

基于YOLOv8m的改进腕部X光片骨折检测算法

彭志博, 陈勇, 崔艳荣

长江大学计算机科学学院, 湖北 荆州 434000

【摘要】目前腕部X光片的骨折检测存在误诊率高、医疗资源不足等问题。为了辅助医生进行骨折诊断,提出了一种基于YOLOv8m的X光片骨折检测方法。首先引入可分离大核注意力机制来提取重要特征信息,抑制不显著特征信息;然后将残差块融入注意力机制,增强注意力机制的作用,增加模型的泛化能力;最后将可切换空洞卷积与C2f模块结合,增加模型的感受野,捕捉不同尺寸的特征信息。实验结果表明,与先进的YOLOv8l改进模型相比,本文模型mAP50提高了1.3%,由于使用了规格更小的YOLOv8m为基础模型,参数量降低了14.3%,浮点运算次数降低了42.7%。此模型能够辅助放射科医生进行腕部X光片的骨折诊断。

【关键词】X射线;骨折检测;深度学习;YOLOv8

【中图分类号】R318;TP391

【文献标志码】A

【文章编号】1005-202X(2025)04-0542-08

Fracture detection in wrist X-ray image using an improved algorithm based on YOLOv8m

PENG Zhibo, CHEN Yong, CUI Yanrong

School of Computer Science, Yangtze University, Jingzhou 434000, China

Abstract: Currently, the fracture detection in wrist X-ray image has high misdiagnosis rates and faces the challenge of inadequate medical resources. To assist doctors in fracture diagnosis, an improved approach based on YOLOv8m for fracture detection in wrist X-ray image is proposed: (1) a large separable kernel attention mechanism is introduced to extract crucial feature information while suppressing insignificant ones; (2) residual block is integrated into the attention mechanism to enhance its effectiveness and the model's generalization ability; (3) switchable atrous convolution is combined with the C2f module to expand the model's receptive field, enabling it to capture multi-scale feature information. Experimental results demonstrate that compared with the improved model based on the advanced YOLOv8l, the proposed approach achieves a 1.3% increase in mAP50. Notably, by adopting the more compact YOLOv8m model as the basic model, parameter count is reduced by 14.3%, and the floating-point operations per second is lowered by 42.7%. The proposed model can effectively aid radiologists in detecting fractures in wrist X-ray image.

Keywords: X-ray; fracture detection; deep learning; YOLOv8

前言

腕部骨折为常见的外伤骨折类型^[1],该骨折多发生在桡骨远端近关节2 cm处。若不针对具体情况及时进行治疗,可能会导致手腕关节畸形,造成手腕活动范围受限、握力减小等功能性问题^[2]。若儿童发生手腕部骨折没有得到正确的治疗,更有可能导致终身不便。

目前,骨折检查普遍使用X射线、计算机断层扫

描(CT)以及核磁共振成像(MRI)3种医学成像设备^[3]。其中,X射线因具有成本低、检查时间短、密度分辨率高等优点,为骨折的首选影像学评价方式之一^[4]。然而,因为腕部骨质结构复杂且排列不规则,骨科医生在阅片时极易造成漏诊、误诊等现象^[5],这类问题在医疗水平欠发达的地区更为突出。计算机辅助诊断技术为解决这类问题提供了一定的帮助。随着深度学习以及医学成像技术的不断发展,越来越多的研究者尝试将神经网络融入到计算机辅助诊断技术。图像处理技术中的目标检测任务能够帮助完成腕部骨折的检测。

YOLO算法^[6]是最常见的目标检测算法之一。事实上,已经有许多研究人员将YOLO系列算法运用在手腕骨折的检测当中。Dibo等^[7]将GAM注意力机制以及Transformer模块插入YOLOv7算法^[8]的颈

【收稿日期】2024-12-20

【基金项目】国家自然科学基金(62077018)

【作者简介】彭志博,硕士研究生,研究方向:深度学习、计算机视觉,
E-mail: 2023720815@yangtzeu.edu.cn

【通信作者】陈勇,高级工程师,硕士生导师,研究方向:WEB信息处理、人工智能,E-mail: 285527563@qq.com

部网络中,加强了网络模型对信息的提取能力并且能对提取的信息进行分类。Chien等^[9]将多种注意力机制以及全局上下文机制分别插入在YOLOv8模型的颈部网络中,对腕部X光片进行骨折检测并比较了这些机制对网络模型的提升,其中全局上下文机制的改进模型效果最好。最近,Chien等^[10]使用最近发布的YOLOv9模型进行改进实验,但效果不如融合了全局上下文机制的YOLOv8l模型。

虽然上述YOLO系列算法的改进方法在精度方面均有提升,但都需要在模型大小为1的基础上性能才能达到最优,大小为1的模型参数量大,训练时间长,对GPU也有较高的要求。尤其是改进后的YOLOv9-C以及YOLOv9-E的参数量也较大,训练将花费较长时间。本文以参数量更少的YOLOv8m模型为基础,先与可分离大核卷积注意力机制(Large Separable Kernel Attention, LSKA)相结合并进行二次改进,然后使用可切换空洞卷积替换部分模块中的卷积操作。经实验证明,与其他改进模型相比较,本文模型不仅进一步提高了基于YOLOv8改进模型的检测精度,参数量以及浮点运算次数大幅度降低,减少了模型在数据集上的训练时间。

1 YOLOv8 目标检测算法

YOLOv8算法^[11]是Ultralytics于2023年推出的YOLO系列算法,YOLOv8的网络结构由输入端(Input)、骨干网络(Backbone)、颈部网络(Neck)以及头部网络(Head)4部分组成。

输入端(Input)与YOLOv5算法基本一致,采用了Mosaic、Mixup数据增强方法以及自适应锚框计算、自适应图片缩放来处理数据^[12]。

骨干网络(Backbone)采用了跨阶段局部(Cross Stage Partial, CSP)的方法^[13]。该方法可以在减少模型计算量、提高计算速度同时,保持模型的精度。YOLOv8算法在骨干网络用C2f模块替换了YOLOv5中的C3模块,C2f模块集成了C3模块和YOLOv7算法中ELAN(Efficient Layer Aggregation Networks)的设计^[14]。其次也将YOLOv5算法中的空间金字塔池化结构(Spatial Pyramid Pooling, SPP)模块替换成快速空间金字塔池化结构(Spatial Pyramid Pooling-Fast, SPPF)模块。两者都有效减少了计算量。

颈部网络(Neck)使用了多种尺度的特征融合,将骨干网络不同阶段的特征图进行融合增强。其保留了YOLOv5模型中PAN(Path Aggregation Network)^[15]加FPN(Feature Pyramid Network)^[16]架构。PAN自底向上提取特征,将下层显著特征信息传递到上层,而FPN自顶向下提取特征,保证下层特

征图包含了丰富的特征信息。

在头部网络(Head),YOLOv8采用了解耦头,将分类和检测任务分离。这种设计能让每个部分专注于特定的任务,提高了分类的准确性和检测的精确度。对于分类任务的损失函数,使用了二元交叉熵(Binary Cross-Entropy, BCE)损失函数,其表达式为:

$$\text{LOSS}_{\text{BCE}} = -w \times [y_n \times \log(x_n) + (1 - y_n) \times \log(1 - x_n)] \quad (1)$$

其中, w 代表权重, y_n 表示标记值, x_n 表示预测值。对于检测任务,YOLOv8采用了分布焦点损失函数(Distributed Focal Loss, DFL)^[17]以及CIoU(Complete Intersection over Union)损失函数^[18],DFL和CIoU损失函数公式分别如下:

$$\text{LOSS}_{\text{DFL}} =$$

$$-\left[(y_{n+1} - y) \times \log\left(\frac{y_{n+1} - y}{y_{n+1} - y_n}\right) + (y - y_n) \times \log\left(\frac{y - y_n}{y_{n+1} - y_n}\right) \right] \quad (2)$$

$$\text{LOSS}_{\text{CIoU}} = 1 - \text{IoU} + \frac{d^2}{c^2} + \frac{v^2}{(1 - \text{IoU}) + v} \quad (3)$$

其中,IoU为预测样本框和真实样本框之间的重叠度, d 表示了预测样本框和真实样本框中心点的欧氏距离, c 表示同时包含预测样本框和真实样本框的最小包围框的对角线长度, v 代表量化长宽比一致性的参数, v 的表达式为:

$$v = \frac{4}{\pi^2} \times \left(\arctan \frac{w_{gt}}{h_{gt}} - \arctan \frac{w_p}{h_p} \right)^2 \quad (4)$$

其中, w 代表权重, h 表示边界框的高度。

2 基于YOLOv8m的改进检测算法

2.1 颈部网络的改进

2.1.1 LSKA 腕部X光片复杂多样,在骨折后会引起的并发症,增加了X光片中的检测种类,使其包含了更为丰富的特征信息,而在众多的特征信息中去关注显著特征,抛弃不必要的特征能有效提高模型的精度。在目标检测任务中,注意力机制能够使模型在观察图像时更有针对性,能够对图像中的重要信息进行提取,并且抑制无用信息。近年来注意力机制在目标检测领域取得了优异的成绩。本文引用了LSKA^[19],该注意力机制是在大核注意力机制(Large Kernel Attention, LKA)的基础上改进而得来的^[20]。具体改进方法为将LSKA机制放置在颈部网络中C2f模块后面。

LKA是一种线性注意力机制,吸收了卷积以及自注意力机制的优点。LKA模块首先利用了标准的深度卷积和小感受野核来捕捉局部信息^[21],使用扩展的深度卷积以及大的感受野核来建立远程依赖,在经过两类卷积操作后,将输出进行 1×1 的卷积操作生成注意力图,最后将该注意力图与最开始的输出特

征相乘来进行自适应特征细化。但 LKA 深度卷积的设计导致其计算效率低下。为了解决该问题。LSKA 的提出者保留了 LKA 模块对输入特征的处理方式以及结构。只是将深度卷积设计成可分离的卷积,将 $K \times K$ 的卷积核平均划分为 $1 \times K$ 和 $K \times 1$ 的卷积核,可分离卷积核的提出减少了随着核大小增加而增加的参数量,提高了计算效率。LKA 以及 LSKA 的结构图如图 1 所示。

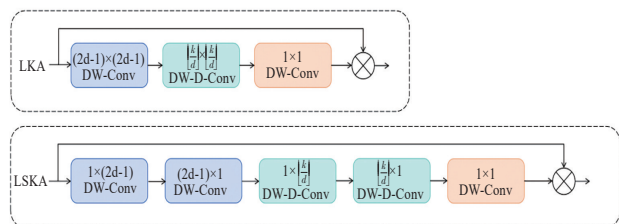


图1 LKA 和 LSKA 结构图

Figure 1 LKA and LSKA structures

LSKA 的输出如下所示:

$$\bar{Z}^C = \sum_{h,w} W_{(2d-1) \times 1}^C * \left(W_{1 \times (2d-1)}^C * F^C \right) \quad (5)$$

$$Z^C = \sum_{h,w} W_{\left\lfloor \frac{k}{d} \right\rfloor \times 1}^C * \left(W_{1 \times \left\lfloor \frac{k}{d} \right\rfloor}^C * \bar{Z}^C \right) \quad (6)$$

其中,*和 \otimes 分别表示卷积操作和阿达玛积, h,w 分别代表特征图的高度和宽度, W 代表卷积核,下标代表卷积核大小, d 表示空洞率,表示卷积核处理数据时各值的间距, F^C 为输入的特征映射,式(5)将特征映射与 $(2d-1) \times 1$ 及 $1 \times (2d-1)$ 的卷积核进行卷积,其作用是捕捉局部信息。式(6)中卷积核的下标 $\left\lfloor \frac{k}{d} \right\rfloor$ 表示 k 除以 d 向下取整, k 表示卷积核的最大感受野。式(6)将式(5)的结果与 $\left\lfloor \frac{k}{d} \right\rfloor \times 1$ 及 $1 \times \left\lfloor \frac{k}{d} \right\rfloor$ 的卷积核进行卷积,负责捕捉 \bar{Z}^C 中的全局信息。

式(7)表示深度卷积的输出 A^C 和 1×1 的卷积核进行卷积操作。LSKA 的输出即为式(8),结果为 A^C 与特征映射 F^C 的阿达玛积。

$$A^C = W_{1 \times 1} * Z^C \quad (7)$$

$$\bar{F}^C = A^C \otimes F^C \quad (8)$$

2.1.2 LSKA 的二次改进 注意力机制的引入会增加网络复杂度,可能会导致模型在深层网络中难以得到更好的训练。因此,本文将残差模块(ResBlock)^[22]与 LSKA 相结合。卷积注意力机制(Convolutional Block Attention Module, CBAM)^[23]的提出者在实验中将 CBAM 集成于残差网络(ResNet)中,后续有研究人员据此将 CBAM 机制与 ResNet 中的 ResBlock

模块融合形成 ResCBAM 机制。

ResBlock 是一种含跳跃连接的卷积模块,可以直接将输入内容直接传输到网络深层,并保持信息的完整性。ResBlock 能够一定程度上解决神经网络在深层训练出现梯度爆炸或者梯度消失的现象,使神经网络能在深层得到更好的训练。受到 CBAM 与 ResBlock 融合的启发,本文将 LSKA 与 ResBlock 相结合(ResLSKA),进一步提高模型在网络深层对信息的提取能力,加强网络在深层的训练,并且同时提高网络的泛化能力(图 2)。

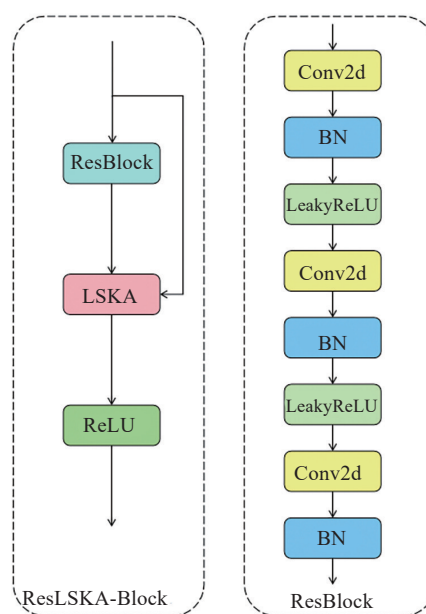


图2 ResLSKA 结构图

Figure 2 ResLSKA structure

2.2 骨干网络(Backbone)的改进

腕部骨折因受伤程度的不同,会导致受伤面积不一致,进行辅助固定的钢板大小也各有不同,这一情况会产生不同尺度的检测目标并且检测目标的变化性较大。在骨干网络中,C2f模块通过多次卷积操作能够有效提取和转换输入数据的特征,经过 C2f 模块的输出更具表征能力,但单一的卷积操作提取出来的特征尺度相同,缺乏多样性。为了使 C2f 模块能够获取更为丰富的不同尺度的空间特征信息,本文选择用可切换空洞卷积(Switchable Atrous Convolution, SAC)来替换 C2f 模块中的部分卷积操作^[24]。其结构如图 3 所示。

SAC 的架构包括 3 部分:在输入输出段都有两个全局上下文(Global Context)模块^[25],中间的 SAC 组件就是两个不同空洞率的 3×3 卷积操作加上一个转化机制组成,转化机制通过一层池化和一层 1×1 的卷积操作来实现。在两层不同空洞率的卷积操作之间

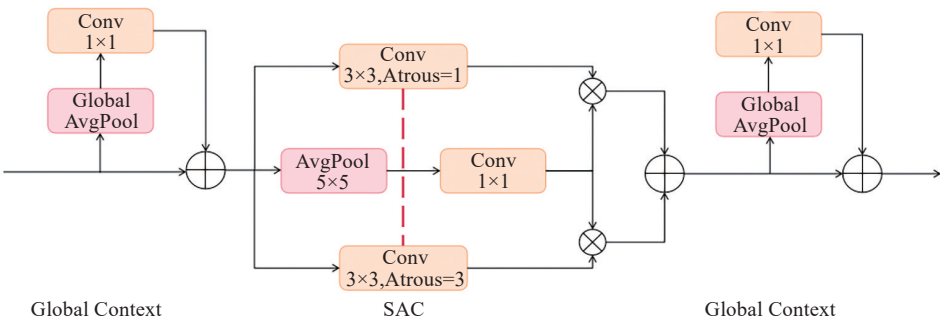


图3 可切换空洞卷积结构图

Figure 3 Switchable atrous convolution structure

有一个开关函数,负责将两个卷积操作的输出结果组合在一起。SAC的核心思想就是将相同的输入特征进行不同空洞率的卷积操作,空洞卷积可以在卷积核中增加额外的空间,增加了感受野,而不同的空洞率可以用来捕捉不同尺度的特征,可以更好地处理数据。C2f模块具体的改进方法是将SAC替换C2f模块中Bottleneck模块中的第二层卷积,其中SAC模块中两个空洞卷积的卷积率根据SAC提出者的原模型进行选择,分别设置为1和3。C2f模块改进后的C2f_SAC模块具体结构如图4所示。

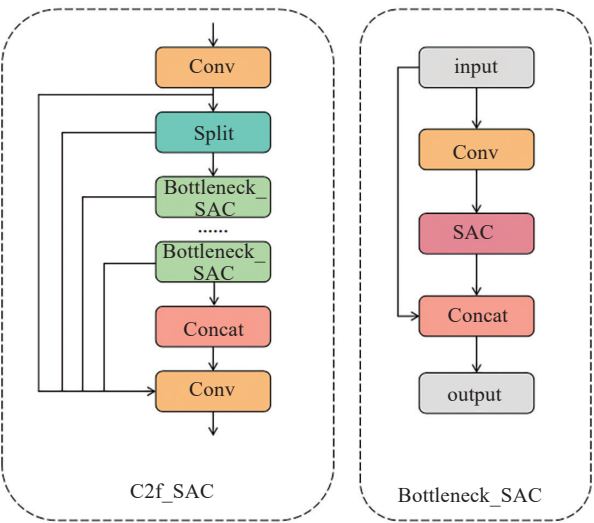


图4 C2f_SAC结构图

Figure 4 C2f_SAC structure

考虑到SAC操作会增加参数量降低计算速度,为了平衡模型的精度和计算效率,只替换掉部分C2f模块。其中,骨干网络的第二、第三个C2f模块的输出将作为颈部网络Concat模块的输入且Concat模块的输出流向颈部网络的C2f模块,为了前后保持一致,替换策略为将第一、第四个C2f模块替换为C2f_SAC模块。最终改进模型结构如图5所示。

3 实验设计与结果分析

3.1 数据集及其预处理

GRAZPEDWRI-DX数据集是格拉茨医科大学公布的公开数据集^[26],这些数据集内容是儿童手腕创伤X光片图像,这些图像是格拉茨医科大学多名放射科医生在2008年~2018年收集的,包含了6 091位患者,一共有20 327张图片、75 559个图片标签以及67 771个标签目标,一共有9类标签。GRAZPEDWRI-DX数据集的发布者并没有设定训练集、验证集以及测试集。所以进行随机划分,训练集、验证集、测试集的比例为7:2:1。该数据集X光片图像的亮度较为一致,仅在这些图片上训练的模型可能在预测其它X光片图像的时候效果不佳。为了提高模型的泛化能力,对训练集进行了数据增强,增强方式为微调图片的亮度以及对比度。具体方法为使用OpenCV中提供的addWeighted函数对图片的对比度和亮度进行微调,其中参数contrast设置为1.2, brightness设置为30。

在全部标签中,含有标签类“foreignbody(异物)”的图片十分稀少,只有8张,在随机划分的过程中很有可能同时分到测试集、验证集或者训练集导致模型无法训练或者预测,为了防止此情况的发生,决定将此类删除。

3.2 实验配置及其参数

本文进行的实验没有设置预训练权重,在预处理后的GRAZPEDWRI-DX数据集上使用改进后的YOLOv8m模型训练。Ju等^[27]的实验表明,该数据集在原始YOLOv8l模型上训练时在60个epoch左右将达到最佳效果,其他以YOLOv8l以及YOLOv9为基础的改进模型设置的epoch为100,由于本研究使用参数量更小的YOLOv8m模型,所以将训练的epoch设为80,同时Ju等在对比优化器的消融实验当中,SGD优化器表现最好,因此优化器选择SGD。设置输入图片大小为640 px, batchsize设为32。其他超参数设置根据YOLOv8模型开发者的默认设置,将权

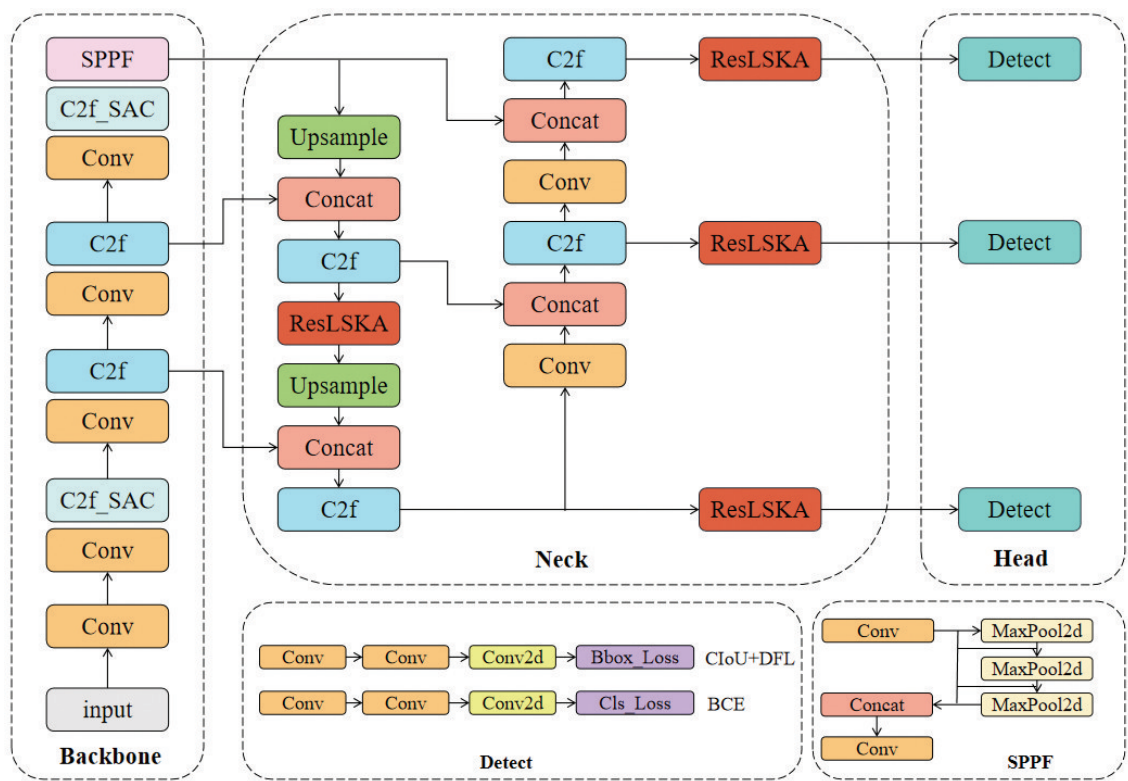


图5 改进后的YOLOv8结构图
Figure 5 Improved YOLOv8 structure

重衰减设为 5e-4, 动量为 0.937, 初始学习率为 1e-2。使用的 GPU 为 RTX 4090。

3.3 模型评价指标

3.3.1 参数 (Params) 模型的参数取决于模型的层数、结构的复杂度、每层的神经元数量等因素。参数越大, 该模型就越大。不同的数据集对模型参数量的要求不同, 更多的参数会增加计算机计算时间。在具体应用中要平衡好模型大小和计算成本。

3.3.2 浮点运算次数 (FLOPs) FLOPs 是指模型中的浮点运算操作总数, 包括加减乘除等操作。它表示模型在训练或推理中实际进行的浮点运算次数。该值越小代表模型所进行的运算次数越少, 能在一定程度上反应模型的计算效率。

3.3.3 全类平均精度 (mAP) mAP 是衡量目标检测模型性能最重要的指标之一, 本文使用的指标是置信度为 0.5 时的平均精度值 (mAP50)。置信度代表了预测样本框和真实样本框之间的重叠程度, 若置信度为 0.5, 表示只有预测样本框与真实样本框的重叠程度达到了 50% 以上, 该预测样本才被认为是被正确识别出的样本。单个标签类别的平均精度 (AP) 结合了精确率和召回率。精确率代表了正样本预测的精确程度, 公式如下:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

(9)

其中, TP 代表正样本且被判定为正样本, FP 为负样

本但被判定为正样本。召回率指的是被正确识别出的所有实际样本中所占有正样本的比例, 公式如下:

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}}$$

(10)

其中, FN 代表正样本但被判定为负样本。计算精确率、召回率后以精确率为 y 轴, 召回率为 x 轴构建精确率-召回率曲线 (P-R 曲线), AP 的值为 P-R 曲线下的面积, 通常利用积分计算, 公式如下:

$$\text{AP} = \int_0^1 f(x) dx$$

(11)

其中, $f(x)$ 为 P-R 曲线的表达式, mAP 就是所有类别 AP 的平均值, 公式如下:

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^n \text{AP}_i$$

(12)

3.4 3 种 YOLOv8 模型的对比

为了找出更适合本文改进方法的 YOLOv8 模型, 分别将融入了 LSKA 以及 ResLSKA 机制的 YOLOv8s、YOLOv8m、YOLOv8l 模型在相同的数据集以及环境下进行训练, 并记录训练后在测试集上的表现。从表 1 得知无论是 LSKA 机制还是 ResLSKA 机制, 改进后的模型都以 YOLOv8m 模型为基础模型时训练效果最好。据此, 使用 YOLOv8m 模型为改进模型的基础模型。

3.5 消融实验

为了验证本文改进策略的有效性, 以基础的

表1 不同大小的YOLOv8模型对比
Table 1 Comparison of YOLOv8 models of different sizes

模型	模型大小	mAP50	Params/M	FLOPs/B
YOLOv8+LSKA	s	0.625	12.4	31.3
YOLOv8+LSKA	m	0.653	37.9	84.3
YOLOv8+LSKA	l	0.647	46.2	173.8
YOLOv8+ResLSKA	s	0.630	16.1	38.9
YOLOv8+ResLSKA	m	0.660	33.9	99.5
YOLOv8+ResLSKA	l	0.652	53.9	198.4

YOLOv8m 模型的训练结果为基准,逐步在 YOLOv8m 模型上添加改进策略,观察模型性能的变化。消融实验中实验设置与前文说明保持一致。选择的优化器为SGD,训练轮次为80次,图片输入大小为640 px,使用的GPU为RTX 4090,bath-size 设为32,其他超参数选择 YOLOv8 模型开发者的默认设置。

从表2可以看出,LSKA 机制的加入(YOLOv8m+LSKA)使 YOLOv8m 模型的 mAP50 有较大提升,体现了 LSKA 提取重要信息,抑制冗余信息的作用;将 LSKA 与 ResBlock 融合之后(YOLOv8m+ResLSKA),进一步提升了改进模型的性能,证明了 ResBlock 有

效提升了 LSKA 机制的作用;最后 C2f_SAC 的替换(YOLOv8m+ResLSKA+C2f_SAC)使得模型能更好地捕捉并处理输入的特征信息。

表2 消融实验
Table 2 Ablation study

模型	mAP50	Params/M	FLOPs/B
YOLOv8m	0.626	25.8	79.1
YOLOv8m+LSKA	0.653	27.9	84.3
YOLOv8m+ResLSKA	0.660	33.9	99.5
YOLOv8m+ResLSKA+C2f_SAC	0.672	37.6	94.9

3.6 与其他改进模型的对比

为了验证本文提出的改进模型对腕部X光片骨折检测的有效性,将本文改进模型(YOLOv8m+ResLSKA+C2f_SAC)与其他 YOLOv8 的改进模型以及基于 YOLOv9 模型在相同数据集上训练后在测试集上的测试结果进行比较。由于基于 YOLOv9 的改进模型对显卡内存需求较大,需要将 batch-size 设为16以适应显卡内存限制,其他模型训练的 batch-size 为32。实验结果如表3所示。

表3 不同模型的实验结果对比
Table 3 Comparison of experimental results of different models

模型	mAP50	Params/M	FLOPs/B	训练时间/h
YOLOv9-C+DualDDetect	0.655	51.02	239.0	16.81
YOLOv9-E+DualDDetect	0.656	69.42	244.9	20.08
YOLOv8l+ResCBAM	0.658	53.87	196.2	9.26
YOLOv8l+Global Context	0.663	43.85	165.6	8.34
YOLOv8m+ResLSKA+C2f_SAC	0.672	37.58	94.9	6.57

从表3可知,在相同的数据集上,本文模型的 mAP50 指标取得了最好的结果,同时相较于以 YOLOv8l 以及 YOLOv9 为基础的改进模型,本文模型减少了参数量以及 FLOPs,提高了改进模型在数据集上的训练速度。

为了验证本文模型能够进行实时检测,将本文模型与其他改进模型在不同型号 GPU 上的推理速度进行比较,结果如表4所示。

表4 不同模型在不同GPU上的推理速度(FPS)
Table 4 Inference speed (FPS) of different models on different GPU

模型	RTX-3090	RTX-4090	A-100
YOLOv9-C+DualDDetect	101.83	169.21	144.92
YOLOv9-E+DualDDetect	81.35	132.97	131.58
YOLOv8l+ResCBAM	149.49	222.73	208.46
YOLOv8l+Global Context	165.57	248.71	218.22
YOLOv8m+ResLSKA+C2f_SAC	193.18	267.16	225.32

将训练后的模型分别在 RTX-3090、RTX-4090 以及 A-100 3 种 GPU 上进行推理速度的比较。实验结果表明,不同模型在3种显卡中的推理速度有一定差距,但所有模型都满足实时检测的需求(>30 FPS)。本文模型训练后在测试集上的表现以及检测实例如图6~图10所示。

从检测效果标签图可以看出,骨折、文本、金属以及骨膜反应这4类标签的检测都有较高的准确率,大部分的目标都能正确检测出来,但也存在一定的误检。比如在骨折类的检测中可能会出现将一个骨折目标分成两个骨折目标来检测,但检测出的骨折目标的范围相比原标签并没有太大变化;再比如金属类中可能会存在较为明显的线性图形,模型可能将这种线性图形误测为文本类。同时由于腕部X线片结构较为复杂多样,也会出现小概率的漏检以及误检,这说明了本文模型对特征信息的提取能力以及不重要信息的抑制能力还有进一步的提高空间。

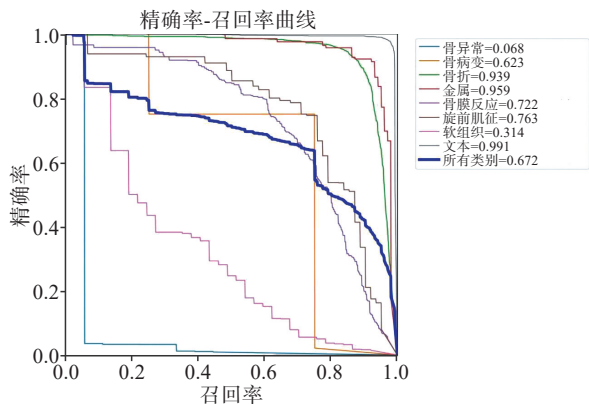


图6 改进模型在测试集上的结果(mAP50)

Figure 6 Results (mAP50) of the improved model on the test set

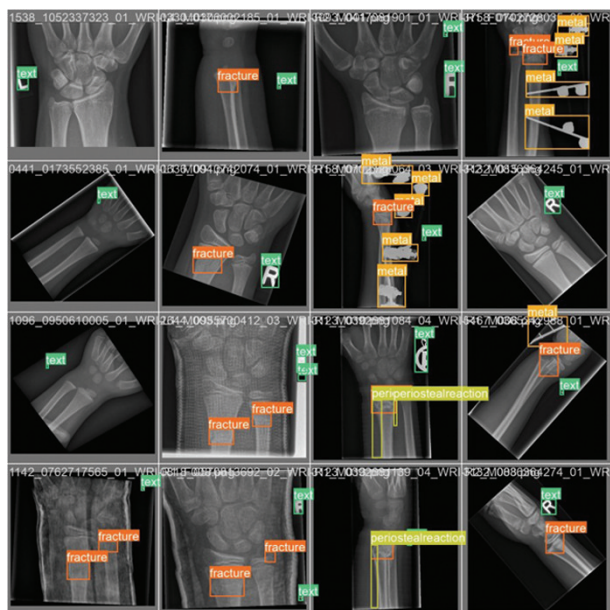


图7 测试集上的真实标签1

Figure 7 True label 1 on the test set

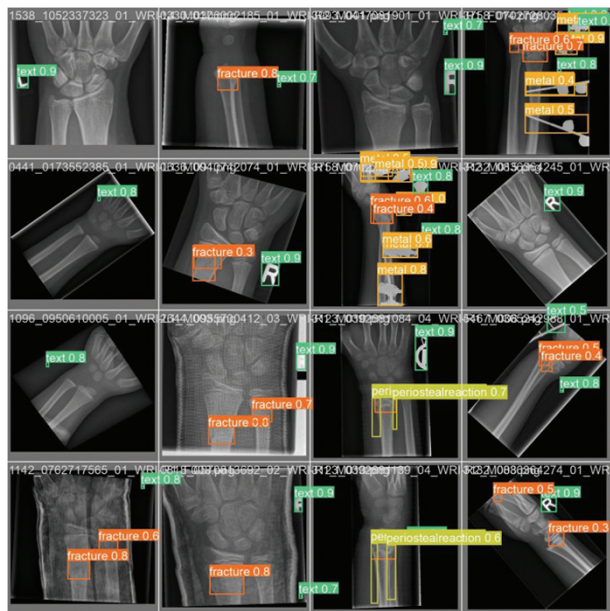


图8 本文模型的检测效果(对应真实标签1)

Figure 8 Detection results of the proposed model for true label 1

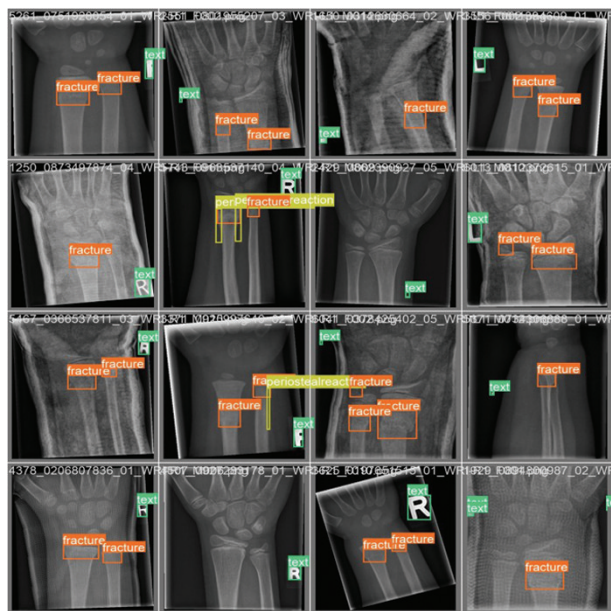


图9 测试集上的真实标签2

Figure 9 True label 2 on the test set

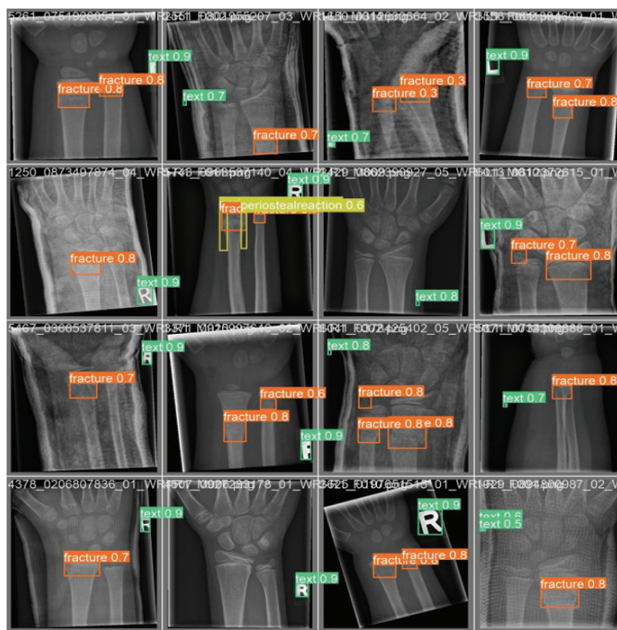


图10 本文模型的检测效果(对应真实标签2)

Figure 10 Detection results of the proposed model for true label 2

4 结论

本文在网络YOLOv8m的基础上进行改进使其能够更好地进行腕部X光片的骨折检测。首先引入了全新的LSKA,提取重要信息,抑制无用信息;将残差模块与该注意力机制融合,进一步提高注意力机制的作用,同时提高模型的泛化能力;将SAC卷积块与C2f模块融合,提高了捕捉输入特征的能力。实验结果表明文本改进模型与YOLOv8l改进模型相比,mAP50提高了1.3%,同时得益于使用参数量更少的

YOLOv8m模型,本文模型的参数量较先进的改进模型减少了14.3%,FLOPs减少了42.7%,减少了模型的训练时间。

同时该模型的检测效果还有进一步的提升空间。从数据集角度来看,该数据集有些目标类的数量稀少,很大程度上拉低了mAP指标,并且很难将这些数量稀少的类平均地分配给训练集、验证集、测试集,在训练过程中,数量较少的类别可能会提前过拟合导致整体精度开始下降。如果引入更多这些类的图片与标签,会一定程度上缓解此问题,同时提高mAP。从模型本身来看,YOLOv8骨干网络的SPPF模块或者Conv模块也有很大的改进空间,后期将引入更好的卷积块来替换SPPF中的Conv模块,或者直接替换骨干网络的Conv模块来提升模型性能。

【参考文献】

- [1] 汤倩媛. 腕关节骨折的X线片、CT及重建技术的诊断价值分析[J]. 影像研究与医学应用, 2022, 6(8): 137-139.
Tang QY. Diagnostic value of X-ray, CT and reconstruction techniques for wrist fracture[J]. Journal of Imaging Research and Medical Applications, 2022, 6(8): 137-139.
- [2] Bamford R, Walker DM. A qualitative investigation into the rehabilitation experience of patients following wrist fracture[J]. Hand Ther, 2010, 15(3): 54-61.
- [3] 叶玲玲, 张念, 王大川. 腕关节外伤X线平片、CT和MRI比较分析[J]. 温州医学院学报, 2008, 38(6): 566-568.
Ye LL, Zhang N, Wang DC. Comparative analysis of X-ray plain film, CT and MRI for wrist joint trauma[J]. Journal of Wenzhou Medical College, 2008, 38(6): 566-568.
- [4] 祁艳梅, 唐治, 周晟. 腕关节损伤的影像学检查研究进展[J]. 西北国防医学杂志, 2020, 41(9): 589-594.
Qi YM, Tang Z, Zhou S. Research progress of imaging examination in wrist injury[J]. Medical Journal of National Defending Forces in Northwest China, 2020, 41(9): 589-594.
- [5] 杨锋, 陈雷, 邢蒙蒙. 基于Faster R-CNN的图像检索手腕骨折分类研究[J]. 中国医疗设备, 2023, 38(2): 1-6.
Yang F, Chen L, Xing MM. Research on classification of wrist fracture based on faster R-CNN image retrieval[J]. China Medical Devices, 2023, 38(2): 1-6.
- [6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 779-788.
- [7] Dibo R, Galichin A, Astashev P, et al. DeepLOC: deep learning-based bone pathology localization and classification in wrist X-ray images[C]//Analysis of Images, Social Networks and Texts. Cham: Springer Nature Switzerland, 2024: 199-211.
- [8] Wang CY, Bochkovskiy A, Liao HY. YOLOv7: trainable bag-of-freebies Sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 7464-7475.
- [9] Chien CT, Ju RY, Chou KY, et al. YOLOv8-AM: YOLOv8 based on effective attention mechanisms for pediatric wrist fracture detection[EB/OL]. (2024-09-28). <https://arxiv.org/abs/2402.09329>.
- [10] Chien CT, Ju RY, Chou KY, et al. YOLOv9 for fracture detection in pediatric wrist trauma X-ray images[J]. Electron Lett, 2024, 60(11): e13248.
- [11] Ultralytics. Ultralytics YOLO11[EB/OL]. <https://github.com/ultralytics/ultralytics>.
- [12] Zhang H, Cisse M, Dauphin YN, et al. Mixup: beyond empirical risk minimization[C]//The 6th International Conference on Learning Representations, 2018.
- [13] Wang CY, Liao HY, Wu YH, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2020: 1571-1580.
- [14] Wang CY, Liao HY, Yeh IH. Designing network design strategies through gradient path analysis[J]. J Inf Sci Eng, 2023, 39(4): 975-995.
- [15] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 936-944.
- [16] Liu S, Qi L, Qin HF, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 8759-8768.
- [17] Li X, Wang WH, Wu LJ, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection [C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2020: 21002-21012.
- [18] Zheng ZH, Wang P, Ren DW, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Trans Cybern, 2022, 52(8): 8574-8586.
- [19] Lau KW, Po LM, Rehman YA. Large separable kernel attention: rethinking the large kernel attention design in CNN[J]. Expert Syst Appl, 2024, 236: 121352.
- [20] Guo MH, Lu CZ, Liu ZN, et al. Visual attention network[J]. Comput Vis Media, 2023, 9(4): 733-752.
- [21] Wang PQ, Chen PF, Yuan Y, et al. Understanding convolution for semantic segmentation[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2018: 1451-1460.
- [22] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [23] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [24] Qiao SY, Chen LC, Yuille A. DetectoRS: detecting objects with recursive feature pyramid and switchable atrous convolution[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021: 10208-10219.
- [25] Cao Y, Xu JR, Lin S, et al. GCNet: non-local networks meet squeeze-excitation networks and beyond[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Piscataway, NJ, USA: IEEE, 2019: 1971-1980.
- [26] Nagy E, Janisch M, Hrzić F, et al. A pediatric wrist trauma X-ray dataset (GRAZPEDWRI-DX) for machine learning[J]. Sci Data, 2022, 9(1): 222.
- [27] Ju RY, Cai WM. Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm[J]. Sci Rep, 2023, 13(1): 20077.

(编辑:薛泽玲)