

# 人工智能单阶段目标检测模型诊断 X 线平片骨折线的效能研究

王亚丽, 乔仔依, 吴浩戈, 陈诗飞, 张冉旭, 吴慧钊\*

(河北医科大学第三医院医学影像科, 河北 石家庄 050051)

**[摘要]** 目的 比较单阶段目标检测法 YOLO-v5、YOLO-v7、YOLO-v8 3 种模型对 X 线平片不同部位骨折线的诊断性能, 选择对细小骨折线、复杂骨折线、多发骨折线精准检测的最佳模型。方法 纳入 6 740 幅骨折数字化 X 线平片影像, 使用 pycharm 集成开发环境配置 YOLO-v5、YOLO-v7、YOLO-v8 目标检测模型, 以端对端的方式输入影像并可视化诊断结果。使用准确率、精准率、召回率、F1 值、平均精度、ROC 曲线下面积等评价指标比较不同模型诊断骨折线的性能差异。由两名放射科医师使用/不使用 3 种诊断模型辅助, 即 8 种诊断方式对测试集进行诊断, 比较诊断时间、准确率、敏感度、特异度等差异, 分析假阳性、假阴性结果的原因。结果 3 种模型中 YOLO-v5 在综合部位组中的骨折诊断准确率最高(78.11%), 四肢长骨骨折组的准确率为 82.47%, 脊柱骨折组的准确率为 83.34%, 但对骨盆和手足部位骨折的检测效能较差, 准确率分别为 60.15%、65.12%。YOLO-v7 模型对四肢长骨、骨盆、手足骨折检测的准确率最低(30.35%、30.11%、47.68%), 但脊柱骨折诊断准确率最高(88.92%)。YOLO-v8 5 组骨折检测准确率分别为 76.20%、81.69%、82.36%、60.52%、55.65%, 性能介于其他 2 个模型之间。对比有无模型辅助诊断, 发现 YOLO-v5、YOLO-v8 模型均能提升诊断医师的工作速度及诊断准确性。结论 3 种模型辅助诊断效果有差异, 选择适宜模型进行辅助诊断可提升医师的诊断效率和准确性。

**[关键词]** 骨折; 诊断; 人工智能 doi:10.3969/j.issn.1007-3205.2025.11.012

**[中图分类号]** R683 **[文献标志码]** A **[文章编号]** 1007-3205(2025)11-1331-09

## Diagnostic performance of artificial intelligence-based single-stage object detection models in fracture lines on X-ray radiographs

WANG Ya-li, QIAO Zi-yi, Wu Hao-ge, CHEN Shi-fei, ZHANG Ran-xu, WU Hui-zhao\*

(Department of Medical Imaging, the Third Hospital of Hebei Medical University, Shijiazhuang 050051, China)

**[Abstract]** **Objective** To analyze the diagnostic performance of single-stage object detection models YOLO-v5, YOLO-v7 and YOLO-v8 in fracture lines across different anatomical regions on X-ray radiographs, and to determine the optimal model for precise detection of subtle, complex, and multiple fracture lines. **Methods** A total of 6 740 digital X-ray radiographs of fractures were included. The YOLO-v5, YOLO-v7 and YOLO-v8 models were configured in the PyCharm integrated development environment to enable end-to-end image input and visualize diagnostic results. Evaluation metrics, including accuracy, precision, recall, F1-score, average precision, and area under the ROC curve, were used to compare the models in

[收稿日期] 2025-07-07

[基金项目] 河北省卫生健康委员会青年科技课题(20230087);  
河北医科大学大学生创新性实验计划项目(USIP2025122)

[作者简介] 王亚丽(1988-), 女, 河北石家庄人, 河北医科大学第三医院主管技师, 医学学士, 从事医学影像技术及人工智能医学应用研究。

\* 通信作者。E-mail: wuhuizhao@hebm.u.edu.cn

diagnosing fracture lines. Two radiologists evaluated the test set using eight diagnostic approaches (with/without each of the three models), comparing diagnostic time, accuracy, sensitivity, specificity, and analyzing causes of false-positive results and false-negative results.

**Results** Among the three models, YOLO-v5 achieved the highest overall diagnostic accuracy (78.11%) for comprehensive anatomical regions. YOLO-v5 demonstrated accuracies of 82.47% for fractures in the long bones of extremities, 83.34% for spinal fractures, but lower performance for pelvic and hand/foot fractures (60.15% and 65.12%, respectively). YOLO-v7 showed the lowest accuracy for long bones of extremities, pelvis, and hand/foot fractures (30.35%, 30.11%, and 47.68%) but achieved the highest accuracy for spinal fractures (88.92%). The fracture detection accuracies of YOLO-v8 for the five groups were 76.20%, 81.69%, 82.36%, 60.52%, and 55.65%, respectively, with its performance lying between the other two models. Compared with the diagnoses with or without model assistance, both YOLO-v5 and YOLO-v8 improved radiologists' workflow speed and diagnostic accuracy. **Conclusion** The three models exhibited varying effectiveness in assisting diagnosis. Selecting appropriate models for specific anatomical regions can significantly enhance diagnostic efficiency and accuracy of radiologists.

[Key words] fracture; diagnosis; artificial intelligence

X线平片是诊断骨折的常规影像学检查手段,可简便快捷地提供高空间分辨率图像,但受解剖结构重叠、细微骨折线对比度较低不易辨认等影响,观察者的一致性较差<sup>[1-2]</sup>。放射科医师的诊断效能受工作负荷、临床经验差异影响,能够在初始X线检查中准确诊断骨折具有重要临床意义<sup>[3]</sup>。人工智能(artificial intelligence, AI)深度学习方法可识别人眼不能辨认的图像特征<sup>[4-5]</sup>,其辅助诊断可为影像科医生提供决策支持,减少图像解释错误,避免工作疲劳产生的误诊、漏诊<sup>[6-8]</sup>。人工智能的计算机视觉算法中,单阶段目标检测YOLO算法快速迭代,综合了检测速度快和精度平衡的优点,在自然图像目标检测领域实现了较快的实时处理速度和较高的检测精度<sup>[9]</sup>。在医学图像分析领域,YOLO系列算法在病灶目标检测(诊断)中显示了在检测速度和检测精度方面的潜力<sup>[10-11]</sup>。Liu等<sup>[12]</sup>在YOLO模型诊断CT图像肋骨骨折的研究结果显示,YOLO模型在较高的诊断准确率、敏感度、特异度下,诊断速度明显优于放射科医师。因此本研究假设YOLO目标检测模型可快速有效地对数字化X线平片中的骨折线进行检测和定位,研究从检测性能、临床适用性出发选择YOLO系列算法中的v5、v7、v8对骨折线进行检测,分析比较各个算法检测性能的差异,筛选最佳X线平片骨折线单阶段法目标检测模型。并将最佳模型的诊断效能与放射科医师进行比较,以验证其临床应用价值。

## 1 资料与方法

1.1 一般资料 选取河北医科大学第三医院2020年6月—2024年3月3780例骨折患者的7805幅数字化X线平片,其中3622例患者的6740幅骨折X线平片被纳入研究。为模拟急诊临床场景,病例纳入标准为外伤性骨折患者(急性、亚急性)。排除标准为炎性骨病、骨骺不闭合、风湿、肿瘤、代谢性疾病、病理性骨折、病史资料不全等。3622例患者中男性2050例,女性1572例,年龄17~78岁,平均(42.2±27.4)岁。所有患者均进行6周~1年的临床随访,手术患者进行了CT检查(占比34.15%)。X线平片中骨折部位分布见表1。

本回顾性研究获医院伦理审查委员会批准(K2024-143-1)。

表1 X线平片骨折部位分布

Table 1 Distribution of fracture sites on X-ray radiographs

骨折部位	独立病例数 (例数, %)	影像数量 (幅, %)	每病例平均 影像数(幅)
四肢长骨	2400(66.26)	4580(67.95)	1.63
脊柱	497(13.72)	520(7.72)	1.00
手足	513(14.16)	1052(15.61)	1.72
骨盆	212(5.85)	588(8.72)	1.88
总计	3622(100.00)	6740(100.00)	1.86

1.2 图像标注 由3名具有10年经验以上的放射科医师使用Labelme目标检测框标注工具骨折线进行标注,所有图像均由高年资医师进行标记审查,以该病例临床随访结果结合所有影像学检查为综合诊断标准。

**1.3 数据集划分及预处理** 所有图像均进行了脱敏处理,脱敏数据集按 7:3 的比例划分为内部数据集和外部数据集,内部数据集用于模型的训练,训练模型参数。使用梯度下降进行参数更新,不断迭代达到该模型的最佳检测性能。外部数据集则以各项评价指标验证最佳模型的检测效能。通过遍历每幅图像,给出目标检测框被分类为“骨折”可能性的置信度分数。网络训练前对 X 线影像进行剪切、噪声扰动等图像增强处理,使用多种组合数据增强变换的方式来产生相似但又不同的训练样本,增加网络鲁棒性。

**1.4 目标检测网络环境配置及训练策略** 数据标注和训练平台使用 CPU 为 Intel Core i9-10900K, GPU 为 NVIDIA GeForce GTX 3080,于 Windows 10 系统上运行基于 Python 的源深度学习框架 PyTorch(3.8.0 版本, <https://pytorch.org/>) 和 CUDA(版本 11.8, Nvidia)。3 种 YOLO 模型下载官网链接: <https://github.com/ultralytics/ultralytics/tree/main>。学习率按步长衰减,初始学习率设置为 0.001,每隔 30 的 epoch 数将学习率乘以 0.1(具体衰减点在 epoch30、epoch60、epoch90),权重衰减系数为 0.0001。

**1.5 骨折线诊断模型性能及人机结合诊断性能的评价** 使用精准率、准确率、敏感度(即召回率)、F1 值、平均精度(average precision, AP)等比较 3 种模型的骨折线检测性能。由 2 名放射科医师使用及不使用 3 种诊断模型辅助,即 8 名“诊断者”对测试集 300 幅(真阳性、真阴性影像为 2:1)进行诊断,比较人、人机结合的诊断时间、准确率、敏感度、特异度等差异,分析不同“诊断者”出现假阳性、假阴性原因。

**1.6 统计学方法** 应用 SPSS 22.0 统计软件分析数据。以 Fleiss' Kappa 系数评价 3 名医师的标注一致性,两两间的标注一致性使用 Cohen's Kappa 系数评价。设置 5 组各 200 例图像(真阳性 150 例、真阴性 50 例)以受试者工作特征曲线(receiver operating characteristic curve, ROC)下面积(area under curve, AUC)评价模型的临床诊断性能,并以

DeLong 检验对 AUC 值进行两两比较。 $P < 0.05$  为差异有统计学意义。

## 2 结 果

**2.1 数据标注的一致性评价** 3 名标注医师内和 3 名标注医师间对数据标注的一致性评价,结果显示:骨盆部位医师之间标注的一致性较好( $K > 0.6$ ),其余部位均获得优等的评价一致性( $K > 0.8$ ),见表 2。

**2.2 3 种目标检测模型诊断骨折线效能的比较** 通过 100 个 epoch 的迭代训练实现了模型收敛,接下来利用 YOLO-v5、YOLO-v7、YOLO-v8 模型分别测试五组内部数据集检测性能,结果显示:YOLO-v5 对综合部位组骨折的诊断准确率达到 78.11%、骨折线召回率为 80.96%;四肢长骨骨折的诊断准确率达到 82.47%、召回率为 88.16%;脊柱骨折的诊断准确率达到 83.34%、召回率为 86.12%;对手足和骨盆部位骨折的诊断效能相对较差,但仍然是 3 种模型中性能最优,准确率分别为 65.12%、60.15%,召回率分别为 71.24%、65.24%。综合部位组 3 个模型中诊断性能最差的是 YOLO-v7 模型,但其对脊柱部位的检测最好,准确率达到 88.92%,见表 3。

**2.3 3 种目标检测模型诊断骨折 AUC 的比较** 3 种目标检测模型各自诊断不同部位骨折的 ROC 曲线见图 1,对同一部位骨折诊断的 ROC 曲线见图 2,诊断各部位骨折 AUC 值的比较见表 4。结果显示:YOLO-v5 在综合部位、四肢长骨、脊柱、骨盆、手足 5 组中均取得高值(AUC 分别为 0.897、0.944、0.951、0.875、0.884, 95% CI: 0.838~0.956、0.905~0.983、0.924~0.992、0.819~0.947、0.804~0.946)说明 YOLO-v5 为最佳模型。YOLO-v7 在脊柱组中取得了最高 AUC 值 0.960(95% CI: 0.924~0.996),其余均低于其他 2 种目标检测模型,说明 YOLO-v7 适合脊柱骨折骨折检测。YOLO-v8 的 AUC 值则介于 2 种模型之间,诊断效能逊于 YOLO-v5。

表 2 3 名标注医师数据标注的一致性分析

Table 2 Consistency analysis of data annotation among three radiologists

部位	医师 1 和医师 2 Cohen's K 值	医师 1 和医师 3 Cohen's K 值	医师 2 和医师 3 Cohen's K 值	整体 Fleiss' K 值
四肢长骨	0.963	0.952	0.974	0.952
脊柱	0.902	0.893	0.912	0.913
手足	0.851	0.862	0.893	0.871
骨盆	0.752	0.742	0.764	0.752
综合部位	0.863	0.881	0.832	0.864

表3 3种目标检测模型对内部数据集的检测性能

Table 3 Detection performance of three object detection models on internal datasets

模型	部位	准确率(%)	精准率(%)	召回率(%)	F1值(%)	AP(%)
YOLO-v5	综合部位	78.11	79.11	80.96	81.12	79.63
	四肢长骨	82.47	85.12	88.16	86.91	86.16
	脊柱	83.34	84.06	86.12	85.32	84.41
	手足	65.12	66.71	71.24	65.15	66.18
	骨盆	60.15	67.46	65.24	69.11	66.81
YOLO-v7	综合部位	45.26	45.36	46.25	46.12	44.26
	四肢长骨	30.35	30.41	31.12	31.09	30.25
	脊柱	88.92	89.15	90.35	90.63	88.63
	手足	47.68	48.41	49.01	49.42	47.32
	骨盆	30.11	30.25	30.62	30.01	30.12
YOLO-v8	综合部位	76.20	75.31	80.60	80.10	78.62
	四肢长骨	81.69	83.76	87.36	85.96	84.96
	脊柱	82.36	83.56	83.62	83.11	82.61
	手足	60.52	63.86	65.52	62.35	65.38
	骨盆	55.65	65.68	70.25	62.20	64.89

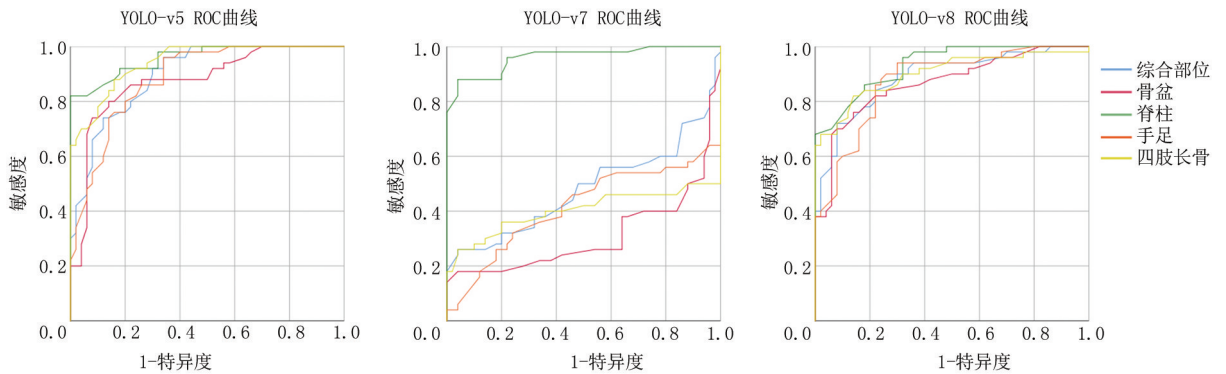


图1 3种目标检测模型各自诊断不同部位骨折的ROC曲线

Figure 1 ROC curves of the three object detection models for diagnosing fractures in different anatomical regions

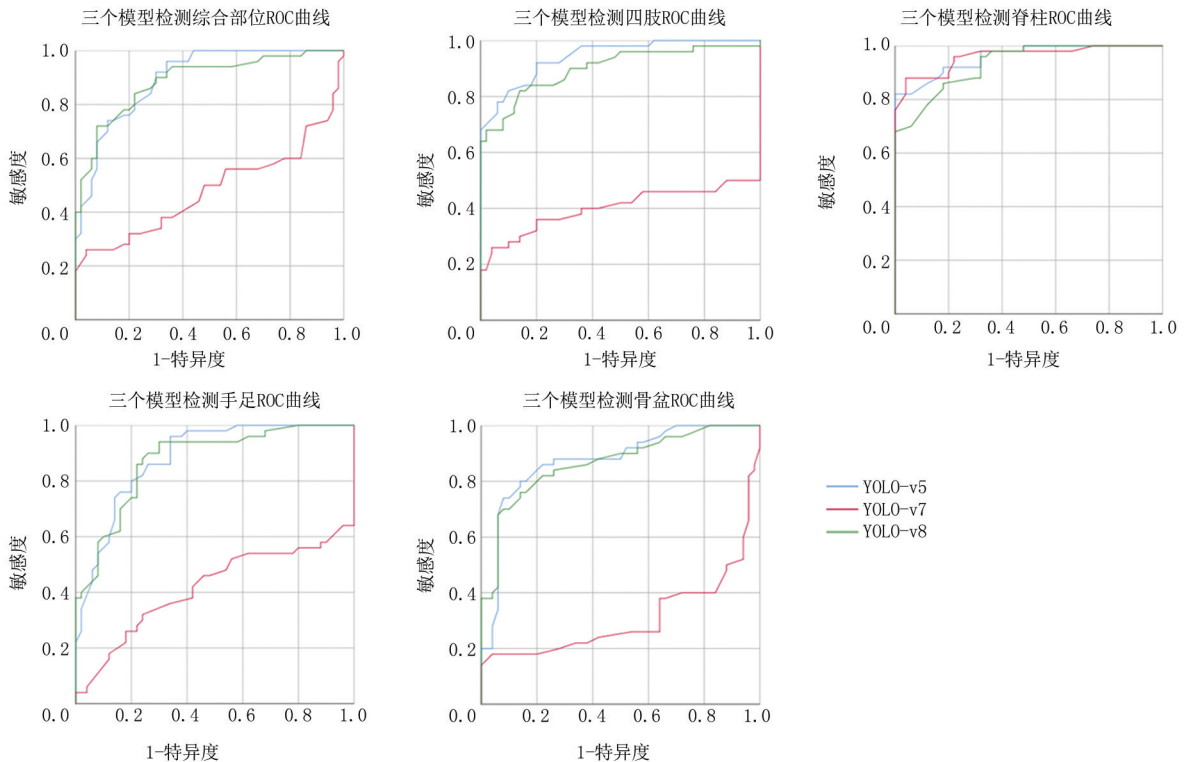


图2 3种目标检测模型诊断同一部位骨折的ROC曲线

**Figure 2 ROC curves of the three object detection models for diagnosing fractures in the same anatomical region****表4 3种目标检测模型诊断不同部位骨折的ROC曲线下面积比较****Table 4 Comparison of areas under the ROC curves among three object detection models for fracture diagnosis across different anatomical regions**

模型	综合部位	四肢长骨	脊柱	骨盆	手足
YOLO-v5	0.897	0.944	0.951	0.875	0.884
YOLO-v7	0.480	0.402	0.960	0.311	0.415
YOLO-v8	0.888	0.901	0.933	0.866	0.874

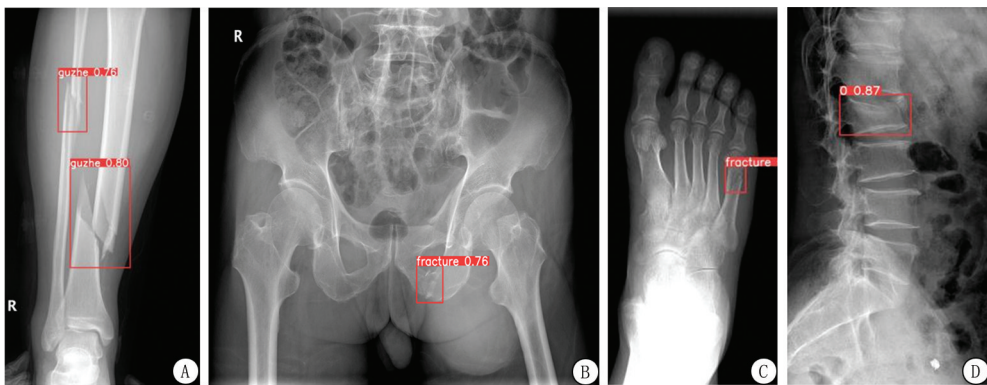
## 2.4 使用及不使用3种模型辅助诊断效能的比较

2名放射科医师使用及不使用3种诊断模型辅助(人及人机结合),共8名“诊断者”对测试集病例进行诊断结果的比较见表5,使用优秀模型辅助诊断提高了医生的诊断速度、准确率及敏感度等。

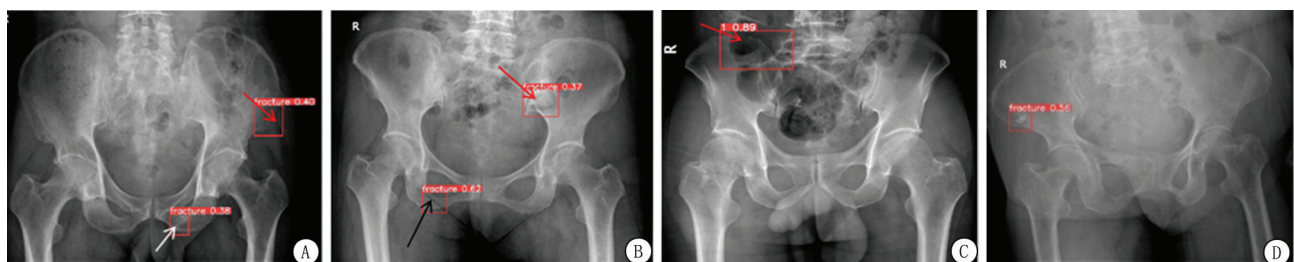
## 2.5 3种模型在测试集中检出骨折线的性能表现

**表5 人及人机结合诊断效能的比较****Table 5 Comparison of diagnostic performance between human readers and human-AI collaboration**

指标	时间(min)	准确率(%)	敏感度(%)	特异度(%)	精准率(%)	F1值(%)
医生1	25	93.61	97.92	86.48	92.52	95.14
医生2	23	91.29	95.81	83.62	91.03	93.29
医生1+YOLO-v5	20	95.72	98.41	90.68	95.04	96.66
医生1+YOLO-v7	26	93.01	96.85	86.21	92.47	94.56
医生1+YOLO-v8	19	95.03	97.37	87.24	94.02	96.22
医生2+YOLO-v5	18	94.68	97.91	88.78	94.01	95.87
医生2+YOLO-v7	26	91.03	95.32	83.38	91.02	93.14
医生2+YOLO-v8	19	94.04	97.36	88.13	93.46	95.37

**图3 3个模型均可正确诊断易分骨折**

A.右胫腓骨X线;B.骨盆X线;C.右足X线;D.脊柱X线

**Figure 3 All three models for correctly diagnosing the readily identifiable fractures**

## 4 常见假阳性示例

A.骨盆X线(红箭头为软组织间隙,白箭头为骨质增生);B.骨盆X线(红箭头为肠气,黑箭头为皮肤褶皱);C.骨盆X线(红箭

及假阳性、假阴性分析 对于易分正样本(容易正确分类的正样本,本研究为错位明显、容易辨认骨折线的病例),3个模型均可正确诊断并给出较高的置信度(图3)。假阳性见于腹部软组织间隙阴影(图4A)、肠气遮挡(图4B、C),以及皮肤褶皱(图4B)、骨质增生(图4A)、骨化(图4D)。足部多发骨折病例诊断中,YOLO-v5成功检出并给出较高的置信度,YOLO-v7显示假阴性结果,YOLO-v8检出但给出较低的置信度分数(图5)。对于细小骨折和多发骨折病例,YOLO-v5表现出较强的诊断准确性,YOLO-v7发生漏诊,YOLOv8则检测不全,未能检出全部骨折线(图6)。脊柱正位图像中YOLO-v5和YOLO-v8未能诊断腰椎骨折,YOLO-v7却能成功召回(图7)。

头为肠气);D.骨盆 X 线(骨化造成骨盆骨折假阳性)

Figure 4 Representative examples of common false positives

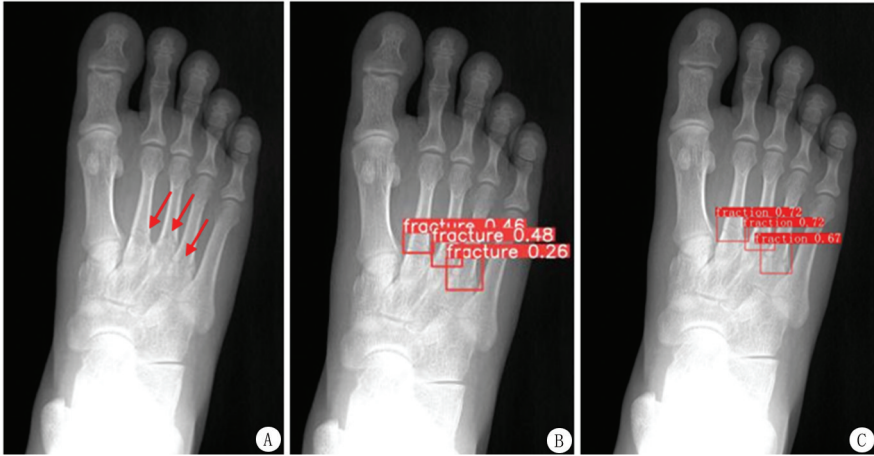


图 5 3 个模型在足部多发骨折病例诊断中的置信度

A.YOLO-v7 未检出为假阴性(红箭);B.YOLO-v8 检出但置信度分数较低;C.YOLO-v5 成功检出并给出较高的置信度分数

Figure 5 The confidence of the three models in the diagnosis of multiple fractures of the foot

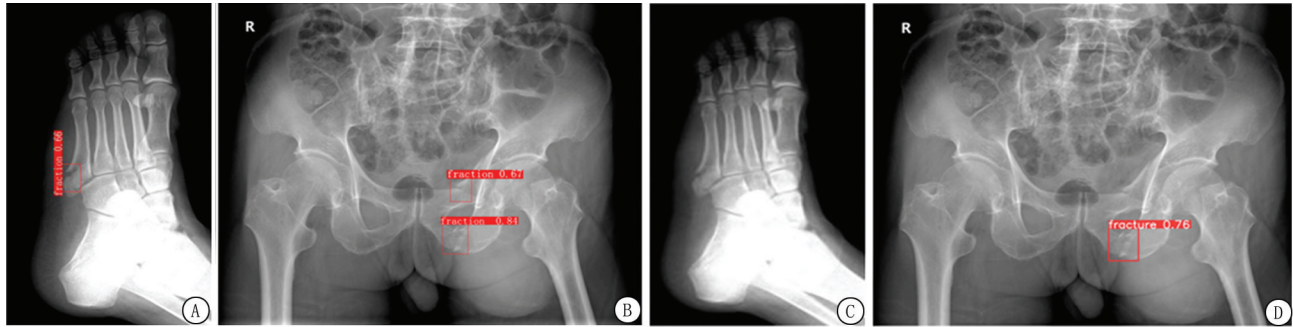


图 6 3 个模型在细小骨折和多发骨折病例中诊断的准确性

A.左足 X 线(YOLO-v5 对细小骨折的准确诊断);B.骨盆 X 线(YOLO-v5 对多发骨折线的准确诊断);C.左足 X 线(YOLO-v7 假阴性诊断);D.骨盆 X 线(YOLO-v8 对骨盆骨折线漏诊)

Figure 6 The diagnostic accuracy of the three models in cases of minor fractures and multiple fractures

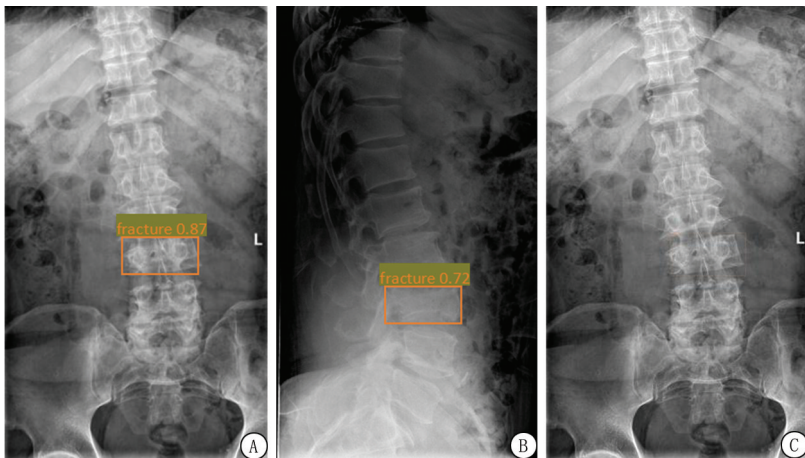


图 7 3 个模型在脊柱骨折诊断中的准确性

A.脊柱 X 线(YOLO-v7 正确诊断脊柱正位腰椎体骨折);B.脊柱 X 线(侧位显示患椎压缩骨折);C.脊柱 X 线(YOLO-v5 和 YOLO-v8 则未能检出)

Figure 7 The accuracy of the three models in the diagnosis of spinal fractures

### 3 讨 论

当前深度学习目标检测算法有两阶段法、单阶

段法、Transformer 3 类,不同算法在精度、速度和鲁棒性等方面各有优势<sup>[13-17]</sup>。现有目标检测模型基于 3 通道的自然图像检测任务开发,应用于单通道

数字 X 线图像需 3 次遍历影像才能完成训练,具有较高的时间成本和运算开销,在临床场景中部署受到限制<sup>[18]</sup>。本课题组前期研究根据骨折线的形态及空间关系特点开发特征歧义减轻算子作用于两阶段目标检测法的第一、二阶段衔接处,对骨折线特征图进行裁剪,减轻骨折线的特征歧义,提高了卷积神经网络的诊断效能,敏感度、特异度和 AUC 达到 77.5%、93.4% 和 86.5%<sup>[19]</sup>。但研究中发现其检测速度较慢,敏感度低于特异度,需开发更佳目标检测算法。本研究基于数字化 X 线平片的单通道数据特点,发挥单阶段目标检测算法的优势应用于骨折线人工智能诊断,设计从头训练以保证网络设计的灵活性。相较于 Lindsey 等<sup>[20]</sup>使用 135 409 幅平片的单部位训练数据集获得 80.8% 的敏感度及 87.5% 的特异度,本研究的多部位训练数据集样本含量控制在 10 000 以内,人机结合获得最高 98.4% 的敏感度和 90.7% 的特异度,以较小数据集取得了同类研究中较好的检测效果<sup>[21-23]</sup>。筛选出优势模型 YOLO-v5 可快速检出门诊、急诊场景下的骨折影像,具有较高的诊断效能,验证了实验假设。

研究中比较 YOLO-v5、YOLO-v7 和 YOLO-v8 3 种单阶段目标检测模型在 X 线平片骨折诊断中的差异发现, YOLO-v5 在四肢长骨(准确率 82.47%, 召回率 88.16%)、骨盆复杂骨折(准确率 60.15%, 召回率 65.24%)及手足细小骨折(准确率 65.15%, 召回率 71.24%)中综合表现最优。而 YOLO-v7 在脊柱不规则骨折诊断中表现突出(准确率 88.92%, F1 值 90.63%),有 90.35% 的脊柱骨折病例实现了召回。这一结果提示,不同模型的骨折线特征提取能力与其网络架构深度、特征融合策略密切相关。YOLO-v5 的浅层特征保留能力(通过 CSPDarknet 和 Focus 模块)使其更擅长捕捉高对比度的线性骨折(如四肢长骨骨折),而 YOLO-v7 的深层网络通过增强语义理解能力,更适合处理脊柱等结构复杂部位的骨折。分析 YOLO-v8 在骨盆多发骨折检测较差(准确率 55.65%)原因,可能与其 C2f 模块在特征金字塔中的多尺度融合效率不足有关,导致小目标和重叠结构的漏检率升高。

归因于使用改进的 CSPDarknet 主干, YOLO-v5 具有浅层特征提取优势,通过结合 Focus、BottleneckCSP 和 SPP 3 个模块,颈部使用 PANet 路径聚合网络进行信息融合,在预测头模块上将分类和回归耦合。CSPDarknet 通过跨阶段部分连接有效减少了计算冗余,同时 Focus 模块通过将输入图像拆分、拼接的切片操作增强了局部细节的捕获

能力。这些设计使其在浅层网络中保留了骨折线的边缘信息等更多高分辨率特征,从而在四肢长骨等易辨识骨折检测中表现优异。由于 YOLO-v5 基于整张图像进行预测<sup>[24]</sup>,其路径聚合网络通过自底向上和自顶向下的双向特征融合,进一步提升了多尺度目标的检测灵敏度,保证了尺度大小不一骨折线的顺利检出。

YOLO-v7 通过扩展高效层聚合网络和更深的残差块设计,增强了对 X 线复杂骨折图像的语义理解能力。在脊柱骨折检测中,深层网络能够整合椎体形态改变、骨折线走行等全局信息,识别不规则骨折的差异。但是,这种设计也导致手足细小骨折线等浅层细节在多次下采样过程中丢失。研究中 YOLO-v7 在手足骨折检测中的召回率仅为 49.01%,较 YOLO-v5 低 22.23%,表明其深层架构对局部特征的敏感度较差。分析原因为相较于其他 YOLO 模型, YOLO-v7 主要增加了网络深度,更胜任自然图像复杂场景的检测任务。但骨折线的数字特征单一,为低密度线,四肢长骨骨折等易辨认骨折线的特征主要在模型中的浅层结构辨认,而 YOLO-v7 的推理依据基于深层网络结构提取的特征,浅层易辨认骨折的特征被 YOLO-v7 忽略,造成漏诊。脊柱解剖结构复杂,形态不规则,适合深层网络语义总结归纳识别,因此适用于 YOLO-v7 提取特征。YOLO-v7 通过增加卷积层和残差块提高模型的特征提取能力和检测精度,这种改进使得 YOLO-v7 能够更好地处理复杂场景中的目标检测任务,减少了脊柱骨折的漏检和误检。

YOLO-v8 引入 C2f 模块替代传统的 C3 模块进行残差学习,利用路径聚合网络和特征金字塔网络建立自上而下和自下而上的特征金字塔,直接对骨干不同阶段的输出特征进行上采样,减少了参数量并优化了梯度流动。尽管这一改进在自然图像检测中表现出色,但在 X 线影像中,其简化的特征金字塔导致多尺度信息融合效率下降。在骨盆多发骨折检测中, YOLO-v8 对重叠组织的区分能力较弱,误将肠气伪影识别为骨折,分析原因可能为其激活函数的平滑性抑制了低对比度特征的响应,进一步加剧了小目标的漏检,造成骨盆多发骨折的召回困难。因此, YOLO-v8 基于工业检测的网络优化限制了细小骨折线及多发骨折线的诊断性能,尽管 C2f 模块通过跨阶段部分连接将参数量降低,更有利于工业配置,但其特征金字塔的简化导致多尺度特征融合能力下降,这导致了骨盆多发骨折线的漏诊。

由于低密度伪影与骨折线具有相似的密度特征,研究中假阳性高发于腹部肠气、皮肤褶皱等低密度伪影中。假阴性易发生于骨折线被重叠组织遮挡或非常规体位影像中,分析原因为单阶段目标检测模型无需生成候选区域,它们通过密集的网络或者锚框直接预测目标的类别和位置,具有了更快的推理速度,但对小目标物体检测精度降低<sup>[25]</sup>。

各部位分组中骨盆骨折的召回率最低,分析原因为骨盆诸骨为不规则骨,骨折线形态多样,低密度骨折线在解剖结构重叠下特征提取困难,由于数据偏差及 YOLO 算法注重局部目标检测,对骨折间的空间关系(如受力传导导致的关联骨折、对冲骨折)感知能力差,网络结构难以捕捉跨区域骨折相关性,模型需同时处理不同尺度的骨折框造成假阴性,盆腔中存在的低密度肠气则造成假阳性。应增加训练数据多样性,结合多体位图像综合诊断,或引入注意力机制抑制背景噪声提升诊断准确率。

比较人、人机结合诊断效能的差异,发现 YOLO-v5 辅助诊断的整体准确率及骨折线召回率等均有提升,受益于 AI 模型快速定位骨折区域, YOLO-v5 每病例检出时间缩短至 75 ms/例。提示 YOLO-v5 作为本研究的优势模型,其辅助诊断有利于提升影像医师的诊断准确性及工作效率,尤其适用于急诊、夜班工作场景中,避免误诊漏诊发生。

本研究的局限性为单一中心研究,以临床随访结合所有影像学检查的综合诊断作为金标准对标签分类,这可能产生纳入偏倚,存在过拟合的风险,模型的泛化能力尚需验证。此缺点需在后续研究中进行多中心合作,加大样本含量,增加模型识别不同类型、不同部位、不同形态骨折的泛化能力。研究所使用模型均未根据 X 线平片的数据特征进行改良,这些应用于自然图像的目标检测分类模型显然并不适合直接应用于具有灰度特征的医学数字化图像,这是研究中诊断模型效能受到限制的重要原因。后续研究中将根据数字 X 线平片的单通道特点及灰度特征进行算法及模型结构上的优化,以继续提升模型的诊断效能。

本研究系统评估了 YOLO-v5、YOLO-v7 及 YOLO-v8 3 种单阶段目标检测模型应用于 X 线平片骨折线诊断中的性能差异,并通过人机协作实验探讨了 AI 辅助诊断的临床应用价值。结果显示 YOLO-v5 模型的综合表现最优;YOLO-v7 在脊柱不规则骨折检测中表现较好;YOLO-v8 在骨盆多发骨折检测中准确率较低,选择适宜模型进行辅助诊断可提升医师的诊断效率和准确性。未来需通过

融合 CT、MRI、X 线平片多模态数据、开发专用单通道灰度图像算法等措施优化模型性能,同时规范 AI 诊断模型与临床应用场景整合流程,以推动 AI 从“诊断辅助工具”向“诊断可信工具”演进。

#### [参考文献]

- [1] Han SM, Cao L, Yang C, et al. Value of the 45-degree reverse oblique view of the carpal palm in diagnosing scaphoid waist fractures[J]. *Injury*, 2022, 53(3): 1049-1056.
- [2] Wu J, Liu N, Li X, et al. Convolutional neural network for detecting rib fractures on chest radiographs: A feasibility study[J]. *BMC Med Imaging*, 2023, 23(1): 18-30.
- [3] Srinivasu PN, Kumari GLA, Narahari SC, et al. Exploring the impact of hyperparameter and data augmentation in YOLO V10 for accurate bone fracture detection from X-ray images [J]. *Sci Rep*, 2025, 15(1): 9828-9851.
- [4] 黄晓斌, 吴正参, 黄莹, 等. 基于深度学习的主动脉根部计算机断层扫描血管造影图像快速分割的质量评估与临床价值[J]. *空军军医大学学报*, 2025, 46(2): 187-191.
- [5] 俞世强. 人工智能在主动脉疾病中的应用与展望[J]. *空军军医大学学报*, 2025, 46(2): 141-144.
- [6] 李跃华. 人工智能在评估颅内动脉瘤中的研究进展[J]. *中华放射学杂志*, 2024, 58(11): 1137-1145.
- [7] 赵凯, 郭丽, 张晓东, 等. 人工智能在腰椎影像中的研究现状和进展[J]. *中华放射学杂志*, 2023, 57(4): 424-427.
- [8] 常沙, 闫东, 杜霞, 等. 基于深度学习的骨龄评估系统对生长发育异常儿童骨龄预测的准确性[J]. *中华放射学杂志*, 2023, 57(4): 364-369.
- [9] Yan J, Zeng Y, Lin J, et al. Enhanced object detection in pediatric bronchoscopy images using YOLO-based algorithms with CBAM attention mechanism [J]. *Heliyon*, 2024, 10(12): e32678.
- [10] Baldini C, Migliorelli L, Berardini D, et al. Improving real-time detection of laryngeal lesions in endoscopic images using a decoupled super-resolution enhanced YOLO [J]. *Comput Methods Programs Biomed*, 2025, 260(3): 108539.
- [11] Alam T, Yeh WC, Hsu FR, et al. An integrated approach using YOLOv8 and ResNet, SeResNet & vision transformer (ViT) algorithms based on ROI fracture prediction in X-ray images of the Elbow [J]. *Curr Med Imaging*, 2024, 20(10): e15734056309890.
- [12] Liu S, Zhang W, Wu G. YOLOv8 algorithm-aided detection of rib fracture on multiplane reconstruction images [J]. *Curr Med Imaging*, 2025, 21(3): e15734056337623.
- [13] Zhou Q, Li X, He L, et al. TransVOD: End-to-end video object detection with spatial-temporal transformers [J]. *IEEE T Pattern Anal*, 2023, 45(6): 7853-7869.
- [14] Zhang H, Mao F, Xue M, et al. Knowledge amalgamation for object detection with transformers [J]. *IEEE Trans Image Process*, 2023, 32(1): 2093-2106.
- [15] Mirzaei B, Nezamabadi-Pour H, Raouf A, et al. Small Object detection and tracking: A comprehensive review [J]. *Sensors*

- (Basel), 2023, 23(15):6887-6901.
- [16] Van Leeuwen MC, Fokkinga EP, Huizinga W, et al. Toward versatile small object detection with Temporal-YOLOv8[J]. *Sensors (Basel)*, 2024, 24(22):7387-7392.
- [17] Yamana M, Bise R, Endo M, et al. Development and evaluation of deep learning models for detecting and classifying various bone tumours in full-field limb radiographs using automated object detection models[J]. *Bone Joint Res*, 2025, 14(9):760-768.
- [18] Cheng YC, Hung YC, Huang GH, et al. Deep learning-based object detection strategies for disease detection and localization in chest X-Ray images[J]. *Diagnostics (Basel)*, 2024, 14(23):2636-2655.
- [19] Wu HZ, Yan LF, Liu XQ, et al. The feature ambiguity mitigate operator model helps improve bone fracture detection on X-ray radiograph[J]. *Sci Rep*, 2021, 11(1):1589-1599.
- [20] Lindsey R, Daluiski A, Chopra S, et al. Deep neural network improves fracture detection by clinicians [J]. *Proc Natl Acad Sci USA*, 2018, 115(45):11591-11596.
- [21] Hendrix N, Scholten E, Vernhout B, et al. Development and validation of a convolutional neural network for automated detection of scaphoid fractures on conventional radiographs [J]. *Radiology Artificial intelligence*, 2021, 3(4):e200260.
- [22] Erne F, Dehncke D, Herath SC, et al. Deep learning in the detection of rare fractures-development of a "deep learning convolutional network" model for detecting acetabular fractures[J]. *Z Orthop Unfall*, 2023, 161(1):42-50.
- [23] Suen K, Zhang R, Kutaiba N. Accuracy of wrist fracture detection on radiographs by artificial intelligence compared to human clinicians. A systematic review and meta-analysis[J]. *Eur J Radiol*, 2024, 178(9):111593.
- [24] Li J, Li S, Li X, et al. Primary bone tumor detection and classification in full-field bone radiographs via YOLO deep learning model[J]. *Eur Radiol*, 2023, 33(6):4237-4248.
- [25] Hong Q, Dong H, Deng W, et al. Education robot object detection with a brain-inspired approach integrating Faster R-CNN, YOLOv3, and semi-supervised learning [J]. *Front Neurorobot*, 2023, 17(5):8104-8114.

(本文编辑:王聪)