DOI:10.19651/j.cnki.emt.2416040

基于改进 Swin Transformer 的膝骨关节炎 X 光影像自动诊断*

许 超 王云健 刘 洋 卢雪梅 丁 勇
(辽宁大学物理学院 沈阳 110036)

摘 要:膝骨关节炎是老年人群体的常见疾病,具有较高的致残性。依托深度学习算法开展膝骨关节炎的自动诊断,具有重要的应用价值。为此,提出了一种基于改进 Swin Transformer 模型的膝骨关节炎 X 光影像自动诊断算法。通过两层全连接层加 ReLU 激活函数的结构替换颈部网络的全局平均池化层,对迁移学习进行保护;在头部网络中添加全连接层与 Tanh 激活函数,组合出更多非线性特征;在数据预处理和模型训练过程中,分别依托 Albumentations 库和添加 Mixup 模块以此实现数据增强处理。实验结果表明,所提算法能够有效提升膝骨关节炎 X 光影像的分类精度,在 Kaggle 网站的公开数据集上诊断精度达到 76.0%;同时,经过在其他膝骨关节炎 X 光影像数据集与不同领域的医学影像数据集上进行泛化实验,结果表明其具有较好的泛化能力,进一步证明所提算法的有效性。

关键词:膝骨关节炎;Swin Transformer;全局平均池化;数据增强
中图分类号:TP391;TN98 **文献标识码**:A **国家标准学科分类代码**:520.6040

Knee osteoarthritis based on improved Swin Transformer X-ray image automatic diagnosis

Xu Chao Wang Yunjian Liu Yang Lu Xuemei Ding Yong
(College of Physics, Liaoning University, Shenyang 110036, China)

Abstract: Knee osteoarthritis is a common disease in the elderly population, which is highly disabling. Automatic diagnosis of knee osteoarthritis based on deep learning algorithm has important application value. Therefore, an automatic diagnosis algorithm of knee osteoarthritis based on improved Swin Transformer model is proposed. The transfer learning is protected by replacing the global average pooling layer of the neck network with a two-layer fully connected layer plus ReLU activation function. Adding full connection layer and Tanh activation function to the head network to combine more nonlinear features; in the process of data preprocessing and model training, data enhancement is realized by relying on Albumentations library and adding Mixup module respectively. The experimental results show that the proposed algorithm can effectively improve the classification accuracy of X-ray images of knee osteoarthritis, and the diagnostic accuracy reaches 76.0% on the public data set of Kaggle website. At the same time, the generalization experiments on other X-ray image data sets of knee osteoarthritis and medical image data sets in different fields show that it has good generalization ability, which further proves the effectiveness of the proposed algorithm.

Keywords: knee osteoarthritis; Swin Transformer; automatic diagnosis; data augmentation

0 引 言

膝骨关节炎(knee osteoarthritis,KOA)属于异质性疾病的一种,多发于老年人群体,尽早地发现和及时治疗对于 KOA 预后有着积极意义[1]。X 射线检查因普及性和经济性成为早期 KOA 检查的通用手段。凯尔格伦和劳伦斯分

收稿日期:2024-05-14
*基金项目:辽宁省研究生教育教学改革研究项目(LNYJG2022010)、辽宁省普通高等教育本科教学改革研究项目(2022-10140-11,2021-10140-11)、辽宁大学研究生优质在线课程建设与教学模式综合改革研究项目(YJG202202095,YJG202301021)、辽宁大学本科教学改革研究项目(JG2023CXCY04)资助

级系统统(kellgren and lawrence grading system,KL)是世界卫生组织认可的最常用 KOA 临床量表,根据是否存在关节间隙狭窄、有无骨赘产生、软骨下骨是否硬化与畸形等,按照轻重程度分成从 0～4 级共 5 个等级。随着深度学习技术的不断发展,运用深度学习算法进行 KOA 严重等级的自动诊断,有着较高的应用价值。Antony 等[2]率先使用深度学习算法做出多种尝试,将 KL 等级任务作为一个回归问题,并且使用均方根损失进行微调;随后 Antony 等[3]使用完全卷积模型并且优化交叉熵和均方损失得加权比率,最终获得 63.6%的分类准确率。Tiulpin 等[4]提出 Deep Siamese 卷积神经网络模型来测量 KOA 严重程度,其使用随机种子选择不同的膝关节连接子区域,通过融合所选区域的预测,获得 66.7%的分类准确率。Gorriz 等[5]使用注意力机制充当膝关节区域的无监督细粒度探测器,获得 64.3%的分类准确率;Chen 等[6]对 VGG 网络进行微调,改进模型损失函数,使用可调整序数损失进行分类,在 Github 网站上公开数据集,获得 69.7% 的分类精度。Cueva 等[7]提出基于 Deep Siamese 和 ResNet-34 的半自动计算机辅助诊断模型,多分类准确率平均值为 61%;Gu 等[8]提出一种基于 YOLO v3-Tiny 定位和使用 VGG-16 模型分类的算法,获得了 68.5%的分类准确率。

针对目前 KOA 的 X 光影像自动诊断识别准确率较低、所使用网络模型较为传统的问题,本文提出了一种基于改进 Swin Transformer 模型的 KOA 自动诊断算法。针对在模型训练过程迁移学习会导致源域与目标域跨度较大、影响模型精度的问题,通过两层全连接层加 ReLU 激活函数的结构替换原模型颈部网络中的全局平均池化层,对迁移学习进行保护。在头部网络中添加全连接层与 Tanh 激活函数,组合出更多非线性特征。在数据预处理和模型训练过程中,分别依托 Albumentations 库和添加 Mixup 模块以此实现数据增强处理,从而提升模型泛化性能。实验结果表明,所提的改进方法能够有效提升模型的分类精度;同时,本文先后在同领域的其他膝骨关节炎 X 光影像数据集与不同领域的结肠无线胶囊内窥镜图像数据集上进行泛化实验,结果表明所提算法具有较好的泛化能力,进一步证明所提算法的有效性。

1　Swin Transformer 网络结构

Swin Transformer 在 Vision Transformer(ViT)模型基础上提出类似卷积神经网络(convolutional neural networks, CNN)的位移窗口方法,通过分层设计使得多个等级窗口结合解决了图像大小的二次方复杂度问题[9],对比 CNN 模型拥有更好的全局建模的能力。

Swin Transformer 模型由 Swin Transformer 块堆叠而成,中间穿插 Patch Merging 下采样。块中独特采用两种多头自注意力机制,一种是基于窗口的多头自注意力机制(windows multi-head self-attention, W-MSA);另一种是基于滑动窗口的多头自注意力机制(shifted windows multi-head self-attention, SW-MSA),两种注意力机制交替使用并穿插残差连接,其结构如图 1 所示。

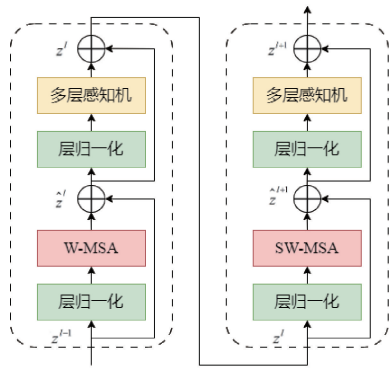


图 1　Swin Transformer 块结构示意图

Fig. 1　The architecture of Swin Transformer block

2　基于改进 Swin Transformer 的自动诊断算法

2.1　框架设计

原始 Swin Transformer 模型颈部网络采用全局平均池化操作,这虽符合如今采用全局平均池化替换全连接层的网络设计趋势,但会损失特征图的空间特征信息,并且全连接层(fully connected, FC)相较于全局平均池化层有着保护迁移学习的作用,对此本文将原始 Swin Transformer 模型颈部网络中的全局平均池化层替换为两层 FC 层加 ReLU 的结构;与此同时,在头部网络中添加全连接层与 Tanh 激活函数,进而在网络提取特征时获取更多的非线性特征组合;本文在头部网络中添加全连接层与 Tanh 激活函数;最后为提升模型的泛化性能,添加 Mixup 模块和基于 Albumentations 库的数据增强操作;改进模型的结构如图 2 所示。

2.2　改进颈部网络

Swin Transformer 模型的颈部网络采用一个全局平均池化层,通过其将主干网络中输出的二维特征图转换为一维特征图,以适应最后分类头部网络中的全连接层,起到良好的衔接作用。但模型中使用池化操作时应当考虑特定任务中出现在图像内的特征尺度,选取不同的池化方式[10],全局平均池化也有其局限性。由于其本质属于平均池化,若输入图像的所有像素值较低,那么经过池化操作会出现对比度方面的信息下降的问题,使特征图像会变得模糊,甚至当大多数区域的值为零时问题会进一步恶化,效果将显著降低[11]。并且全局池化操作实现对输入信息中局部特征的无序表示,无论要识别的特征信息处于什么位置都能用少量的信息来捕获区分特征,但采用此方式便完全忽略了局部特征位置,这种少量信息表示可能会面临一些失败的分类问题。对于医学影像分类问题,位置信息也是有用的,所以使用全局平均池化操作,虽然符合流行趋势但也存

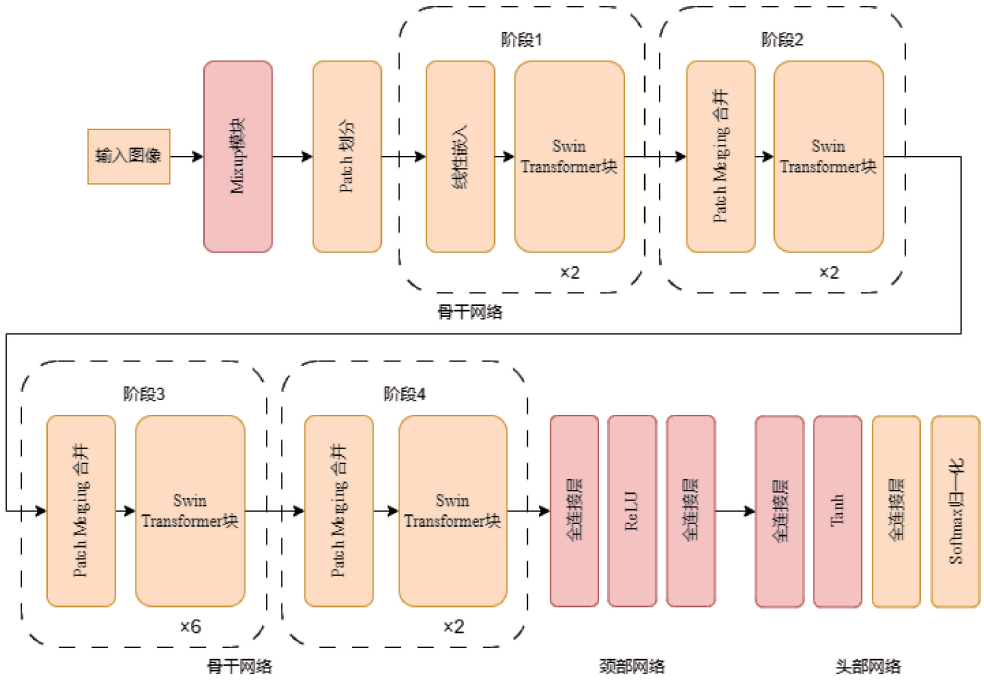


图 2　改进 Swin Transformer 的自动诊断算法整体网络结构

Fig. 2　Overall network structure diagram of automatic diagnosis algorithm for improved Swin Transformer

在一些局限性。

目前在网络训练过程中使用迁移学习方法来提升模型收敛速度及分类准确率是较为常见的,各类模型在 ImageNet 数据集上的训练权重非常容易获取。本文在实验过程中依靠基于 ImageNet-1K 数据集的迁移学习,该数据集中有 1 000 种类别,包含生活当中常见的物种,那么该数据集的源域为生活化场景,而将其应用于膝骨关节炎 KL 等级分类的场景下,目标域便为医学影像场景。如此源域与目标域可处于不同的向量空间或遵循着不同的数据分布,导致迁移学习效果变差。虽然 GoogLeNet 和 ResNet 模型的提出使得后续模型研究中都把 FC 层替换为全局平均池化层,但是 VGG 模型一直使用 FC 层,其实力依然不可小觑,作为传统卷积的代表至今仍有计算机视觉任务使用 VGG 模型作为骨干网络。GoogLeNet 和 ResNet 在具有大数据集的任务或迁移学习的源域与目标域相近的任务中表现良好,但在领域跨度较大的任务中效果并不理想,甚至出现失败,由此推断 VGG 模型中的全连接层保护视觉任务中迁移学习的高精度[12]。

针对上述问题本文对原模型颈部网络进行改进,将全局平均池化替换为两个 FC 层加 ReLU 激活函数的多层感知机。具体方式如下:首先设置两层 FC 层和 ReLU 激活函数的类别,定义该类别的初始化方式,设置输入层与输出层的通道数,由此衔接骨干网络与最后的分类头部网络。其次创建包含两个 FC 层与 ReLU 激活函数的多层感知机模型。最终定义网络的前向传播过程,调用函数对输入特征进行预处理,为将其送入最终的分类其中做准备。

2.3　改进头部网络

在原始模型中,头部网络是结构较为简单的线性分类头,在使用迁移学习的情况下,获取到的预训练权重便是在分类头上进行微调。线性分类头的结构仅为单层 FC 层,由于 FC 层的作用使得每个输入特征都根据权重进行加权组合,所以并未设置激活函数,使用 FC 层将特征映射至分类的标签空间。

改进头部网络的方式如下:首先,将头部网络中 FC 层输出特征的维度转换为嵌入维度,使得输入特征映射至更高的维度空间中,便于后续操作中更好地组合特征。其次,因为嵌入维度即是新增加一个维度为 768 的隐藏层,通过附加激活函数可以获得更多的非线性特征,所以增添与颈部网络不同的 Tanh 激活函数。再次,再经过一层 FC 层将隐藏层特征映射至输出类别上,与输出类别对应。可以说,通过增加嵌入维度的隐藏层使得改进后的头部网络更加适应前面骨干网络中的 Swin Transformer 模型,将骨干网络提取到的特征充分利用并映射至类别空间中进行最终分类。

2.4　数据增强

在进行膝骨关节炎 KL 等级分类任务时采用的数据集包含数据量较小,但少量数据使得模型难以学到较多的特征。原有模型中采用的数据增强方式较为简单,仅采用随机缩放裁剪,随机翻转及像素值归一化。对此本文采用一种 Mixup[13] 和基于 Albumentations 库结合的数据增强方法。

1) Mixup 数据增强模块

Mixup 为一种简单即用即插的与原始数据无关的数

据增强方法，其将两张图片通过随机的线性插值方法合成新的训练样本，属于混合图像数据增强范畴。在成对的数据集图像和与其对应标签的凸组合上训练神经网络，具体过程如下：

$$\tilde{x} = \lambda x_i + (1-\lambda)x_j \tag{1}$$

$$\tilde{y} = \lambda y_i + (1-\lambda)y_j \tag{2}$$

其中，x_i 和 x_j 为在训练过程随机输入两张图片，y_i 和 y_j 为相对应的标签，根据这两对示例 (x_i, y_i) 和 (x_j, y_j) 合成新的训练样本 $\tilde{x}$ 和对应标签 $\tilde{y}$[14]，由此方法扩大训练过程中的样本空间，增加特征的多样性，并且使模型训练时更加注重样本的决策边界[15]。由于添加了原本不属于训练集中的样本信息，可通过降低生成样本的权重减小其对最终输出的不利影响[16]。

2）Albumentations 数据增强库

Albumentations 是能对计算机视觉任务进行快速灵活的数据增强开源库[17]，对比 Pytorch 框架中自带的图像预处理方法，Albumentations 能提供更多选择。其应用简单且功能强大，不仅支持图像分类任务，还为图像的检测和分割提供接口。能够对图像进行像素级或空间级的变换，常见的有添加噪声、RGB 通道转换、色彩抖动、随即亮度、对比度等 70 多种图像处理方法。

针对膝骨关节炎 X 光影像数据集，本文采用随机缩放旋转操作，限制图像在水平垂直方向的移动比例为 6.25%，概率为 0.5；随机打乱图像通道，概率为 0.1；添加模糊和中值模糊两种处理方式，并设置最大模糊度为 3，概率分别为 0.1。通过以上图像处理方式，将放入模型的数据集进行数据增强操作，增加数据复杂度，降低模型过拟合风险，提升模型的泛化能力[18]，处理后图像如图 3 所示。

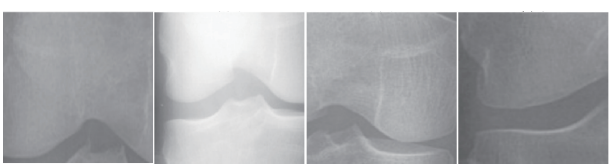


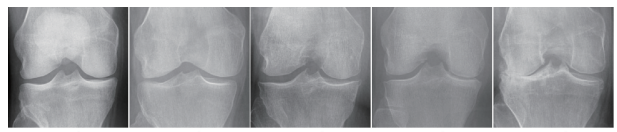
图 3　数据预处理示意图

Fig. 3　Data preprocessing schematic diagram

3　模型训练

3.1　膝骨关节炎 X 光影像数据集

本文使用的 KOA 数据集是从 Kaggle 网站上获取的公开数据集，世界上膝骨关节 X 光图像使用(osteoarthritis initiative，OAI)作为评估标准。OAI 为一项针对 KOA 的前瞻性、纵向及多中心的观察研究，共有 4 796 名年龄在 45 岁到 79 岁的医学学者参加，目的是明确发病进展的生物标志。分类标准采用 KL 等级，其根据 KOA 的严重程度共分为 0～4 级，共 5 个等级：0 等级代表关节间隙正常且边界光滑，属于正常的膝关节；1 等级代表有可疑的膝关节关节间隙狭窄；2 等级代表明确出现小骨赘及可能的关节间隙狭窄；3 等级代表有中等量的骨赘形成，明确出现关节间隙狭窄，可能的膝关节骨性畸形及软骨下骨硬化；4 等级代表严重的关节间隙狭窄骨赘较大、明显的膝关节骨性畸形及严重硬化[19]。数据集中各等级图像如图 4 所示。



(a) 0级 (a) Grade 0　(b) 1级 (b) Grade 1　(c) 2级 (c) Grade 2　(d) 3级 (d) Grade 3　(e) 4级 (e) Grade 4

图 4　膝骨关节炎 X 光影像数据集各等级图像

Fig. 4　Knee joint samples of all KL grades

由于 OAI 是一项多中心研究，从基线队列收集到的原始的膝骨关节炎 X 光片的分辨率和尺寸并不一致，因此需要先将未经筛选的 X 光影像调整为具有相同分辨率的图像，以保证图像大小相同并且膝骨关节位于图像中心，由此制作出具有 7 828 张图片尺寸为 224×224 的数据集。获取数据集后将所有膝骨关节 X 光影像随机拆分为训练集、测试集分别进行模型的训练及测试，按照 8∶2 的比例拆分。拆分后训练集包含 6 265 张图片，测试集包含 1 563 张图片，具体数量如表 1 所示。

表 1　数据集各类数据分布

Table 1　Data distribution of data sets

数据集	0 级	1 级	2 级	3 级	4 级
训练集	2 469	1 133	1 650	824	189
测试集	616	283	412	205	47
合计	3 085	1 416	2 062	1 029	236

3.2　模型配置及训练过程

具体实验配置如下：中央处理器为 Inter Core I9-10900K，内存 32 G，图形处理器为 NVIDIA GeForce RTX 3090 24 G，操作系统为 Ubuntu20.04 LTS 64 位。实验环境为 CUDA 11.7、CuDNN 8.5、Python3.10.11、Pytorch2.0.1。实验过程的超参数设置如表 2 所示。

表 2　模型超参数设置

Table 2　Model hyperp parameter setting

超参数	数值
迭代次数	300
批次大小	16
优化算法	AdamW
初始学习率	0.001

优化算法使用 AdamW，衰减系数 0.05。学习率策略设置两个阶段，首先在前 20 次迭代进行学习率升温 Warmup 方法，初始学习率因子为 0.001，每次迭代都更新学习率；20 次迭代后采用余弦退火策略，并设置最小学习

率为 0.000 01。

模型的训练过程可以分为前向、反向传播,其中前向传播包含线性归一化、计算 W-MSA 和 SW-MSA 和损失函数。对于一次迭代中的每一个 batchsize 数据,执行以下操作:首先通过骨干网络的 Swin Transformer Block 堆叠使用两种多头自注意力机制进行由浅入深的特征提取,其次进入由全局平均池化构成的模型颈部,将特征图拉成一维向量并且送入线性分类头进行分类,前向传播得到预测结果和误差,再通过反向传播求出网络的梯度;最后进行网络的权重更新,利用交叉熵损失函数调整权重和偏差,不断减小损失函数值。使得预测值更加接近真实值,提升模型分类准确率。

4　实验验证

4.1　评价指标

评估深度学习模型的分类性能是研究项目的重要部分,依靠准确率这一指标来度量模型可以给出令人满意的结果,但是当对照精度或者其他角度而言,单纯使用准确率会给出较差的结果。因此,在大多数情况下会应用各种指标来衡量和比较网络性能。本研究为对膝骨关节炎 KL 等级分类模型的构建,属于图像分类中的单标签分类任务,评价此类任务的指标有准确率(Accuracy)、精准率(Precision)、召回率(Recall)、F1 分数(F1-score)。

本文中 KOA 严重程度分类分为 5 个类别,准确率(Accuracy)衡量模型正确识别所有类型的能力:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{3}$$

精准率(Precision)也称为阳性预测值,描述预测的正样本中确定为此类 KOA 严重等级的比例算式如下:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

召回率(Recall)描述模型对于基础事实的正面预测完整性,它表示的是正样本被找出的比例:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

对于精确率与召回率在理想情况下两个参数都高,模型越好,但是它们彼此是相互对立的,Precision 高对应 Recall 低。为了在保证精确率的条件下,提升召回率引入新的指标 *F1-score*,其代表二者的调和平均值:

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

其中,真阳性(true positive, TP)表示根据给定标签的严重等级类别被正确预测为给定的类别的数量;真阴性(true negative, TN)则表示根据给定标签正确预测不属于该严重等级类别的数量;假阳性(false positive, FP)是模型错误预测属于给定标签类别的数量;假阴性(false negative, FN)是模型错误预测不属于给定标签类别的数量。正样本为预测的严重程度即为真实标签类别,负样本为预测的严重程度不属于此类真实标签。对于准确率、精准率、召回率和 F1 分数这 4 个指标而言,越接近于 1 代表模型性能越好。

4.2　消融实验

为验证改进算法和添加数据增强模块的有效性,进行消融实验。本文对 Swin Transformer 模型结构有两点改进,分别在颈部和头部网络当中,在颈部网络中使用两层 FC 层和 ReLU 激活函数替换掉全局平均池化层,在头部网络中在原本线性分类器的单层 FC 层的基础上添加新的隐藏层和 Tanh 激活函数,对此将改进的颈部和头部网络分别命名为 FC Neck 和 FC Head。在原模型基础上逐个加以改进,进行如下 5 组实验,实验结果如表 3 所示。

表 3　消融实验结果

Table 3　Results of ablation experiment

Swin Transformer	FC Head	FC Neck	Augment	FLOPs/G	参数量/M	准确率/%
1	—	—	—	106.49	199.68	68.778 0
2	√	—	—	106.49	199.69	73.704 4
3	—	√	—	106.49	199.73	74.600 1
4	√	√	—	106.49	199.74	75.303 9
5	√	√	√	106.49	199.74	76.071 7

实验 1 为原模型准确率为 68.7%,实验 2 在原模型基础上改进头部网络,在头部网络中添加隐藏层,使得网络中的输出特征有更多的组合机会,对于最终的分类结果有很大提升,提升接近 5 个百分点;实验 3 在原模型基础上使用 FC 层替换掉颈部网络中的全局平均池化层,经过对比可以看出,FC 层相比全局平均池化层具有保护源域与目标域跨度较大的迁移学习的能力,准确率提升接近 6 个百分点。实验 4 中同时改进颈部网络与头部网络,经过主干网络的特征提取,后续层中添加了 3 层 FC 层和 2 种不同的损失函数,其中特征进行更多的非线性组合,可以看出效果有所提升但没有达到预计的良好水平,非线性组合特征有些是相似的在原模型基础上提升 6 个以上百分点,在实验 3 基础上提

升 0.6 个百分点。实验 5 中本文将两种模型结构的改进和两种数据增强方法同时作用在原模型中，3 种改进方式提升 7 个以上百分点，获得了在 Kaggle 网站的膝骨关节炎 X 光影像公开数据集中最佳的分类精度，达到了 76%。对比各实验的参数量变化，可以看出添加 FC 层与激活函数层的方式模型参数量与浮点运算量并无大幅度增大。以上实验证明本文对 Swin Transformer 模型的三点改进能够有效提升模型的分类精度，提升模型的泛化能力。

4.3　对比实验

为评估不同网络模型的性能，本文分别选取包含传统卷积神经网络、31×31 的大卷积核网络、纯卷积网络、基于 Transformer 的多种变体网络、Swin Transformer 以及本文改进后的 Swin Transformer 在内的共计 17 种网络模型进行对比实验。各模型在训练过程中均使用基于 ImageNet-1K 数据集的预训练权重进行迁移学习，以加快模型收敛速度及提高模型准确率，结果如表 4 所示。

表 4　对比实验结果

Table 4　Comparative experimental results

网络名称	精确度	召回率	F1 分数	FLOPs/G	参数量/M	准确率/%
ResNet-50	63.704 3	56.585 9	57.792 0	5.400	25.444	62.955 9
VGG-19	0.321 5	0.304 7	0.309 5	19.662	147.450	63.147 8
Inception-V3	64.671 4	56.442 1	57.372 7	2.845	21.796	61.868 2
RepLKNet	58.621 3	57.329 1	53.207 3	63.242	340.990	60.268 7
RepMLP	66.069 0	61.109 7	62.763 6	9.693	95.682	66.794 6
ConvNeXt-V2	68.656 9	63.619 3	65.267 9	117.760	672.760	65.707 0
DaViT	65.989 5	62.120 5	63.752 2	15.510	86.935	64.633 3
MViT	64.407 4	59.875 7	61.446 4	10.158	50.708	64.555 3
MobileViT	65.391 5	62.309 9	63.147 8	1.441	4.941	64.683 3
XCiT	65.722 3	63.117 7	63.792 9	63.525	83.813	65.323 1
RIFormer	66.465 9	58.839 6	61.187 3	11.596	72.744	63.915 6
PoolFormer	64.821 9	57.854 2	60.078 3	11.590	72.708	64.043 5
Conformer	63.951 4	59.481 0	60.947 1	22.864	81.186	63.915 6
Twins	66.082 1	61.161 4	62.661 5	8.475	55.306	64.235 5
TNT	69.200 3	67.455 5	68.104 5	5.240	23.372	67.882 3
Swin Transformer	68.730 5	67.281 9	67.223 6	106.490	199.680	68.778 0
Improve-Swin Transformer	77.878 6	75.588 0	76.279 9	106.490	199.740	76.071 7

通过对比实验可以看出，Swin Transformer 模型相比于传统卷积神经网络有着更好的全局建模能力，在各类模型当中获得最佳分类精度。改进后的 Swin Transformer 模型相比原模型有着稍大的浮点运算量，模型较为复杂。在精准率、召回率、F1 分数以及准确率方面都有着最好的指标，证明本文的改进算法性能的优越性。

4.4　可视化分析

本文使用梯度类激活热力图 Grad-CAM 进行可视化，其利用特征图加权求和形成类激活图，显示出图片中某个区域对分类结果的重要程度。图 5 为膝骨关节炎 X 光影像数据集中各类图像在原模型与改进模型中的类激活图，类激活图颜色越深代表该区域对识别结果的贡献越大。可以看出对比原模型改进模型的关注区域更加精准，感兴趣区域集中于骨骼间隙位置，提升模型的算力分匹配能力。

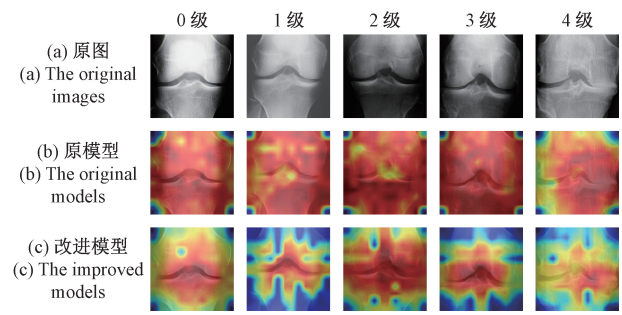


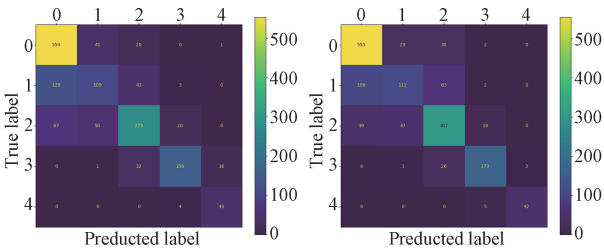
图 5　原模型与改进模型的 Grad-CAM 类激活热力图

Fig. 5　Grad-CAM activation heat map of the original model and the improved model

4.5　混淆矩阵

本文绘制原模型与改进模型的混淆矩阵，如图 6 所示，其能够直观显示出模型预测各类别的准确程度，从而发现模型在哪些类别的预测上表现不佳。混淆矩阵的纵轴为真实标签信息，横轴为模型的预测标签信息，处于矩阵对角线上的元素为预测正确的样本数量，矩阵中每行元素相加为测试集中各类别的样本总数。

根据混淆矩阵之间的对比可以推算出模型各等级的分



(a) 原模型混淆矩阵
(a) Original model confusion matrix

(b) 改进模型混淆矩阵
(b) Improved model confusion matrix

图 6　混淆矩阵对比

Fig. 6　Confusion matrix comparison

类准确率如表 5 所示。可以看出,总体准确率有所上升的同时第 4 级的分类精度稍有下降。经过对比分析 KL 等级各类别的特征发现各等级界限较为模糊。各类别准确率中 1 级分类准确率最低,原始模型容易将 1 级预测为 0 级,经过模型改进预测正确的样本数量有所提升,而将 1 级预测为 2 级的错误率却也有所提高。从 0～3 级准确率皆有所提升,其中 2、3 级提升较大,提升 7%以上。唯有 4 级稍有下降,但总体上模型的分类精度有着较大提升。通过混淆矩阵验证模型改进能够有效提升 2、3 级预测效果。

表 5　原模型与改进模型各类准确率对比

Table 5　Comparison of various accuracy rates　%

模型	0 级	1 级	2 级	3 级	4 级
原模型	89.9	38.5	66.7	76.1	91.5
改进模型	90.1	39.6	74.5	84.4	89.4

4.6　泛化实验

为验证本文改进的 Swin Transformer 算法对于其他医学影像数据集的泛化能力与稳定性,本文分别开展同领域其他膝骨关节炎 X 光影像数据集与不同领域医学影像数据集的泛化实验。分别选取在对比实验中表现较好的 4 种模型与本文的改进模型进行泛化能力测试,在保持原本数据处理流程和超参数设置不变的情况下将泛化数据集送入模型中进行实验。

1) 同领域其他膝骨关节炎 X 光影像数据集泛化实验

为验证本文算法在同领域其他数据集上的性能,在 Github 网站获取 Chen 等使用的膝骨关节炎 X 光影像数据集,在此数据集上其获得 69.7%的分类准确率。该数据集按照 8∶2 的比例拆分为训练集与测试集,各类别数据分布和实验结果分别如表 6 和 7 所示。

表 6　泛化膝骨关节炎 X 光影像数据集数据分布

Table 6　Data distribution of generalized dataset

数据集	0 级	1 级	2 级	3 级	4 级
训练集	2 286	1 046	1 516	757	173
测试集	639	296	447	223	51
合计	2 925	1 342	1 963	980	224

从实验结果可以看出,其他 4 种模型,在原始数据集上有着 65%以上的精度,但在泛化数据集上表现一般。而本文改进的 Swin Transformer 模型获得 73.4%的分类准确率,对比 Chen 等最佳的 69.7%分类准确率,提升 3.5 个百分点。实验结果表明,本文的改进模型在同领域其他膝骨关节炎 X 光影像数据集上具有较好的泛化能力。

表 7　泛化膝骨关节炎 X 光影像数据集实验结果

Table 7　Experimental results of generalized X-ray image data set of knee osteoarthritis

网络名称	精确度	召回率	F1 分数	FLOPs/G	参数量/M	准确率/%
ConvNeXt-V2	63.860 3	58.508 2	59.945 3	117.760	672.760	62.500 0
TNT	62.040 7	58.836 5	59.528 7	5.240	23.372	63.768 1
XCiT	68.254 1	55.418 6	57.484 7	63.525	83.813	61.352 7
RepMLP	62.295 1	55.084 4	54.438 0	9.693	95.682	61.594 2
Swin Transformer-FC	74.925 9	71.785 4	72.795 6	106.490	199.740	73.369 6

2) 不同领域医学影像泛化实验

为验证本文算法在其他领域医学影像数据集上的泛化能力,本文在 Kaggle 网站上获取结肠疾病无线胶囊内窥镜图像数据集,在该数据集上 Gregson 等获得的最佳分类精度为 99.0%。该数据集仍按照 8∶2 的比例拆分为训练集与测试集,各类别数据分布和实验结果分别如表 8 和表 9 所示。

本组实验中改进的 Swin Transformer 模型与轻量化的 Transformer(transformer in transformer, TNT)模型获得了 99.6%的分类精度,此结果超过当前在该数据集上的最佳分类精度。实验结果证明,由于膝骨关节炎 X 光影像自动诊断任务的特殊性,导致改进模型的分类精度仅为

表 8　结肠疾病无线胶囊内窥镜图像数据集数据分布

Table 8　Data distribution of wireless capsule endoscope image dataset for colon diseases

数据集	溃疡	息肉	食道炎	正常
训练集	800	800	800	800
测试集	200	200	200	200
合计	1 000	1 000	1 000	1 000

表 9　结肠疾病无线胶囊内窥镜图像数据集结果

Table 9　Results of wireless capsule endoscope image data set for colon diseases

网络名称	精确度	召回率	F1 分数	FLOPs/G	参数量/M	准确率/%
ConvNeXt-V2	99.002 4	99.000	99.000 0	117.760	672.760	99.000
TNT	79.704 4	79.700	79.700 0	5.240	23.372	99.625
XCiT	98.889 9	98.875	98.874 8	63.525	83.813	98.875
RepMLP	99.038 5	99.000	98.999 6	9.693	95.682	99.000
Swin Transformer-FC	99.625 6	99.625	99.625 0	106.490	199.740	99.625

76.0%。这一结果表明，膝骨关节炎 X 光影像自动诊断任务具有独特的挑战，比如模糊的分类界限以及不均衡的数据分布，这些因素影响模型的分类性能，若将该模型用于其他领域医学影像数据集便能有较好的性能，也再次证明所提出的改进模型在性能上的卓越表现不仅限于特定的数据集。通过多样化的数据集测试，表明该模型能够在各种医学影像应用场景下保持高水平的性能。

5　结　　论

为实现膝骨关节炎 X 光影像自动诊断，本文提出一种改进 Swin Transformer 的深度学习算法。通过两层全连接层加 ReLU 激活函数的结构替换颈部网络的全局平均池化层，对迁移学习进行保护；在头部网络中添加全连接层与 Tanh 激活函数，组合出更多非线性特征；在数据预处理和模型训练过程中，分别依托 Albumentations 库和添加 Mixup 模块以此实现数据增强处理。实验结果表明，所提出的改进方法能有效提升模型性能，使得模型更加关注图像中骨关节间隙位置；同时，其具备处理其他医学影像的能力，有着较好的稳定性与泛化能力。

参考文献

[1] CONAGHAN P G, PORCHERET M, KINGSBURY S R, et al. Impact and therapy of osteoarthritis: The arthritis care OA nation 2012 survey[J]. Clinical Rheumatology, 2015, 34: 1581-1588.

[2] ANTONY J, MCGUINNESS K, O'CONNOR N E, et al. Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks[C]. 2016 23rd International Conference on Pattern Recognition(ICPR), IEEE, 2016: 1195-1200.

[3] ANTONY J, MCGUINNESS K, MORAN K, et al. Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks[C]. Machine Learning and Data Mining in Pattern Recognition: 13th International Conference, 2017: 376-390.

[4] TIULPIN A, THEVENOT J, RAHTU E, et al. Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach[J]. Scientific Reports, 2018, 8(1): 1727.

[5] GORRIZ M, ANTONY J, MCGUINNESS K, et al. Assessing knee OA severity with CNN attention-based end-to-end architectures[C]. International Conference on Medical Imaging with Deep Learning, PMLR, 2019: 197-214.

[6] CHEN P, GAO L, SHI X, 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.

[7] CUEVA J H, CASTILLO D, ESPINOS M H, et al. Detection and classification of knee osteoarthritis[J]. Diagnostics, 2022, 12(10): 2362.

[8] GU H, LI K, COLGLAZIER R J, et al. Knee arthritis severity measurement using deep learning: a publicly available algorithm with a multi-institutional validation showing radiologist-level performance[J]. ArXiv preprint arXiv:2203.08914, 2022.

[9] 王国桢，卢国杰，王桂棠. 无人化起重装卸的目标物实例分割模型研究[J]. 电子测量技术，2023，46(18)：139-146.
WANG G ZH, LU G J, WANG G T. Instance segmentation model of uncertain object in unmanned lifting and handling scenarios [J]. Electric Measurement Technology, 2023, 46(18): 139-146.

[10] NIRTHIKA R, MANIVANNAN S, RAMANAN A, et al. Pooling in convolutional neural networks for medical image analysis: A survey and an empirical study[J]. Neural Computing and Applications, 2022, 34(7): 5321-5347.

[11] ZAFAR A, AAMIR M, MOHDNAWI N, et al. A comparison of pooling methods for convolutional neural networks[J]. Applied Sciences, 2022, 12(17): 8643.

[12] ZHANG C L, LUO J H, WEI X S, et al. In defense of fully connected layers in visual representation transfer[C]. Pacific Rim Conference on Multimedia, Cham: Springer International Publishing, 2017: 807-817.

[13] ZHANG H, CISSE M, DAUPHIN Y N, et al.

mixup: Beyond empirical risk minimization[J]. ArXiv preprint arXiv:1710.09412,2017.

[14] 吕佳,邱小龙.基于 K-means 聚类和特征空间增强的噪声标签深度学习算法[J].智能系统学报,2024,19(2):267-277.
LYU J, QIU X L. A noisy label deep learning algorithm based on K-means clustering and feature space augmentation [J]. CAAI Transactions on Intelligent Systems, 2024,19(2): 267-277.

[15] 李华超.混合图像数据增强方法在细粒度分类的研究[D].南京:南京邮电大学,2023.
LI H CH. Research on fine-grained classification based on hybrid image data enhancement method[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2023.

[16] 陈仁祥,张旭,徐向阳,等.噪声标签下注意力特征混合的旋转机械故障诊断[J].仪器仪表学报,2023,44(9):257-264.
CHEN R X, ZHANG X, XU X Y, et al. Fault diagnosis of rotating machinery with attentive feature mixup in noisy labels[J]. Chinese Journal of Scientific Instrument, 2023, 44(9):257-264.

[17] BUSLARV A, IGLOVIKOV V I, KHVEDCHENYA E, et al. Albumentations: Fast and flexible image augmentations[J]. Information, 2020, 11(2): 125.

[18] KAMARDI C, LAKSANA I K P B, ANGGREAINY M S, et al. Classification of alzheimer's disease using random oversampling and albumentations on convolutional neural network [C]. 2023 Eighth International Conference on Informatics and Computing(ICIC), IEEE, 2023: 1-6.

[19] 刘伟强,罗林开,彭洪,等.基于自适应序列罚权深度神经网络的膝骨关节炎等级评分[J].仪器仪表学报,2021,42(7):145-154.
LIU W Q, LUO L K, PENG H, et al. Grading scoring of knee osteoarthritis based on adaptive ordinal penalty weighted deep neural networks[J]. Chinese Journal of Scientific Instrument, 2021, 42(7): 145-154.

作者简介

许超,高级实验师,硕士,主要研究方向为智能感知与信息处理。
E-mail: xuchao@lnu.edu.cn

王云健,硕士研究生,主要研究方向为图像检测。
E-mail: wangyunjianlnu@163.com

刘洋,本科,主要研究方向为图像处理。
E-mail: 2992701431@qq.com

卢雪梅,副教授,博士,主要研究方向为电子信息工程。
E-mail: luxuemei@lnu.edu.cn

丁勇(通信作者),教授,博士生导师,主要研究方向为光学图像认知理论研究。
E-mail: dingyong@lnu.edu.cn