

文章编号 1004-924X(2024)21-3244-12

基于多模态融合的 3DRes-ViT 膝骨关节炎 分类模型

宋宇¹, 徐睿¹, 才晓东², 王昕^{1*}

(1. 长春工业大学 计算机科学与工程学院, 吉林 长春 134000;

2. 吉林省前卫医院 信息科, 吉林 长春 130012)

摘要:针对膝骨关节炎的多分类准确率低,对膝关节图像特征提取不充分的问题,提出了基于多模态融合的 3DRes-ViT 网络模型。设计一种三维卷积神经网络(3D Convolutional Neural Networks, 3D CNN),分别提取双回声稳态(DESS)和快速自旋回波(TSE)磁共振成像(Magnetic Resonance Imaging, MRI)序列的三维浅层特征。研究发现这两种信息具有互补性,将这些特征进行融合。通过高效通道注意力(Efficient Channel Attention, ECA)模块捕捉融合后特征通道间的依赖关系,并输入到 Vision Transformer(ViT)编码器中,结合 3DCNN 和 ViT 的优势高效聚合两个模态的局部特征和全局特征。最后,ViT 的输出再与二维卷积神经网络(2D CNN)提取的 X 光图像特征进行融合,以进一步提升分类性能。实验结果表明,本文方法在 KOA 四分类任务中表现优异,平均分类准确率达到 91.2%、平均精度为 91.6%,F1 分数为 0.914,平均绝对误差降低至 8.8%,显著提高了膝骨关节炎多分类的准确率。

关键词:医学图像处理;深度学习;膝骨关节炎;多模态融合;X 光;磁共振成像;视觉转换

中图分类号:TP399 **文献标识码:**A **doi:**10.37188/OPE.20243221.3244

3DRes-ViT knee osteoarthritis classification model based on multimodal fusion

SONG Yu¹, XU Rui¹, CAI Xiaodong², WANG Xin^{1*}

(1. College of Computer Science and Engineering, Changchun University of Technology,

Changchun 134000, China;

2. Information Department, Jilin Qianwei Hospital, Changchun 130012, China)

* Corresponding author, E-mail: wangxin315@ccut.edu.cn

Abstract: Aiming at the problems of low accuracy of multiple classification in Knee osteoarthritis (KOA) and insufficient feature extraction of knee joint images, the 3DRes-ViT network model based on multimodal fusion was proposed in this paper. Firstly, the 3D Convolutional Neural Networks (3D CNN) is designed to extract the 3D shallow features of the two magnetic resonance imaging (MRI) sequences respectively, including dual echo steady state (DESS) and fast spin echo (TSE). The study found that the two kinds of information are complementary, and then these features are fused. Secondly, the dependencies among the fused feature channels are captured by the Efficient Channel Attention (ECA) module and

收稿日期:2024-07-12;修订日期:2024-08-12.

基金项目:吉林省科技厅自然科学基金面上项目(No. 20220101128JC)

fed into the Vision Transformer (ViT) encoders, which combines the advantages of 3DCNN and ViT to efficiently aggregate the local and global features of the two modalities. Finally, the output of ViT is then fused with the X-ray image features extracted by the 2D convolutional neural network (2D CNN) to further enhance the classification performance. Experimental results show that our method performs excellently in the KOA four-classification task, with an average classification accuracy of 91.2%, an average precision of 91.6%, an F1 score of 0.914, and a reduction of the average absolute error to 8.8%. The proposed model surpasses the mainstream methods in the current field and significantly improves the multiple classification accuracy of knee osteoarthritis.

Key words: medical image processing; deep learning; knee osteoarthritis; multi-modal fusion; X-ray; magnetic resonance imaging; vision transformer

1 引 言

膝关节骨性关节炎(Knee Osteoarthritis, KOA)是最常见的关节炎形式,是膝关节的一种退行性疾病,会引起疼痛和活动受限,影响患者的生活质量^[1]。在膝关节炎的晚期,影像学上所显示的异常通常是无法逆转的,晚期 KOA 只能采用全膝关节置换术进行治疗。然而,在 KOA 早期,膝关节的功能退化程度是可以通过适当的措施来预防的^[2]。因此,早期诊断 KOA 对于病理学、临床决策、预测疾病进展以及治疗至关重要。基于 Kellgren 和 Lawrence(KL)分级系统^[3],OA 的严重程度被分为 5 个等级:正常(0)、可疑(1)、轻微(2)、中度(3)和重度(4)。传统的 KOA 诊断需要训练有素或经验丰富的放射科医生来完成。计算机辅助诊断 KOA 可以帮助医生提高诊断效率。

多年来,X 射线一直广泛应用于诊断 OA。当关节软骨磨损时,患有 KOA 的关节 X 射线图像中膝关节骨之间的空间通常会变窄。然而,在标准 X 射线中能看到损伤之前,KOA 的症状可能已经出现。Roemer 等^[4]描述了 X 射线无法显示 KOA 的某些结构表现,并且无法检测到一些可能预示着疾病会迅速恶化的迹象。因此,X 射线具有局限性,无法诊断早期 KOA 进展,只对中重度 KOA 的诊断具有较高的敏感性。MRI 为膝关节提供了非侵入性和高分辨率的筛查手段。MRI 能够精确展示关节内各种组织的细微结构和形态,这对于早期识别关节软骨、滑膜以及韧带等组织的损伤和病变具有重要意义。医院的放射科医生经常使用 MRI 这种高分辨率成像技

术来诊断早期骨关节炎(OA)^[5]。这有助于及时发现病变,防止病情进一步恶化。

在 3D MRI 成像技术中,双回声稳态(Double-Echo Steady-State, DESS)和快速自旋回波(Turbo Spin Echo, TSE)在膝关节的软骨和其他软组织成像上能够提供非常有价值的信息。通过 DESS 技术,可以清晰显示关节软骨的层次结构,帮助医生评估软骨的健康状况和损伤程度。TSE 技术能够清晰显示膝关节的肌肉、肌腱、韧带等软组织的细节,帮助医生评估这些组织的健康状况^[6]。这两种不同的 MRI 序列对 KOA 的判断具有互补性,融合两者的特征能显著提高对膝骨关节炎的分类能力。

X 射线和 MRI 分别从骨骼及软组织对膝骨关节炎诊断提供了信息,本文通过结合使用这两种成像技术,实现两者的优势互补,使 KOA 诊断更具可靠性。

2 相关工作

在 KOA 诊断过程中,往往借助 X 光和核磁共振成像(Magnetic Resonance Imaging, MRI)等影像学检查,与膝关节关节镜的有创性相比,影像学方法评估 KOA 更安全、经济 and 高效。由于 X 光比 MRI 价格更便宜,它成为 OA 诊断中常用的一种检查方式。目前,KOA 的分类算法也多集中在用 2D CNN 对 X 光图像进行处理。Norman 等^[7]使用 Dense-Net CNN 来检测是否患有 KOA 并利用膝关节后前位的 X 射线图像分类 KOA 的严重程度,该方法产生的二次 Kappa 系数为 0.83,多类别分类的平均准确率为 66.71%。

Thomas 等^[8]使用 Dense-Net CNN 来检测和评估 KOA 并对其严重程度进行分类。该 CNN 在 36 190 个膝关节上进行训练,并在 OAI 数据库的 4 090 个膝关节的保留测试数据集上进行评估,深度学习方法的精确度、召回率、F1 分数和准确率分别为 68%, 64%, 64% 和 66%。Swiecicki 等^[9]设计了多输入 CNN 架构,同时分析膝关节前后位和侧位 X 射线图像,评估是否患有 KOA 或存在患病风险的人群中 KOA 的严重程度。该架构得到了 0.91 的加权 kappa 系数和 72% 的平均多类准确率。Tiulpin 等^[10]集成了两个独立的带有 SE 模块的 ResNet 50 CNNs 来检测和评估是否患有 KOA。该方法在检测是否患有 KOA 时, AUC(曲线下面积)达到了 0.98。Alshareef 等^[11]采用基于 Vision Transformer(ViT)的迁移学习方法进行膝关节炎的 KL 分级,在较少训练数据的情况下分类准确率达到 70%。

X 光图像对早期 OA 并不敏感, MRI 可以更好地呈现软骨、骨髓病变和积液等软组织^[12]。Guida 等^[12]在 2021 年首次提出了基于 MRI 的膝关节炎分类算法,并利用 3DCNN 对 MRI 序列进行 KOA 分类。在一个包含 1 100 个膝盖的数据集上,该算法在二分类(有 OA 和无 OA)上达到了 83.0% 的准确率,在五分类上实现了 54% 的准确率。3D CNN 虽然具有捕获序列切片间的空间关系的优势,但由于 MRI 数据集样本数目有限,模型出现了过拟合现象,导致分类准确率不理想。Yeoh^[13]等采用迁移学习的方法区分有、无骨关节炎。将 2D 预训练的权值转换为 3D 的初始权值,这项研究表明,迁移学习确实提高了模型的性能,其中 ResNet34 达到了 0.875 的最佳整体准确率和 0.871 的 F1 分数。由于单一模态图像不能准确地描述 OA,多模态医学图像可以提供丰富的互补信息。Guida 等^[14]利用 X 线、MRI 和患者临床信息的多模态中间融合改善 KOA 的分类,多分类准确率提升至 76%。Karim 等^[15]提出了一种基于 X 射线和 MRI 的可解释 KOA 诊断模型 Deep Knee Explainer,并在 MOST(Multicenter osteoarthritis study)数据集上利用 2D CNN 网络训练 DenseNet 和 VGG 架构,选择关节间隙变化的一种分级系统 JSN(Joint Space Narrowing)进行四分类,平均分类准

确率达到 91%。

目前,国内对膝关节炎的分类研究并不多。吴穗岚^[16]利用 CNN 针对膝关节的红外图像进行健康和有病的二分类,达到了 99% 的准确率。何方舟^[17]自建 X 光数据集并采用 DCNN 模型对有无膝关节骨肿瘤进行检测。曹传贵^[18]基于 VGG 网络模型对膝关节 SPECT(单光子发射计算机断层成像术)图像检测有无 OA,达到了 99% 的准确率。任会峰^[19]采用旋转不变非线性局部模糊编码方法提取 MRI T2 加权图像的纹理特征对 KOA 进行诊断,提高了诊断的准确率。

当前的研究集中在利用 CNN 模型实现 KOA 的分类,但有限的感受野使 CNN 难以直接捕捉全局特征,建立长距离依赖关系。而在医学图像分类任务中,全局特征和长距离依赖关系对于准确分类至关重要。在膝关节炎分类任务中, CNN 主要关注局部特征的提取和聚合,网络能够专注于学习输入数据中的边缘、纹理等局部特征,但它很难捕获 MRI 序列间的长期依赖关系。因此,虽然对 KOA 的分类已经取得比较多的研究成果,但是 KOA 的多分类准确率仍然不理想。Dosovitskiy 等^[20]提出了一种完全基于自注意力机制的 ViT 模型实现图像分类任务。Transformer 的自注意力机制可以捕捉长距离依赖关系,具备更强的全局建模能力^[20],但在图像块嵌入时直接将图像块进行序列化,局部特征的提取不足。

针对上述问题,本文提出一种用于膝关节炎分类的多模态融合 3DRes-ViT 网络。多模态融合分类模型通过融合两种不同的 MRI 序列及 X 光的特征,能够同时学习膝关节的骨骼和软组织特征,提高了分类性能。利用串行结构将 3D CNN 与基于自注意力机制的 ViT 模型结合,3D CNN 提取的 MRI 浅层特征作为 ViT 的输入,使 ViT 在捕获全局信息的同时保持局部特征提取能力,在一定程度上减少 ViT 对大量训练数据的需求,充分发挥了 3D CNN 和 ViT 各自的优势。通过引入 3D 高效通道注意力(Efficient Channel Attention, ECA)模块,对两种不同的 MRI 序列(SAG 3D-DESS 和 COR IW-TSE)的融合特征进行通道注意力加权,在不增加过多参数和计算成本的情况下,提升了网络模型的表征能力。

3 网络模型设计

3.1 模型架构

图 1 展示了 3DRes-ViT 多模态 KOA 分类网络模型的整体架构。该模型主要包括 MRI 特征提取网络和 X 光特征提取网络。

在 3D MRI 特征提取网络中,首先设计 3DCNN 分别提取 SAG 3D-DESS 和 COR IW-TSE 这两种不同 MRI 序列的浅层特征。3D 浅层特征提取网络主要由 3D Res Blocks 组成,通过 3D 卷积和残差连接进行特征提取。为了提取 SAG 3D-DESS 和 COR IW-TSE 的差异信息,分别为两者构建并训练 3DCNN 网络,它们具有相同的网络结构。将两个 3DCNN 网络提取的浅层特征进行拼接后进入通道注意力模块 3D ECA,以自适应地学习每个特征通道的权重。然后,3D ECA 模块输出的特征送入三层 Vision Transformer 的编码器中,来提取各种特征之间的交互信息。

X 光特征提取网络改进了 VGG-19 架构。文

献[21]使用来自骨关节炎倡议(OAI)数据集的基线 X 光图像数据库,比较了 ResNet, VGG, DenseNet 和 Inception 的变体,在所有比较的 CNN 架构中,VGG 分类器在膝关节 KL 分级任务上获得了比其他分类器更好的性能,其中 VGG-19 的准确率最高。因此,本文在 VGG-19 架构的基础上搭建了 X 光分类模型。该模型共包含 5 个卷积块,通过堆叠 3×3 步距为 1 填充为 1 的小尺度卷积层来代替大卷积层,在不降低感受野的基础上减少网络参数。与 VGG-19 网络不同的是,在每个模块的卷积层后均加入 BN 层增强泛化能力,并在 BN 层后使用 ReLU 激活函数。在前 4 个卷积块之后使用大小为 2×2 ,步距为 2 的最大池化层降维,在第 5 个卷积块后对 X 光特征进行全局平均池化。

最后,将 Vision Transformer 编码器得到的 MRI 模态全局交互特征与 X 光通过 2DCNN 提取的特征进行融合,并送入 MLP 层进行分类。在端到端的框架中,3DCNN 和 ViT 框架是同时被训练的。

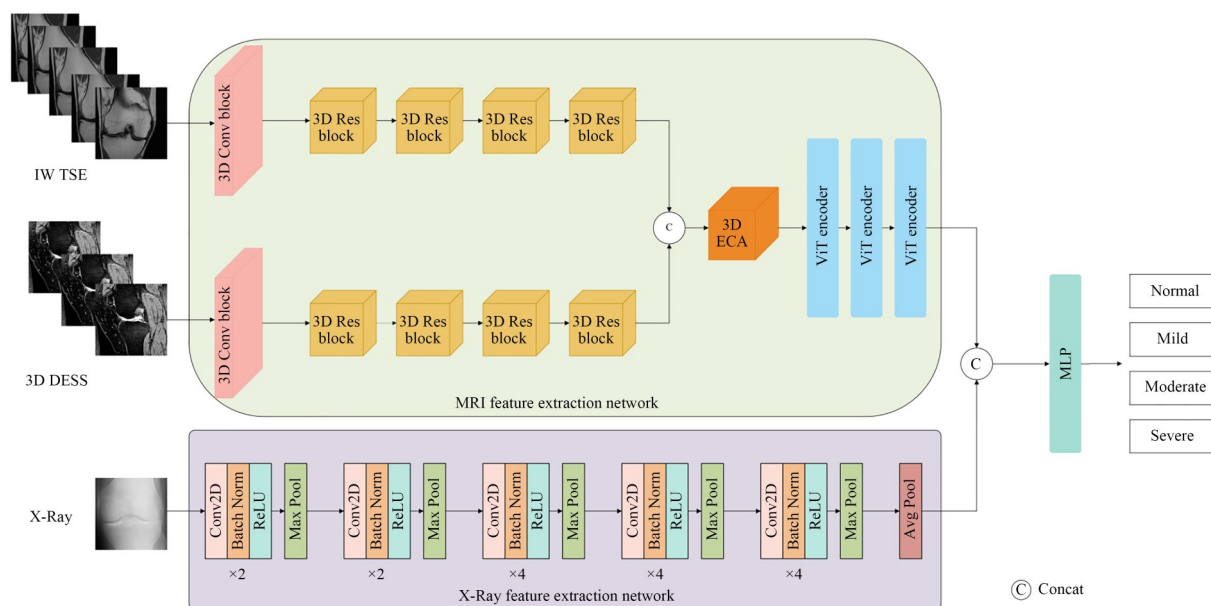


图 1 3DRes-ViT 多模态 KOA 分类网络模型总体架构

Fig. 1 Framework of 3DRes-ViT multi-mode KOA classification network model

3.2 特征提取网络

3.2.1 MRI 序列特征提取网络 3D Res CNN

在处理具有三维空间结构的 MRI 图像时,3DCNN 能够同时捕捉 MRI 图像的多个连续切

片之间的空间信息,而 2DCNN 只能使用 MR 图像的 2D 切片数据,丢失了连续图像切片之间的上下文信息。3D CNN 通过三维卷积核直接在三维空间上进行卷积运算,提高了特征提取的效

率。因此,采用3D CNN提取膝关节MRI的特征,能更准确地识别和分析病变区域。

如图2(a)所示,3D Conv block包括卷积核尺寸为 $5 \times 5 \times 3$,步距为 $(2, 2, 1)$ 的3D卷积层,BN层以及ReLU激活函数。受到残差网络ResNet的启发,本文设计了3D Res block,将3D卷积得到的特征进行跳跃连接。这个模块使用两个卷积块,每个block都有一个尺寸为 $3 \times 3 \times 3$,步距为 $(1, 1, 1)$ 的卷积层都进行了相同的填充操作。然后进行批量归一化和ReLU激活函数。通过交替堆叠3D Res block分别获取SAG 3D-DESS以及COR IW-TSE两个MRI序列的特征,然后进行拼接,实现这两种成像方式的优势互补,3D Res block如图2(b)所示。为了使两个模态的特征能够顺利拼接,通过控制每个block中第一个卷积层的步距来实现两个模态最后的特征维度相等,SAG 3D-DESS的深度步距为 $(1, 2, 2, 2)$,COR IW-TSE的深度步距为 $(1, 1, 1, 1)$ 。

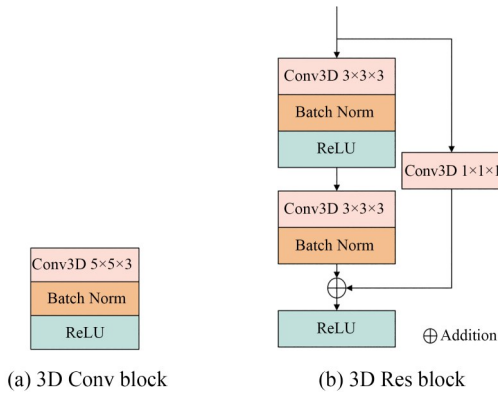


图2 3D Res CNN结构

Fig. 2 Framework of 3D Res CNN

第 i 层位于位置 (x, y, z) 的第 j 个特征立方体(特征图)的计算公式如下:

$$V_{ij}^{xyz} = \sigma \left(\sum_m \sum_{h=0}^{H_i-1} \sum_{w=0}^{W_i-1} \sum_{r=0}^{R_i-1} \omega_{ijk}^{hwr} V_{(i-1)k}^{(x+h)(y+w)(z+r)} + b_{ij} \right), \quad (1)$$

其中: m 表示第 $i-1$ 层中与当前第 j 个特征图相连接的特征图;空间卷积核的高度和宽度分别用 H_i 和 W_i 表示,深度用 R_i 表示; ω_{ijk}^{hwr} 是连接到第 m 个特征图的 (h, w, r) 位置上的权重, b_{ij} 是偏置, $\sigma(\cdot)$ 代表激活函数。

3D CNN模型捕捉的是三维数据的空间特征,因此通常复杂度较高。当MRI数据集有限时,模型可能无法充分学习到数据的内在规律和特征,很容易过拟合,降低3D CNN模型的分类性能。因此,只利用3D CNN提取两个MRI序列的低层(浅层)特征。

3.2.2 3D ECA

在两个MRI序列的3D特征融合后的网络中加入3D ECA块,以自适应地学习每个特征通道的权重。ECA是一种轻量级的通道注意力机制^[22],如图3所示, χ 代表输入特征图,高度、宽度、深度以及通道分别用 H, W, S, C 表示, k 代表自适应卷积核尺寸。ECA首先通过全局平均池化操作将输入特征图的通道信息压缩成一个一维向量,然后利用一个快速的一维卷积操作生成通道注意力权重。这个一维卷积的内核大小是根据通道尺寸的非线性映射自适应确定的,因此不需要进行降维操作。最后,将生成的通道注意力权重与原始输入特征图相乘,得到带有注意力加权的输出特征图 $\tilde{\chi}$ 。ECA利用一维卷积实现了局部跨通道交互,从而能够有效地提取通道间的依赖关系,提升特征表示的能力。

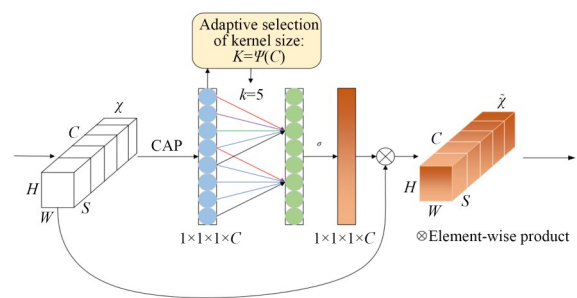


图3 3D 高效通道注意力(ECA)模块结构

Fig. 3 Framework of 3D efficient channel attention (ECA) module

3D ECA通道注意力权重计算公式如下:

$$\omega = \sigma \left(C1D_k(g(\chi)) \right), \quad (2)$$

其中: $g(\chi)$ 代表对输入特征图的全局平均池化(GAP), $C1D$ 代表一维卷积, k 是自适应卷积核大小, σ 是Sigmoid激活函数。

k 与通道 C 的映射关系如下:

$$k = \psi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}}, \quad (3)$$

其中: $\lfloor t \rfloor_{\text{odd}}$ 表示 t 的最近奇数, γ, b 通常设置为 2 和 1, 在本文模型中拼接后的 MRI 特征输入通道数 C 为 1 024, 故 k 为 5。

3D ECA 块是网络结构的重要组成部分, 它负责根据不同通道特征图之间的相互依赖性重新校准网络的特征响应, 从而在不增加过多参数和计算成本的情况下, 有效地增强网络的表征能力, 提高模型的判别能力。

3. 2. 3 Vision Transformer (ViT) 编码器

本文只使用 ViT 的编码器。经过 ECA 模块得到的特征图维度为 (1 024, 10, 10, 10), 直接把得到的每个 3D 特征图映射成向量并进行位置嵌入作为 ViT 编码器的输入。ViT 编码器用于学习输入之间的全局相关信息。

每个 ViT 编码器层通常包含两个归一化层 (LN)、一个多头自注意力 (MSA) 模块和一个多层感知器 (MLP) 层, 具体结构如图 4 所示。MSA 用于捕获特征序列之间的相关性。在 MSA 和 MLP 之前使用残差连接结构, 用于加速训练和提高模型的稳定性。预先定义 3 个可学习的权重矩阵 (W_Q, W_K, W_V), 将输入矩阵线性映射到查询 (Q)、键 (K) 和值 (V) 3 个 3D 矩阵。自注意力 (SA) 的计算公式如下:

$$SA = \text{Attention}(Q, K, V) = \text{soft max} \left(\frac{QK^T}{\sqrt{d_k}} \right) V, \quad (4)$$

其中 d_k 是 K 的维度。

使用 3D Res CNN 网络分别对两个模态的 MRI 进行特征提取并融合, 然后送入 3D ECA 模

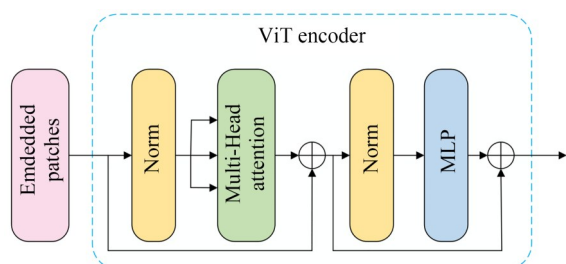


图 4 Vision Transformer 编码器的结构

Fig. 4 Structure of Vision Transformer encoder

块后经过不同层数的 ViT 编码器进行分类, 通过比较模型性能来确定编码器层数。

如图 5 所示, AA 代表平均准确率, AP 代表平均精确率, F1 代表平均 F1 分数。对于两个模态的 MRI 浅层特征, 使用三层编码器能够更好地提取全局特征, 达到最佳的分类性能。因此, 本文使用三层 ViT 编码器进行全局特征提取。

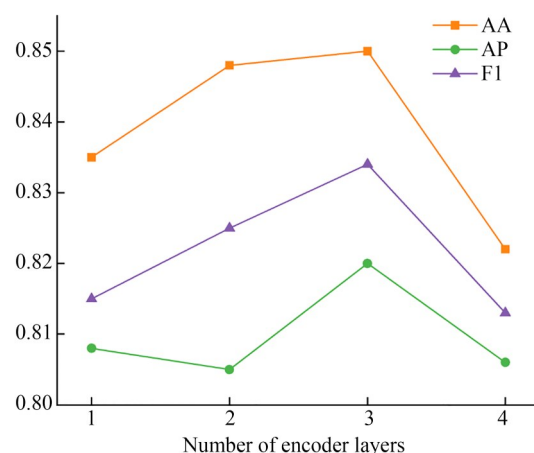


图 5 不同层数编码器性能对比

Fig. 5 Performance comparison of encoders with different layers

3. 2. 4 多模态特征融合

在 MRI 序列中, 本文采用膝关节的矢状面 (sagittal plane, SAG) 和冠状面 (coronal plane, COR) 两个视图进行融合。因为一个物体的多视图图像信息可以互补, 从而增强对该物体的分析和解释。骨关节炎在膝关节内的表现形式因位置和方向而异, 因此, 通过融合矢状面和冠状面这两个视图的信息, 能够获取膝关节更全面的状况, 从而提高网络的特征检测能力。如图 1 所示, 该框架采用两种不同的 MRI 膝关节扫描方式, 即矢状面扫描 (SAG 3D-DESS) 以及冠状面扫描 (COR IW-TSE)。对于每个视图的图像, 3D CNN 都会提取出独特的特征。这些特征随后通过拼接操作进行融合, 融合后的特征依次进入 3D ECA 和 ViT 编码器中进行进一步的全局特征提取。

通过 2D CNN 提取 X 光的特征, 与 ViT 编码器输出的 MRI 特征进行拼接, 送入 KOA 四分类器中进行分类。

4 实验和结果分析

4.1 数据集与预处理

4.1.1 数据集

本研究使用的数据来自骨关节炎倡议(OAI)数据集,源于一项多中心纵向骨关节炎研究(<https://nda.nih.gov/oai/>)。在 OAI 数据集中选择矢状三维双回声稳态回波(SAG 3D-DESS),冠状加权快速自旋回波(COR IW-TSE),以及双侧后前位固定屈曲膝关节 X 光 3 个模态的样本构建本文的数据集,总计包含了 4 157 名基线检查受试者,7 232 个膝盖样本的数据,每个样本都有三个模态的图像。在该数据集中,按照 7:1:2 划分训练集、验证集和测试集。为了保证数据集各类样本分布一致,采用分层抽样来划分数据集,具体分布情况如表 1 所示,正常、轻度、中度和重度的比例分别为 47%,32%,17% 及 4%。

表 1 数据集分布

Tab. 1 Dataset distribution

Class	Normal	Mild OA	Moderate OA	Severe OA	Total
Training	2 387	1 624	849	201	5 061
Validation	341	232	122	29	724
Testing	683	464	243	57	1 447
Total	3 411	2 320	1 214	287	7 232

4.1.2 预处理

膝关节 X 光片是尺寸为 $2\,000\times 2\,000$ 的 16 位 DICOM 文件。为了定位左右两个膝盖的感兴趣区域(Region of Interest, ROI),使用文献[23]中的定位模型标注 OAI 数据集中所有 X 光图像的边界框。根据标注框的中心点,执行中心裁剪提取出两幅尺寸为 560×560 的图像并转换为 PNG 格式,如图 6 所示。

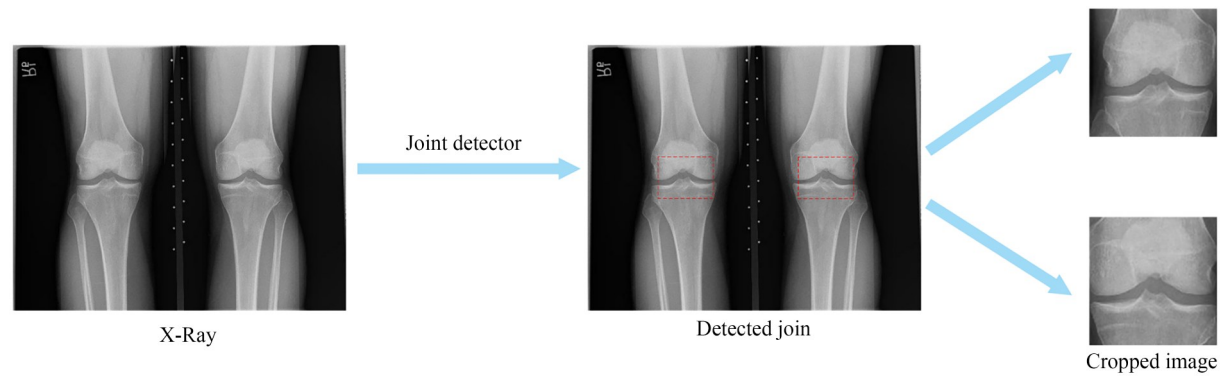


图 6 X 光图像的 ROI 区域提取
Fig. 6 Extraction of ROI from X-ray images

对于两种不同的 MRI 序列,首先将 DICOM 序列转换为神经成像信息学技术倡议(Nifti)的标准格式。如图 7 所示,通过裁剪图像的子区域来去除与分类膝关节 OA 不相关的区域,在这些区域中观察不到骨赘和软骨丢失,而且裁剪以后可以降低图像的输入维度^[12]。这种预处理方法使得分类模型能够更专注于膝关节的相关特征,从而提高分类的准确性。将 SAG 3D-DESS 切片尺寸从 384×384 裁剪到 160×160 ,将 COR IW-TSE 切片尺寸从 384×384 裁剪到 256×256 ,并缩放到 160×160 。

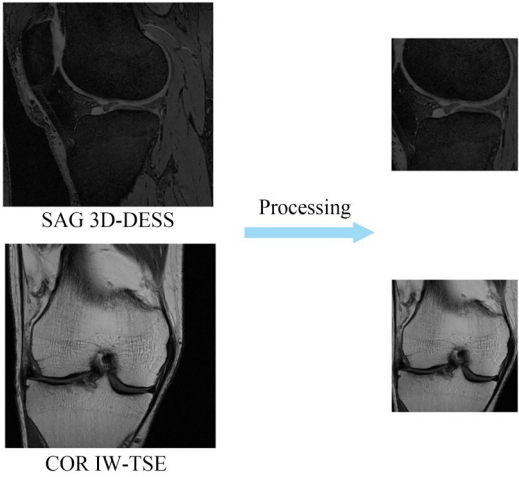


图 7 MRI 切片裁剪示例
Fig. 7 MRI slices cropping example

MRI 序列如图 8 所示,可以看出,开始和结束的切片不包含骨骼或软骨信息,而且这些切

片包含过多的噪声,会干扰特征提取。因此,在预处理期间移除这些切片十分重要。表 2 给

出了预处理后两种 MRI 序列切片对应的参数。

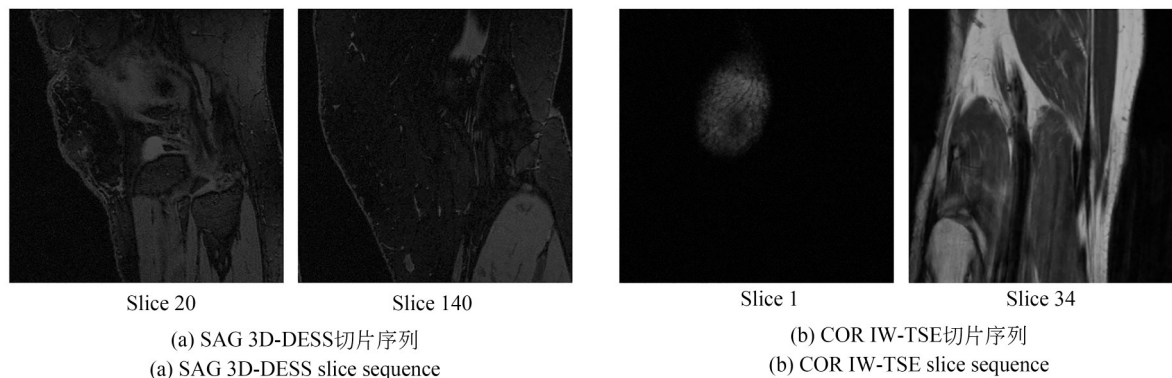


图 8 MRI切片序列

Fig. 8 MRI slice sequences

将所有右膝样本水平翻转后,对 3 种模态数据进行最小-最大归一化和 Z-score 标准化,

并通过随机旋转和伽马校正来增强训练集样本。

表 2 预处理后的切片参数

Tab. 2 Slice parameters after preprocessing

MRI acquisition	Slices size	Nb. of slices	Crop size	Scaling ratio	Removed slices
SAG 3D-DESS	352×352	160	160	[1,1]	[1:40];[120,160]
COR IW-TSE	352×352	35	256	[0.625,0.625]	[1:12];[23:35]

4.2 参数设置与评价指标

实验在一台高性能的 DELL 工作站上进行,该工作站配备 NVIDIA GeForce RTX 3090 显卡,搭载 Intel Core i9-13900K 处理器,主频为 5.80 GHz,操作系统为 Ubuntu 20.04.6 LTS。使用 Python 3.8.0 语言,基于 Pytorch1.10.0 搭建深度学习网络。为了解决不同类别样本之间的不平衡问题,采用 Focal loss 损失函数^[24]($\gamma=2$),并在训练过程中使用加权随机采样器进行采样,使用早停策略来避免过拟合现象。使用 Adam 优化器,权重衰减 1×10^{-4} ,批处理大小为 8,初始学习率为 1×10^{-4} ,学习率预热 5 个 epoch,训练周期为 100 epoch。

因为 OAI 数据集各类别样本分布不平衡,为客观地评价本文提出模型的分类性能,并与其他相关的研究成果进行对比,使用平均准确率(Average Accuracy, AA),平均精确率(Average Pre-

cision, AP), 平均 F1 分数(F1 Score, F1)和平均绝对误差(Mean Absolute Error, MAE)4 个指标进行评估。在临床诊断中,将膝关节的等级误判为邻近等级(如从 1 级误判为 2 级)的严重性,远低于误判为相差较远的等级(如从 1 级误判为 4 级)。因此,仅将分级准确性作为唯一的评估指标是不够的,需要采用 MAE 作为膝关节 KL 等级分类评估的另一个评价指标^[21]。上述评价指标对应的计算公式如下:

$$AA = \frac{1}{C} \sum_{i=1}^C \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i}, \quad (5)$$

$$AP = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FP_i}, \quad (6)$$

$$F1 = \frac{1}{C} \sum_{i=1}^C \frac{2 \times P_i \times R_i}{P_i + R_i}, \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y'_i - y_i|, \quad (8)$$

式中： C 为类别数， TP 为真阳例， TN 为真阴例， FP 为假阳例， FN 为假阴例， P 为单一类别精确率， R 为单一类别召回率， $R = \frac{TP}{TP + FN}$ ， n 为样本总数， y'_i 为预测值， y_i 为真实值。

4.3 消融实验

为了验证 3DRes-ViT 多模态融合分类网络各模块的有效性，进行渐进式消融实验，结果如表 3 所示。以 3DRes 网络为基础设计实验。Net-

work1: 3DRes 对 SAG 3DDESS 进行单模态分类；Network2: 3DRes 对 COR IW TSE 进行单模态分类；Network3: 3DRes 网络对两模态 MRI 浅层特征进行融合分类；Network4: 3DRes-ViT 网络对两模态特征进行融合分类，不加入 3D ECA 模块；Network5: 在 Network4 的基础上又融合了 X 光特征提取网络（不加 3D ECA 模块）；Network6: 在 Network5 的基础上加入 3D ECA 模块，即本文提出的模型。

表 3 消融实验结果

Tab. 3 Results of ablation experiments

Model	Data	ECA	AA	AP	F1	MAE
Network1	SAG 3D-DESS	×	0.677	0.580	0.604	0.367
Network2	COR IW-TSE	×	0.679	0.599	0.599	0.345
Network3	COR IW-TSE, SAG 3D-DESS	×	0.729	0.709	0.717	0.275
Network4	COR IW-TSE, SAG 3D-DESS	×	0.847	0.809	0.825	0.164
Network5	COR IW-TSE, SAG 3D-DESS, X-ray	×	0.889	0.892	0.890	0.112
Network6	COR IW-TSE, SAG 3D-DESS, X-ray	✓	0.912	0.916	0.914	0.088

Network3 将 SAG 3D DESS 与 COR IW TSE 两个模态的特征拼接进行分类，准确率、精确率和 F1 分数相比于 Network1 分别提升了 5.2%，12.9%，11.3%，MAE 降低 9.2%；相比于 Network2 分别提升 5.0%，11.0%，11.8%，MAE 降低 7%。可以看出，将两种不同的 MRI 序列得到的浅层特征进行融合可以提高模型的分类性能。Network4 使用 ViT 编码器对两个 MRI 序列融合后的特征进行全局交互，相比于 Network3 的准确率，精确率和 F1 分数分别提升了 11.8%，10.0%，10.8%，MAE 降低 11.1%。Network5 引入了 X 光特征提取网络，相比于 Network4 准确率提升了 4.2%，证明融合 X 光模态提取的特征可以进一步提升分类性能，减小 MAE。Network6 在两个 MRI 序列的融合特征送入 ViT 编码器之前加入了 ECA 注意力模块进行各通道特征交互，进一步提高了分类准确率。消融实验结果对比如图 9 所示。

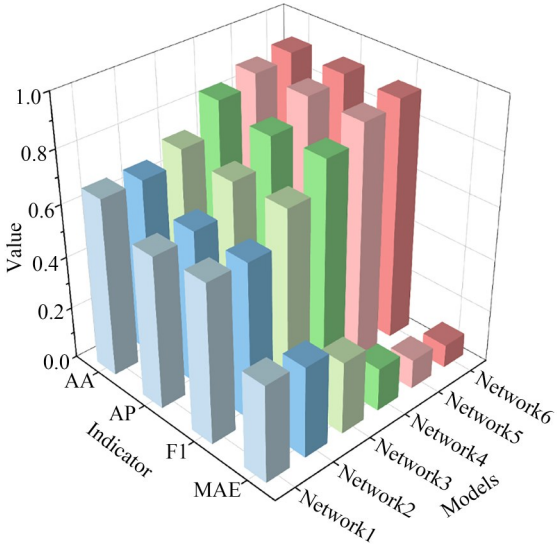


图 9 消融实验结果对比

Fig. 9 Comparison of ablation experiments

4.4 对比实验与分析

为了验证 3DRes-ViT 模型的有效性，与其他 KOA 多分类模型进行了实验对比，其中包括 3D CNN^[12]，Multi-Input 2D CNN^[9]，ViT^[11]，Fusion Model^[14]和 Deep Knee Explainer^[15]。不同模型的

分类结果如表 4 所示。由表 4 可以看出,与其他模型相比,3DRes-ViT 的客观评价指标最优。不同分类模型的评价指标对比如图 10 所示。由图可知,3DRes-ViT 多模态融合网络表现出更好的分类能力。

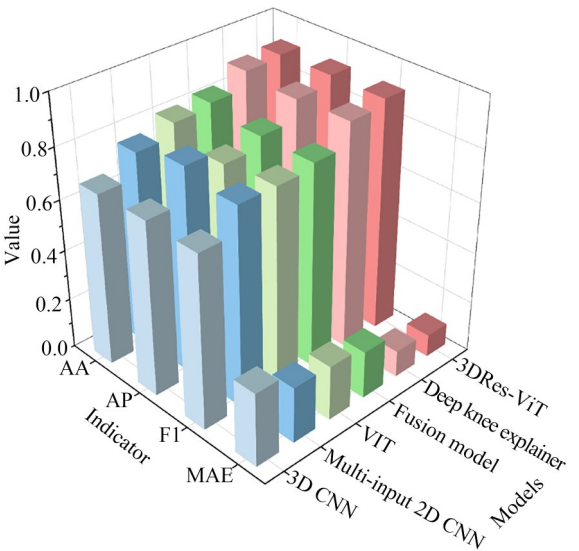


图 10 不同模型的分类评价指标对比
Fig. 10 Comparison of classification evaluation metrics for different models

此外,在表 4 中还对比了模型复杂度和运行效率。与其他 5 种模型相比,3DRes-ViT 的计算复杂度偏高,尽管在运行效率上不占优势,它的分类性能最佳。

图 11 展示了 3DRes-ViT 的分类混淆矩阵,正常患者的分类准确率为 94.9%,轻度患者 85.3%,中度患者 91.8%,重度患者 93.0%。

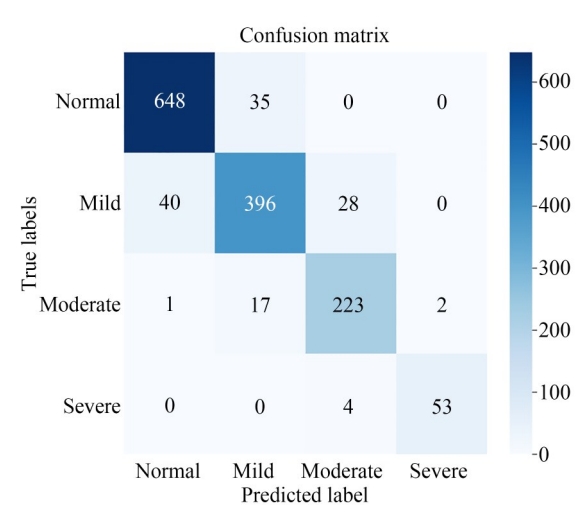


图 11 3DRes-ViT 混淆矩阵
Fig. 11 Confusion matrix of 3DRes-ViT

表 4 不同模型的分类对比实验结果

Tab. 4 Experimental results of classification with different models

Model	Data	AA	AP	F1	MAE	Params (M)	FLOPs (G)
3D CNN ^[12]	SAG 3D-DESS	0.680	0.689	0.683	0.291	33.16	23.63
Multi-Input 2D CNN ^[9]	X-Ray	0.771	0.819	0.783	0.221	4.45	6.23
ViT ^[11]	X-Ray	0.821	0.762	0.782	0.216	86.65	16.86
Fusion Model ^[14]	X-Ray, SAG 3D-DESS, Clinical data	0.838	0.804	0.802	0.187	46.44	25.46
Deep Knee Explainer ^[15]	X-Ray, SAG 3D-DESS	0.903	0.882	0.890	0.104	79.49	30.52
3DRes-ViT	COR IW-TSE, SAG 3D-DESS, X-Ray	0.912	0.916	0.914	0.088	69.09	127.28

5 结 论

本文提出了一种基于多模态融合的 3DRes-ViT 网络模型,用于 KOA 的四分类任务。该模型通过对 X 射线和 MRI 图像进行精确的预处理和特征提取,有效地结合 3D CNN 和 ViT 的优势,显著提高了 KOA 的分类准确率。实验结果显示,3DRes-ViT 模型在分类任务中表现优异,平均准确率达到了 91.2%,平均精确率为 91.6%,F1 分

数为 0.914,平均绝对误差降至 8.8%,证明了模型在处理复杂医学图像数据方面的强大能力。然而,本研究也存在一定的局限性。首先,尽管模型的分类准确性显著提高,但其计算复杂度偏高,这使得模型难以在资源受限的设备上部署。为了克服这一问题,未来的研究将探索轻量化模型架构,例如使用 ViT 的轻量化变种,考虑替换更高效的 3D CNN 架构,进行模型剪枝等以降低计算开销和内存需求。此外,本研究在应对数据集不平衡

问题上,在训练过程中采用加权随机采样器,这一方法并未彻底解决样本不平衡的问题。由于带标注的医学图像获取困难,未来计划利用生成对抗

网络^[25](Generative Adversarial Networks, GAN)来合成少数类别样本,以进一步平衡数据集,提高模型的泛化能力和实际应用效果。

参考文献:

- [1] LESPASIO M J, PIUZZI N S, HUSNI M E, *et al.* Knee Osteoarthritis: A Primer [J]. *Perm J*, 2017, 21: 16-183.
- [2] OEI E H G, RUNHAAR J. Imaging of early-stage osteoarthritis: the needs and challenges for diagnosis and classification[J]. *Skeletal Radiology*, 2023, 52(11): 2031-2036.
- [3] KELLGREN J H, LAWRENCE J S. Radiological assessment of osteo-arthritis [J]. *Annals of the Rheumatic Diseases*, 1957, 16(4): 494-502.
- [4] ROEMER F W, KWOH C K, HAYASHI D, *et al.* The role of radiography and MRI for eligibility assessment in DMOAD trials of knee OA [J]. *Nature Reviews Rheumatology*, 2018, 14(6): 372-380.
- [5] JURAS V, CHANG G, REGATTE R R. Current status of functional MRI of osteoarthritis for diagnosis and prognosis [J]. *Current Opinion in Rheumatology*, 2020, 32(1): 102-109.
- [6] HUNTER D J, GUERMAZI A, LO G H, *et al.* Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI osteoarthritis knee score) [J]. *Osteoarthritis and Cartilage*, 2011, 19(8): 990-1002.
- [7] NORMAN B, PEDIOIA V, NOWOROLSKI A, *et al.* Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs [J]. *Journal of Digital Imaging*, 2019, 32(3): 471-477.
- [8] THOMAS K A, KIDZIŃSKI Ł, HALILAJ E, *et al.* Automated classification of radiographic knee osteoarthritis severity using deep neural networks [J]. *Radiology Artificial Intelligence*, 2020, 2(2): e190065.
- [9] SWIECICKI A, LI N Y, O'DONNELL J, *et al.* Deep learning-based algorithm for assessment of knee osteoarthritis severity in radiographs matches performance of radiologists [J]. *Computers in Biology and Medicine*, 2021, 133: 104334.
- [10] TIJLPIN A, SAARAKKALA S. Automatic grading of individual knee osteoarthritis features in plain radiographs using deep convolutional neural networks [J]. *Diagnostics*, 2020, 10(11): 932.
- [11] ALSHAREEF E A, EBRAHIM F, LAMAMI Y, *et al.* Knee osteoarthritis severity grading using vision transformer [J]. *Journal of Intelligent & Fuzzy Systems*, 2022, 43(6): 8303-8313.
- [12] GUIDA C, ZHANG M, SHAN J. Knee osteoarthritis classification using 3D CNN and MRI [J]. *Applied Sciences*, 2021, 11(11): 5196.
- [13] YEOH P S Q, LAI K W, GOH S L, *et al.* Transfer learning-assisted 3D deep learning models for knee osteoarthritis detection: data from the osteoarthritis initiative [J]. *Frontiers in Bioengineering and Biotechnology*, 2023, 11: 1164655.
- [14] GUIDA C, ZHANG M, SHAN J. Improving knee osteoarthritis classification using multimodal intermediate fusion of X-ray, MRI, and clinical information [J]. *Neural Computing and Applications*, 2023, 35(13): 9763-9772.
- [15] KARIM M R, JIAO J, DOHMEN T, *et al.* DeepKneeExplainer: explainable knee osteoarthritis diagnosis from radiographs and magnetic resonance imaging [J]. *IEEE Access*, 2021, 9: 39757-39780.
- [16] 吴穗岚, 陈乐, 曾涛. 基于卷积神经网络的膝关节炎患者的膝盖红外图像分类 [J]. *中国计量大学学报*, 2019, 30(2): 185-190.
- WU S L, CHEN L, ZENG T. Knee infrared image classification of knee arthritis patients based on convolutional neural networks [J]. *Journal of China University of Metrology*, 2019, 30(2): 185-190. (in Chinese)
- [17] 何方舟, 牛凯, 唐顺, 等. 基于X线图像的膝关节周围原发性骨肿瘤辅助诊断的机器学习模型研究 [J]. *现代生物医学进展*, 2021, 21(15): 2842-2847.
- HE F ZH, NIU K, TANG SH, *et al.* Study on machine learning model of primary bone tumor around knee joint assisted diagnosis based on X-ray images [J]. *Progress in Modern Biomedicine*, 2021, 21(15): 2842-2847. (in Chinese)

- [18] 曹传贵,林强,满正行,等. 基于 VGG 的 SPECT 骨扫描图像关节炎分类[J]. 西北民族大学学报(自然科学版), 2021, 42(2): 36-45.
CAO CH G, LIN Q, MAN ZH X, *et al.* VGG-based classification of arthritis in bone SPECT images [J]. *Journal of Northwest Minzu University (Natural Science)*, 2021, 42(2): 36-45. (in Chinese)
- [19] 任会峰,伏建雄,鄢锋,等. 基于旋转不变非线性局部模糊编码的膝骨关节炎辅助诊断[J]. 软件导刊, 2022, 21(4): 25-31.
REN H F, FU J X, YAN F, *et al.* Knee osteoarthritis aided diagnoses based on rotation invariant nonlinear local fuzzy coding [J]. *Software Guide*, 2022, 21(4): 25-31. (in Chinese)
- [20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An image is worth 16x16 words: transformers for image recognition at scale [J]. *ArXiv e-Prints*, 2020; arXiv: 2010.11929.
- [21] CHEN P J, GAO L L, SHI X S, *et al.* Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss [J]. *Computerized Medical Imaging and Graphics*, 2019, 75: 84-92.
- [22] WANG Q L, WU B G, ZHU P F, *et al.* ECA-net: efficient channel attention for deep convolutional neural networks [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020. Seattle, WA, USA. IEEE, 2020.
- [23] ZHANG B F, TAN J M, CHO K, *et al.* Attention-based CNN for KL grade classification: data from the osteoarthritis initiative [C]. 2020 *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. April 3-7, 2020. Iowa City, IA, USA. IEEE, 2020.
- [24] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection [C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. October 22-29, 2017. Venice. IEEE, 2017.
- [25] BHAGWAN S, MANIVASAGAM, SONY A. Generative adversarial networks for synthesizing high-quality medical images [C]. 2024 *International Conference on Optimization Computing and Wireless Communication (ICOCWC)*. January 29-30, 2024. Debre Tabor, Ethiopia. IEEE, 2024.

作者简介:



宋 宇(1969—),男,吉林长春人,硕士,教授,硕士生导师,1996年于沈阳理工大学获得硕士学位,主要从事人工智能与机器学习方面的研究。E-mail: songyu@ccut.edu.cn

通讯作者:



王 昕(1972—),女,辽宁大连人,博士,副教授,硕士生导师,2003年于长春理工大学获得硕士学位,2007年于中国科学院长春光学精密机械与物理研究所获得博士学位,2015年于吉林大学博士后出站,主要从事图像处理与机器学习方面的研究。E-mail: wangxin315@ccut.edu.cn