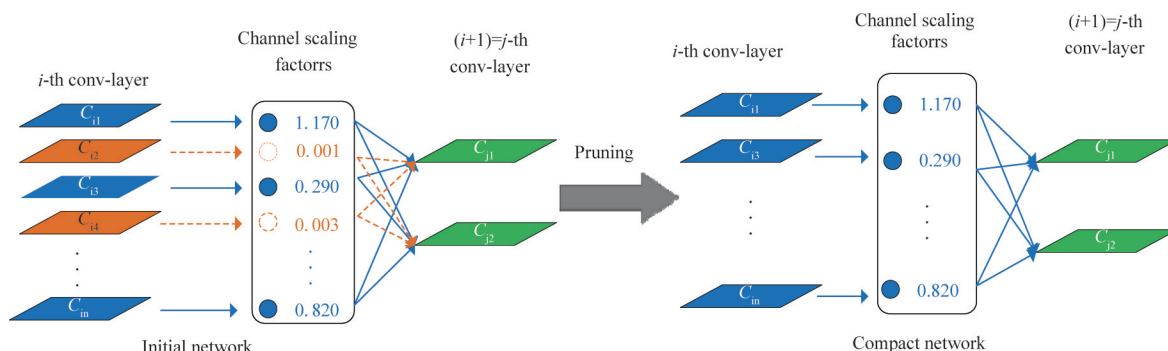


图 4 剪枝原理

Figure 4 Pruning principle

剪枝比例下,剪枝函数将权重张量中绝对值最小的部分置为零,认为这些权重对输出贡献较小,因此可以剪除而不显著影响模型性能。剪枝操作生成一个与权重张量同形状的掩码,掩码中为0或1表示权重是否保留,剪枝掉的权重在前向传播时不参与计算。这种方法优先去除影响小的权重,有效消除冗余连接,降低模型参数量,便于移动端部署。

2 结果与分析

2.1 实验数据及预处理

本研究通过回顾安徽省立医院总院及南区2009年1月—2018年10月接受因股骨颈骨折接受闭合复位和空心螺钉复位的470例患者临床资料为研究对象,该数据来自于骨折空心螺钉复位术后的DR图像,由两位专业医师通过DR图像评估股骨头的状态共同进行特异性标定,由185张训练集图像和64张测试图像组成,分为有股骨头骨坏死组139例和无股骨头骨坏死组64例。

由于DR图像较难获得且数据集样本量较少,为了减少小规模样本训练过程中的过拟合风险,引入水平/垂直翻转、平移变换等几何操作作为正则化手段。通过旋转和镜像操作,可以生成多种不同方向和视角的图像,这样模型在训练过程中能够接触到更多样化的数据,进而提高其对未见样本的泛化能力。扩充后的数据集共1 665幅,训练集为1 333幅,验证集为332幅。数据集信息如表1所示。

2.2 实现细节

首先将图像调整为256×256大小,然后使用SG-MobileViT模型将数据分成小批次,通过数据加载器加载图片和标签。特征图经过3×3卷积层提取局部

表 1 DR 图像数据

Table 1 DR image data

患者术后股骨头健康情况	原数据		扩充后数据	
	训练集	验证集	训练集	验证集
正常	104	35	749	187
术后股骨头坏死	81	29	584	145

特征,再用1×1卷积层调整通道数。接着,采用Unfold→Transformer→Fold结构进行全局特征建模,再通过1×1卷积层恢复通道数。最后,将原始输入特征图与处理后的特征图进行通道拼接,并通过3×3卷积层融合,得到最终输出。

本文所用的实验环境主要在GPU下进行,GPU型号为NVIDIA GeForce RTX3080Ti(12GB),内存为90 GB,使用Python环境3.8,深度学习框架PyTorch 1.10.0,Cuda版本11.3来实现网络搭建与模型训练。经多次反复实验,设置初始学习率(learningrate)0.000 1,采用Adamw优化器实现对模型参数的优化和正则化,其权重衰减设置为0.01。批处理大小(batch size)设置为32,训练轮次(epoch)设置为100。

2.3 实验结果与分析

本节将进行对比实验、消融实验、剪枝实验来评估模型的性能与有效性。

2.3.1 对比实验

为了全面评估和验证所提出的SG-MobileViT模型的性能,选取多个受到广泛认可的基础网络模型,在相同数据集上进行训练后对比分析它们与SG-MobileViT模型的性能。分别选取轻量级CNN模型EfficientNet^[11]、MobileNet^[12]、ShuffleNet^[13]和重量级模型AlexNet^[14]、ResNet^[15]、ConvNeXt^[16]对本文

DR图像数据集进行识别并与本研究模型进行对比实验。本实验采用加载预训练模型的方法,与其他经典模型进行对比,与轻量级、重量级CNN准确率对比如图5、图6所示,具体评估指标对比如表2、表3所示。

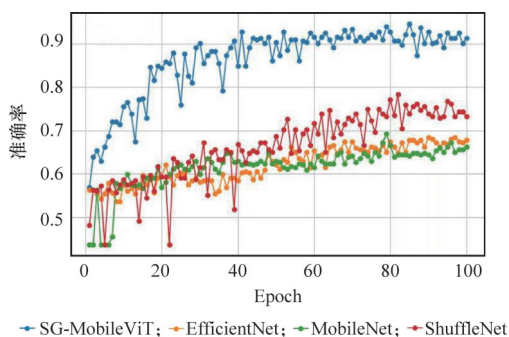


图5 与轻量级CNN模型准确率对比曲线

Figure 5 Comparison of accuracy curve with lightweight CNN model

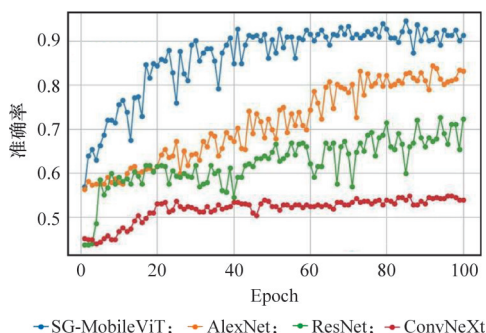


图6 与重量级CNN准确率对比曲线

Figure 6 Comparison of accuracy curve with heavyweight CNN

表2 与轻量级CNN对比

Table 2 Comparison with lightweight CNN

模型	准确率/%	精确率/%	召回率/%	F1分数/%	模型参数量/M
EfficientNet	68.37	71.91	68.37	65.21	15.30
MobileNet	78.31	78.35	78.31	78.09	8.69
ShuffleNet	69.27	69.27	69.27	68.71	4.80
SG-MobileViT	95.78	96.26	96.55	95.23	3.63

由图5和表2可以直观地看出与传统的轻量级CNN相比,在相似参数量的情况下,本文改进的SG-MobileViT网络模型准确率、精确率、召回率、F1分数分别是95.78%、96.26%、96.55%、95.23%,都高于

其他几个轻量级模型。与MoileNeT相比,SG-MobileViT模型参数量小5.06 M,准确率高了17个百分点;与ShuffleNet模型相比,SG-MobileViT模型参数量小约1.2 M,准确率高26个百分点。综上所述,与上述轻量级模型相比来说,SG-MobileViT精度高,模型尺寸小,可以更好地应用于DR医学图像的识别。

表3 与重量级CNN对比

Table 3 Comparison with heavyweight CNN

模型	准确率/%	精确率/%	召回率/%	F1分数/%	模型参数量/M
AlexNet	84.33	84.31	84.34	84.30	55.64
ResNet	72.59	72.54	72.59	78.09	78.30
ConvNext	69.27	69.27	69.27	72.20	106.11
SG-MobileViT	95.78	96.26	96.55	95.23	3.63

由图6和表3可以看出,本文研究的SG-MobileViT网络模型在模型参数量、准确率等评估指标都优于上述的重量级CNN模型,由此可见,本文对MobileViT进行改进提出的SG-MobileViT,实现了识别精度与识别速度的权衡,相比于其他卷积神经网络,模型参数量更少,具有更好的分类效果。

2.3.2 消融实验

本研究通过消融实验验证模型改进方案的有效性。在MobileViT基准模型基础上,分别引入SENet注意力模块和GGCA全局上下文模块进行独立测试,结果显示单一模块对准确率、精确率和召回率的提升有限(分别为+0.6%、+0.4%、+0.8%)。然而,采用双模块协同架构的SG-MobileViT模型时,这三项指标均显著提升,证实了SENet与GGCA模块在特征提取和上下文感知方面的互补效应。这一结果支持了复合模块设计的合理性,表明两者的协同作用对模型性能提升至关重要。消融实验结果如表4所示。

表4 消融实验结果

Table 4 Ablation experiment results

模型	准确率/%	精确率/%	召回率/%	F1分数/%
MobileViT	95.18	95.82	95.78	95.77
MobileViT+SENet	87.95	88.11	87.95	87.98
MobileViT+GGCA	94.88	94.94	94.87	94.88
SG-MobileViT	95.78	96.26	96.55	95.24

2.3.3 剪枝实验结果分析

本研究中使用的是 L1 非结构化剪枝,即基于权重的 L1 范数。具体来说,对于每个卷积层和全连接层,会计算各个权重的绝对值,并根据这些绝对值大小来判断其重要性。为了降低网络模型的参数量与计算量,通过剪枝减少网络层数,使模型更轻量化,从而减少内存占用并加速推理过程,具体实验结果数据如表 5 所示。

表 5 不同剪枝率下模型评估指标

Table 5 Model evaluation indicators under different pruning rates

模型	准确率/%	召回率/%	F1 分数/%	非零参数量/ 10^5
基础模型	95.78	96.55	95.23	9.52
剪枝 0.1	94.87	95.17	94.19	8.53
剪枝 0.2	93.67	93.79	92.83	7.59
剪枝 0.3	94.57	94.48	93.83	6.65
剪枝 0.4	90.96	84.82	89.13	5.71
剪枝 0.5	81.62	64.82	75.50	3.83
剪枝 0.6	65.06	44.13	52.45	3.83
剪枝 0.7	63.25	33.79	44.54	2.89
剪枝 0.8	60.54	42.75	48.62	1.95
剪枝 0.9	54.21	53.79	50.64	1.01

注:本表是模型在不同剪枝率参数下的性能。

由表 5 可以看出当剪枝率为 0.3 时,模型的准确率、召回率、F1 分数分别为:94.57%、94.48%、93.83%,相比于基础模型非零参数量下降了约 2.85×10^5 ,而且准确率、召回率、F1 分数仅下降 1.21、1.38、1.40 个百分点,在降低模型的参数量的同时,尽可能地保持了模型原有的精度。综上所述,课题组选择剪枝率为 0.3 的模型,有效地降低对存储带宽和计算资源的需求,同时减少能源消耗,在不显著损失性能的前提下提升模型效率,降低训练和推理成本。

3 讨论

本研究提出了一种融合轻量化与多尺度特征增强的 MobileViT 优化模型。该模型基于 MobileViT 的 CNN-Transformer 混合架构,引入 SENet 注意力机制进行特征通道重校准,并设计了改进的全局分组坐标注意力模块,通过并行全局池化和双路径特征融合,显著提升了长程依赖建模能力。两种注意力

机制的协同作用实现了局部细节与全局上下文的多尺度特征互补,增强了医学影像解析的特征表示。为满足移动端部署需求,模型将 ReLU 激活函数替换为 SiLU 函数以增强鲁棒性,并通过参数剪枝技术对 Transformer 层进行稀疏化,最终实现了 30% 的参数量降低,获得了紧凑型架构。

本研究的方法存在局限性。首先,DR 图像数据中不同病变类型的发病率差异导致临床数据高度不均衡,模型性能受原始数据分布限制。后续需收集多中心数据作为训练和测试样本,以全面评估方法。其次,所采用的剪枝策略可能降低了模型性能,未来应探索更优的优化策略,以降低计算成本并提升模型在自动诊断医疗影像方面的性能。

4 结论

针对 DR 医学图像识别任务,本文提出了一种基于改进 MobileViT 的模型,结合 SENet 模块和改进的全局分组坐标注意力模块。该模型通过加载预训练模型进行训练,并采用 L1 非结构化剪枝技术进行压缩,显著降低了参数量,便于在移动端部署。实验结果表明,该模型在保持轻量化特性的同时,识别准确率达到 94.57%,非零参数量为 6.65×10^5 ,参数量较改进前降低 30.14%。与常用 CNN 模型(如 AlexNet、ResNet、MobileNet、ShuffleNet)相比,识别准确率分别提高了 11%、23%、17% 和 26%。该模型的轻量化设计在确保运算效率的同时,实现了 DR 图像的高效解析,为疾病早期筛查提供了可靠的技术支持,有效拓展了医学影像在预防医学领域的应用。

参考文献

- [1] 石磊,籍庆余,陈清威,等.视觉 Transformer 在医学图像分析中的应用研究综述[J].计算机工程与应用,2023,59(8):41-55.
Shi L, Ji QY, Chen QW, et al. Review of research on application of vision transformer in medical image analysis [J]. Computer Engineering and Applications, 2023, 59(8): 41-55.
- [2] 徐锦凡.基于注意力机制与多尺度卷积的医学图像分割方法研究[D].桂林:桂林理工大学,2025.
Xu JF. Research on medical image segmentation methods based on attention mechanism and multi-scale convolution [D]. Guilin: Guilin University of Technology, 2025.
- [3] 宋宇,徐睿,才晓东,等.基于多模态融合的 3DRes-ViT 膝骨

- 关节炎分类模型[J]. 光学精密工程, 2024, 32(21): 3244-3255.
- Song Y, Xu R, Cai XD, et al. 3DRes-ViT knee osteoarthritis classification model based on multimodal fusion[J]. Optics and Precision Engineering, 2024, 32(21): 3244-3255.
- [4] 蒋舒颖, 李志明, 莫贤, 等. 融合注意力的主动迭代优化白细胞图像分类模型[J]. 中国生物医学工程学报, 2024, 43(4): 408-418.
- Jiang SY, Li ZM, Mo X, et al. Active iterative optimization of leukocyte image classification model with fused attention[J]. Chinese Journal of Biomedical Engineering, 2024, 43(4): 408-418.
- [5] 平金如, 孙子文. 特征融合的MV2-Transformer肺炎X光图像分类模型[J]. 计算机应用, 2025, 45(12): 4030-4036.
- Ping JR, Sun ZW. Pneumonia X-ray image classification model by MV2-Transformer with feature fusion[J]. Journal of Computer Applications, 2025, 45(12): 4030-4036.
- [6] 周杰. 基于深度学习的新冠肺炎CT图像分类[D]. 南京: 南京信息工程大学, 2023.
- [7] 魏榕剑, 邵剑飞, 温剑, 等. 融合SE和多尺度卷积的轻量级CT图像分类模型[J]. 云南大学学报(自然科学版), 2023, 45(3): 583-589.
- Wei RJ, Shao JF, Wen J, et al. A lightweight CT image classification model for fusion SE and multi-scale convolution[J]. Journal of Yunnan University: Natural Sciences Edition, 2023, 45(3): 583-589.
- [8] 赖裕, 李锵, 聂为之, 等. 基于混合知识蒸馏的轻量级胸部疾病分类算法[J]. 光电子·激光, 2024, 35(9): 993-1000.
- Lai Y, Li Q, Nie WZ. Lightweight thoracic disease classification algorithm based on mixed knowledge distillation[J]. Journal of Optoelectronics·Laser, 2024, 35(9): 993-1000.
- [9] 郑洋, 蒋晓天, 付东豪, 等. 基于剪枝-蒸馏的视觉Transformer模型压缩[J]. 西安电子科技大学学报, 2025, 52(4): 55-65.
- Zheng Y, Jiang XT, Fu DH, et al. Vision transformer model compression based on pruning-distillation[J]. 西安电子科技大学学报, 2025, 52(4): 55-65.
- [10] 苏申申, 周卫, 周淋芋, 等. 基于MobileViT轻量化网络的蘑菇图像分类算法改进[J]. 现代计算机, 2024, 30(21): 69-73.
- Su SS, Zhou W, Zhou LY, et al. Improvement of mushroom image classification algorithm based on lightweight MobileViT network[J]. Modern Computer, 2024, 30(21): 69-73.
- [11] Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks [C]// 36th International Conference on Machine Learning. Long Beach, California, USA: PMLR 97, 2019: 6105-6114.
- [12] Sinha D, El-Sharkawy M. Thin mobilenet: An enhanced mobilenet architecture[C]//2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). New York, NY, USA: IEEE Press, 2019: 280-285.
- [13] Zhang X, Zhou X, Lin M, et al. Shufflenet: an extremely efficient convolutional neural network for mobile devices [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE Press, 2018: 6848-6856.
- [14] Zhang X. The AlexNet, LeNet-5 and VGG NET applied to CIFAR-10[C]//2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE). Nanchang, China: IEEE Press, 2021: 414-419.
- [15] 李秋宇, 李其洋. 基于卷积神经网络的图像分类算法优化[J/OL]. 智能物联技术. [2026-04-14]. <https://cnki.net/urlid/33.1411.TP.20260413.1813.008>.
- Li QY, Li QY. Optimization of image classification algorithm based on convolutional neural network [J/OL]. Technology of IoT & AI. [2026-04-14]. <https://cnki.net/urlid/33.1411.TP.20260413.1813.008>.
- [16] Woo S, Debnath S, Hu R, et al. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE Press, 2023: 16133-16142.

(2025-06-20收稿, 2025-09-17修回)