

## ✧ 医学物理与工程学

## Modified RT-DETR model for automatically differentiating Brucellosis spondylitis and spinal tuberculosis based on fat suppression T2WI

ZHANG Zhaojing<sup>1</sup>, SU Lei<sup>2</sup>, HAO Bin<sup>2</sup>, ZHAO Pengfei<sup>3</sup>, LI Yupu<sup>1</sup>, WANG Ziyi<sup>1</sup>,  
ZHANG Xiaojuan<sup>4</sup>, QIAO Pengfei<sup>3,5\*</sup>

(1. The First Clinical College, 5. Medical Simulation Center, Inner Mongolia Medical University, Hohhot 010000, China;  
2. School of Digital and Intelligent Industry, Inner Mongolia University of Science & Technology, Baotou 014000,  
China; 3. Department of Imaging Diagnosis, the Affiliated Hospital of Inner Mongolia Medical University,  
Hohhot 010000, China; 4. Department of Imaging Diagnosis, Inner Mongolia Hospital of  
Beijing Hospital of Traditional Chinese Medicine, Bayannur 015000, China)

**[Abstract]** **Objective** To observe the value of modified RT-DETR model for automatically differentiating Brucellosis spondylitis (BS) and spinal tuberculosis (ST) based on fat suppression (FS) T2WI. **Methods** Lumbar MRI data of 141 BS and 106 ST patients was retrospectively collected, and the involved vertebrae of BS or ST were labeled on FS T2WI. Then 2 408 labels were obtained and taken as the dataset, including 1 150 BS and 1 258 ST labels, which were divided into training, validation and test sets at ratio of 8 : 1 : 1. Feature enhanced decoder (FED) was introduced into baseline model RT-DETR-R18 framework to construct the FED-based RT-DETR (RT-DETR-FED) model for automatically identifying BS and ST in FS T2WI. Comparative experiment was performed in YOLOv5-S, YOLOX-S, PPYOLOE-S, YOLOv6-S, YOLOv8-S, YOLOv9-S, ViTDet, SwinV2-T and RT-DETR-R18 models, while ablation experiment was performed in RT-DETR-FED model to evaluate their diagnostic efficacy. **Results** Comparative experiment showed that RT-DETR-FED model had the highest accuracy, recall, mean average precision (mAP) at intersection over union (IoU) threshold of 0.5 (mAP@0.5), mAP averaged over IoU thresholds from 0.5 to 0.95 (mAP@0.5 : 0.95) in test set, which was 87.16%, 92.76%, 96.36% and 67.12%, respectively, and its floating point operations per second (FLOPs) was 50.77G. Ablation experiment showed that RT-DETR-FED model achieved the best performance in test set. **Conclusion** Modified RT-DETR model could effectively and automatically differentiate BS and ST based on FS T2WI.

**[Keywords]** spondylitis; Brucellosis; tuberculosis; magnetic resonance imaging; artificial intelligence

**DOI:**10.13929/j.issn.1003-3289.2026.01.022

## 改良 RT-DETR 模型基于脂肪抑制 T2WI 自动鉴别 布鲁氏菌性脊柱炎与脊柱结核

张昭静<sup>1</sup>, 苏蕾<sup>2</sup>, 郝斌<sup>2</sup>, 赵鹏飞<sup>3</sup>, 李宇璞<sup>1</sup>, 王梓怡<sup>1</sup>, 张小娟<sup>4</sup>, 乔鹏飞<sup>3,5\*</sup>

(1. 内蒙古医科大学第一临床医学院, 5. 医学模拟中心, 内蒙古呼和浩特 010000; 2. 内蒙古科技大学数智产业学院,  
内蒙古包头 014000; 3. 内蒙古医科大学附属医院影像诊断科, 内蒙古呼和浩特 010000;  
4. 北京中医医院内蒙古医院影像诊断科, 内蒙古巴彦淖尔 015000)

**[摘要]** **目的** 观察改良 RT-DETR 模型基于脂肪抑制 (FS) T2WI 自动鉴别布鲁氏菌性脊柱炎 (BS) 与脊柱结核 (ST)

**[基金项目]** 内蒙古医科大学附属医院院级科研项目 (2023NYFYYPY007)、内蒙古自治区自然科学基金 (2023QN08044)、内蒙古医学科学院公立医院  
科研联合基金科技项目 (2024GLLH0356)。

**[第一作者]** 张昭静 (1999—), 女, 内蒙古锡林郭勒人, 在读硕士。研究方向: 脊柱炎影像学。E-mail: 1422470312@qq.com

**[通信作者]** 乔鹏飞, 内蒙古医科大学附属医院影像诊断科, 010000; 内蒙古医科大学医学模拟中心, 010000。E-mail: qpfff@126.com

**[收稿日期]** 2025-01-20 **[修回日期]** 2025-05-24

的价值。**方法** 回顾性分析 141 例 BS 及 106 例 ST 的腰椎 MRI 资料,于 FS T2WI 中标注受累椎体,获得 2 408 个标签,包括 1 150 个 BS、1 258 个 ST 作为数据集,并按 8:1:1 比例划分训练集、验证集及测试集。向基线模型 RT-DETR-R18 框架中引入特征增强解码器(FED),得到基于 FDE 的 RT-DETR(RT-DETR-FED)模型用于自动鉴别 BS 与 ST;以对比实验评估 YOLOv5-S、YOLOX-S、PPYOLOE-S、YOLOv6-S、YOLOv8-S、YOLOv9-S、ViTDet、SwinV2-T 及 RT-DETR-R18 模型,并以消融实验评估 RT-DETR-FED 模型的鉴别效能。**结果** 对比实验显示,RT-DETR-FED 模型在测试集的精确率、召回率、交并比(IoU)阈值为 0.5 时的平均精度均值(mAP)(mAP@0.5)及 IoU 阈值为 0.5~0.95 时的 mAP(mAP@0.5:0.95)均最高,分别为 87.16%、92.76%、96.36%及 67.12%,且其浮点运算次数(FLOPs)为 50.77G。消融实验显示,测试集中 RT-DETR-FED 模型效能最佳。**结论** 改良 RT-DETR 模型能 FS T2WI 自动鉴别 BS 与 ST。

**[关键词]** 脊柱炎;布鲁氏菌病;结核;磁共振成像;人工智能

**[中图分类号]** R681.51; R445.2 **[文献标识码]** A **[文章编号]** 1003-3289(2026)01-0107-06

布鲁氏菌性脊柱炎(Brucellosis spondylitis, BS)及脊柱结核(spinal tuberculosis, ST)在我国北部农牧区呈现流行趋势;其早期临床症状及影像学表现均相似,而诊治不及时可致严重并发症<sup>[1]</sup>。人工智能(artificial intelligence, AI)辅助提升了影像学诊断的准确性<sup>[2]</sup>;作为视觉领域核心架构的 YOLO、Swin Transformer 及 ViT 等模型已较为广泛地用于目标检测任务,但其在复杂背景下检测小目标的准确率仍有限<sup>[3]</sup>。实时目标检测 Transformer(real-time detection transformer, RT-DETR)算法<sup>[4]</sup>是基于深度学习的实时端到端目标检测模型,可在保持高精度的同时实现实时性能。本研究观察改良 RT-DETR 模型基于脂肪抑制(fat suppression, FS)T2WI 自动鉴别 BS 与 ST 的价值。

## 1 资料与方法

**1.1 研究对象** 回顾性分析 2015 年 1 月—2024 年 6 月内蒙古医科大学附属医院及北京中医医院内蒙古医院共 141 例 BS(BS 组)及 106 例 ST(ST 组)腰椎 MRI 资料。BS 组男 98 例、女 43 例,年龄 19~84 岁、平均(57.7±10.5)岁;ST 组男 56 例、女 50 例,年龄 25~90 岁、平均(62.7±12.9)岁。纳入标准:①首次确诊<sup>[5]</sup>、接受腰椎 MR 检查并符合 BS 或 ST 诊断标准;②临床、MRI 及实验室资料,以及病原学检测或病理检查结果完善。本研究经院医学伦理委员会批准(YKD202301072、BJZ20241201),免除患者知情同意。

**1.2 仪器与方法** 采用 Siemens Skyra 3.0T、Philips Ingenia 3.0T 或 GE Signa 3.0T 超导型 MR 仪、全脊柱相控阵表面线圈行全脊柱或腰椎扫描,采集常规矢状位 T1WI、T2WI 及矢状位 FS T2WI。矢状位 FS T2WI 参数:采用反转恢复序列、频率选择性反转恢复序列或快速自旋回波序列,TR 1 940~4 000 ms,TE 40~90 ms,FOV 18~28 cm×28~30 cm,层厚 4 mm,层间距 0.4~1.0 mm。

**1.3 图像标注** 对腰椎矢状位 FS T2WI 数据先行 Z-score 归一化预处理后,由 2 名具有 5 年以上 MRI 诊断经验的主治医师采用盲法利用 labellmg 软件(<https://github.com/heartexlabs/labellmg>)于矢状位 FS T2WI 中逐层框选并以 YOLO 格式标注受累椎体,再由 1 名具有 10 年以上工作经验的影像科副主任医师审核决定,共获得 2 408 个标签作为数据集,包括 1 150 个 BS 和 1 258 个 ST 标签;按 8:1:1 比例划分训练集、验证集与测试集;训练集含 920 个 BS、1 006 个 ST 标签,验证集含 115 个 BS、126 个 ST 标签,测试集含 115 个 BS、126 个 ST 标签。

**1.4 构建改良 RT-DETR 模型** 向基线模型 RT-DETR-R18 框架中引入特征增强解码器(feature enhancement decoder, FED)以构建改良 RT-DETR 模型,即基于 FED 的 RT-DETR(RT-DETR with feature enhancement decoder, RT-DETR-FED)模型(图 1);其间通过高效注意力机制 HiLo Attention(图 2)<sup>[6]</sup>优化 AIFI 模块,联合应用高频与低频特征信息建模以提高模型在复杂背景下提取病灶关键特征能力,而引入的自研特征增强扩散网络(feature enhancement diffusion network, FEDNet)可提升模型表示多尺度特征能力。

FEDNet 通过特征增强模块和特征扩散机制而提高模型对多尺度特征的代表能力。特征增强(feature enhancement, FE)模块接收 3 个不同尺度输入(P3、P4、P5)并执行连接操作(Concat),随后通过一组并行深度卷积层(卷积核大小分别为 5、7、9 和 11)捕获多尺度特征、提升检测目标性能;将卷积结果逐层相加后,经过 1×1 卷积层进一步精简,最后与输入特征相加,生成最终的增强特征(图 3A、3B);利用特征扩散机制通过强化不同尺度特征之间的交互以有效解决传统多尺度检测方法中常见的信息丢失问题。

**1.5 实验环境及模型训练设置** 以 Ubuntu 20.04 操作系统及 Intel® Xeon® Silver 4210(主频 2.20 GHz)处

理器为实验环境,在 PyCharm 2023. 3. 5 集成开发环境中开发与调试模型;于 NVIDIA A100 PCIE GPU(显存 40 GB)上完成模型训练及推理;采用 Python 3. 8 编程语言实现算法,深度学习框架为 PyTorch 1. 13. 1 (CUDA 11. 6)。

训练过程中对训练集数据以基于医学影像特征的数据增强技术进行扩充,包括矢状面/轴向旋转( $\pm 15^\circ$ )模拟临床多角度阅片、平移( $\pm 20\%$ 宽度/高度范围)以覆盖病变椎体定位偏差、弹性形变(弹性系数  $\alpha = 50$ , 标准差  $\sigma = 5$ )以反映软组织生理形变与病理结构扭曲、尺度缩放(0. 8~1. 2 倍)及灰度抖动( $\pm 15\%$ 强度)以补偿不同设备所获 MRI 信号差异,并确保椎旁、椎小关节炎等病变结构的解剖学合理性<sup>[7]</sup>;以训练集数据对模型进行训练,于验证集和测试集验证模型鉴别效能。

采用 Adam 优化器训练模型,设置初始学习率 0. 001, 权重衰减 0. 000 5, batch size 16, 图像尺

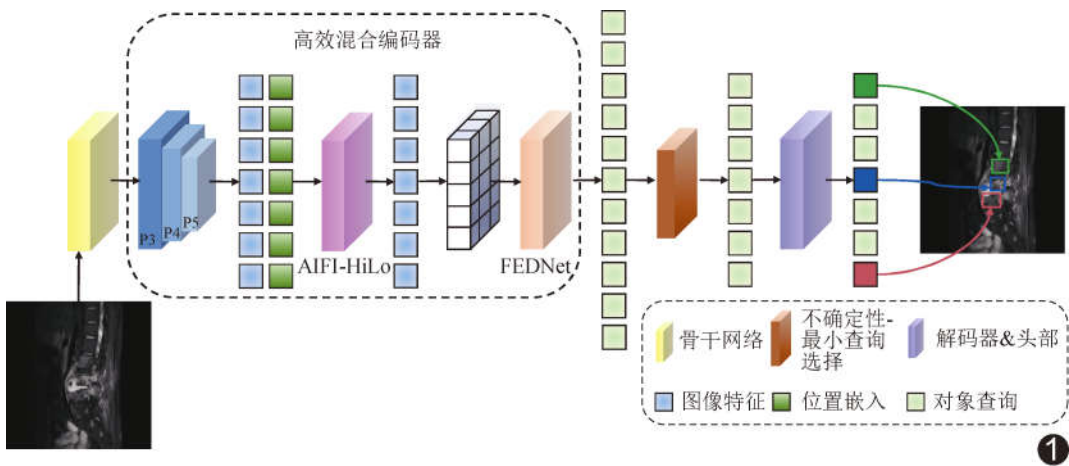


图 1 RT-DETR-FED 网络结构图

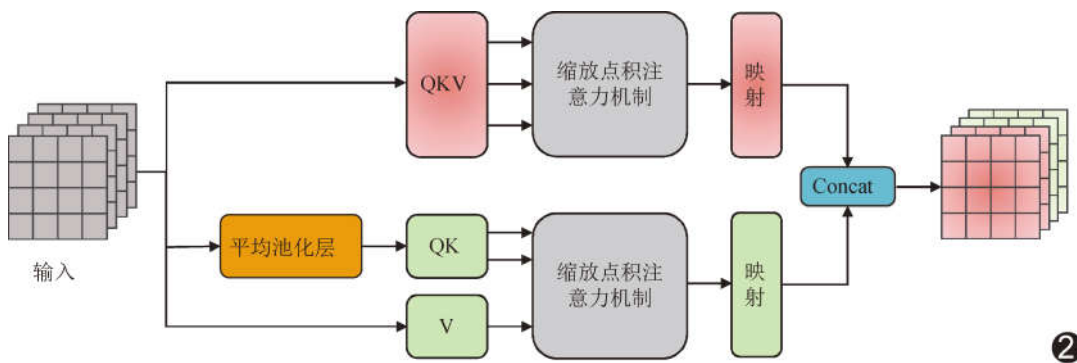


图 2 HiLo Attention 模块结构图

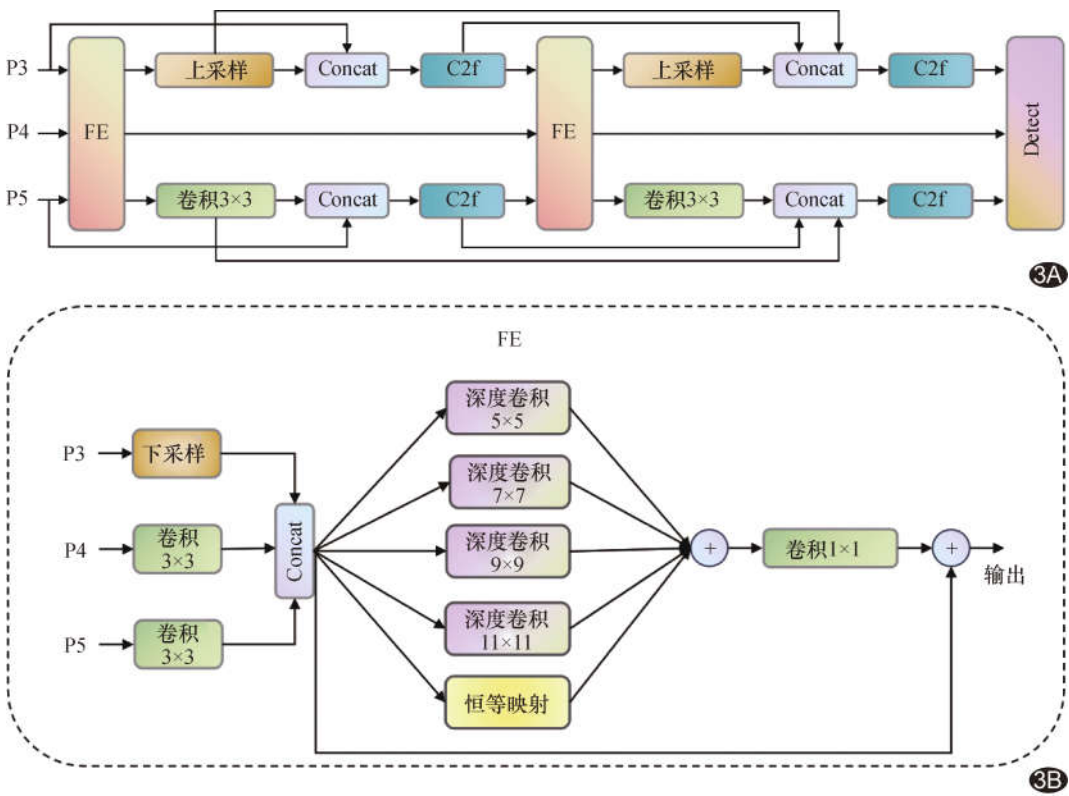


图 3 FEDNet 模块(A)及 FE 模块(B)结构图

寸 640×640, worker 数量为 8。每 5 轮次验证 1 次模型效能;若在验证集中连续 3 轮次未见模型效能提升,则适当降低学习率。整个训练过程持续 50 轮次,直至模型在验证集达到最优性能。

1.6 评价模型效能 选择近年主流目标检测模型 YOLOv5-S、YOLOX-S<sup>[8]</sup>、PPYOLOE-S<sup>[9]</sup>、YOLOv6-S<sup>[10]</sup>、YOLOv8-S、YOLOv9-S<sup>[11]</sup>、ViTDet (ViT-T)<sup>[12]</sup> 及 SwinV2-T<sup>[13]</sup> 于相同训练集和测试集条件下进行训练和测试。采用训练轮次、精确率 (precision, Pr)、召回率 (recall, Rc)、浮点运算次数 (floating-point operations per second, FLOPs)、交并比 (intersection over union, IoU) 阈值为 0.5 时的平均精度均值 (mean average precision, mAP) (mAP at IoU threshold of 0.5, mAP@0.5) 及 IoU 阈值为 0.5 ~ 0.95 时的 mAP (mAP averaged over IoU thresholds from 0.5 to 0.95, mAP@0.5 : 0.95) 对比评估 RT-DETR-FED 与上述模型基于 FS T2WI 自动鉴别 BS 与 TS 的效能。以消融实验评估 RT-DETR-FED 模型效能,以梯度加权类激活映射 (gradient-weighted class activation mapping, Grad-CAM) 可视化模型输出,观察模型对病灶区域的关注情况,并分析训练过程中基线模型与 RT-DETR-FED 模型的收敛速度及稳定性。

## 2 结果

2.1 对比实验 各主流目标检测模型基于 FS T2WI 鉴别测试集 BS 与 ST 的 Pr 为 80.70% ~ 84.79%、Rc 为 80.05% ~ 85.93%、mAP@0.5 为 83.03% ~ 90.02%、mAP@0.5 : 0.95 为 58.87% ~ 62.94%; RT-DETR-R18 效能较之提高,mAP@0.5 和 mAP@0.5 : 0.95 分别达 90.88% 和 65.68%,而 RT-DETR-FED 模型的 Pr、Rc、mAP@0.5 及 mAP@0.5 : 0.95 均最高,分别为 87.16%、92.76%、96.36% 及

67.12%,且其 FLOPs 为 50.77G,显著优于 ViTDet 和 SwinV2-T 等 Transformer 模型。见表 1。

2.2 消融实验 消融实验显示,利用 FEDNet 及 HiLo Attention 改良 RT-DETR-R18 模型后,RT-DETR-FED 模型在测试集的效能最佳,其 mAP@0.5 达 96.39%,mAP@0.5 : 0.95 为 67.18%,见表 2。

2.3 Grad-CAM 及模型损失收敛分析 Grad-CAM 可视化显示,相比 RT-DETR-R18 模型,RT-DETR-FED 用于目标检测区域具有更高激活强度,见图 4。训练集和验证集中,RT-DETR-FED 模型在 50 轮次内的收敛表现均优于 RT-DETR-R18 模型,尤其分类损失具有显著优势,且其在验证集的广义 IoU (generalized IoU, GIoU)、分类及 L1 损失曲线均更为平滑、稳定(图 5)。

表 1 各模型基于 FS T2WI 鉴别测试集 BS 与 ST 的效能

模型	轮次	Pr(%)	Rc(%)	FLOPs	mAP@0.5 (%)	mAP@0.5 : 0.95 (%)
YOLOv5-S	200	83.23	85.75	15.35G	85.21	—
YOLOX-S	200	80.70	80.05	24.24G	83.03	60.35
PPYOLOE-S	200	83.12	83.12	17.04G	83.12	61.60
YOLOv6-S	200	82.73	85.93	45.63G	87.63	60.93
YOLOv8-S	200	84.57	84.55	27.28G	90.02	62.79
YOLOv9-S	200	84.08	82.78	26.44G	89.45	62.94
ViTDet	100	81.40	83.54	93.05G	87.22	61.55
SwinV2-T	100	84.79	80.27	145.68G	88.05	58.87
RT-DETR-R18	50	84.95	86.59	56.98G	90.88	65.68
RT-DETR-FED	50	87.16	92.76	50.77G	96.36	67.12

表 2 HiLo Attention 及 FEDNet 改良 RT-DETR 模型在测试集的消融实验结果

HiLo Attention	FEDNet	Pr(%)	Rc(%)	FLOPs	mAP@0.5 (%)	mAP@0.5 : 0.95 (%)
×	×	84.91	86.55	56.94G	90.81	65.64
×	√	86.25	90.47	51.05G	94.54	66.45
√	×	85.34	87.85	56.07G	93.95	65.87
√	√	87.13	92.73	50.71G	96.39	67.18

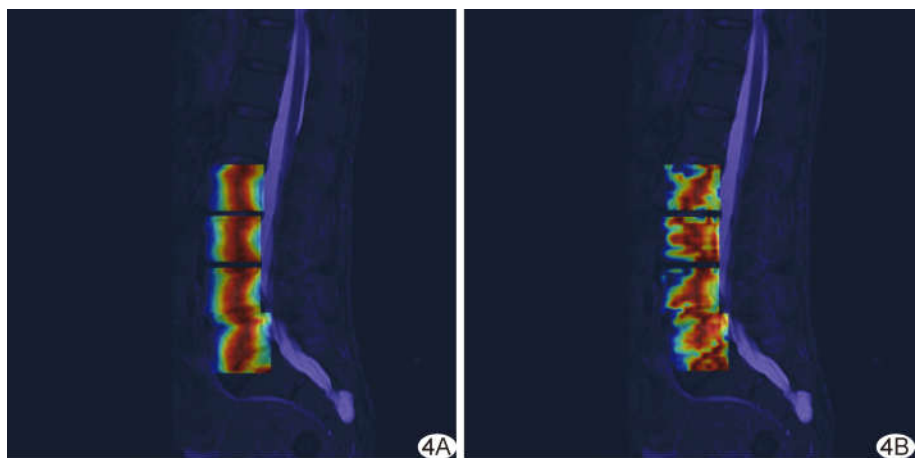


图 4 RT-DETR-R18(A)及 RT-DETR-FED 模型(B)目标检测区域的 Grad-CAM 可视化热图

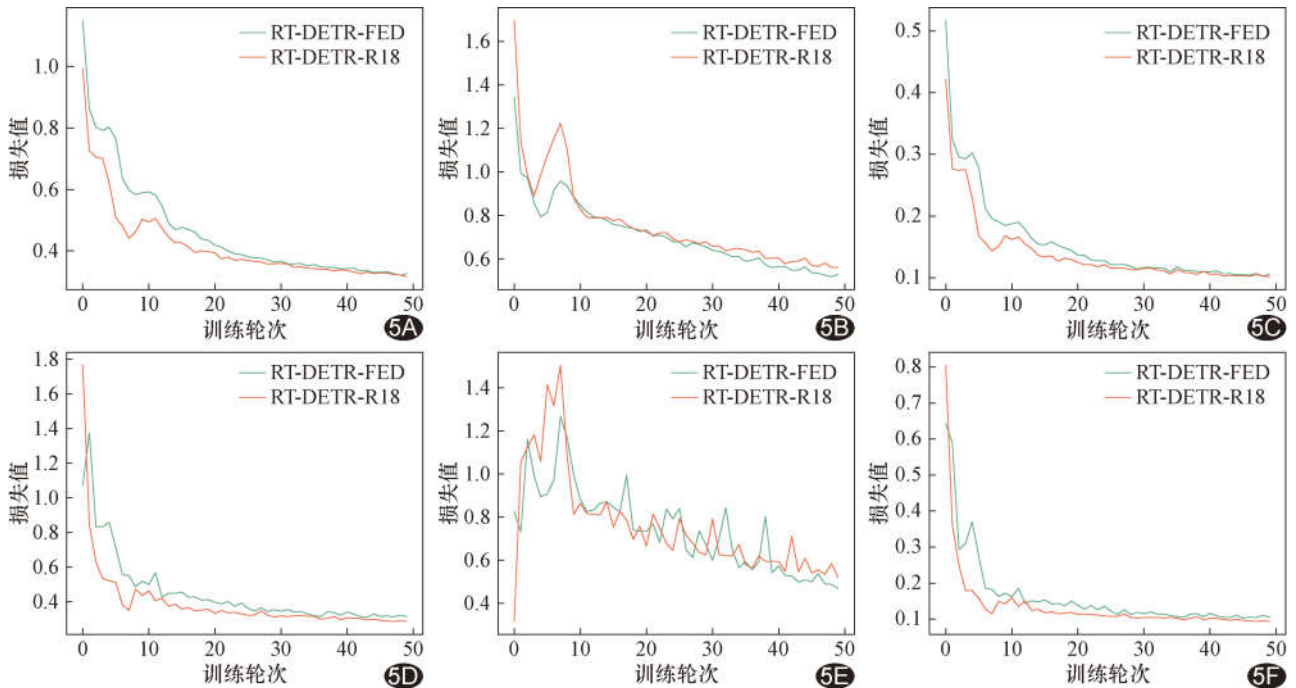


图5 对比分析 RT-DETR-FED 与 RT-DETR-R18 模型的损失收敛 A~C. 在训练集的 GloU (A)、分类 (B) 及 L1 (C) 损失曲线图; D~F. 在验证集的 GloU (D)、分类 (E) 及 L1 (F) 损失曲线图

### 3 讨论

BS 与 ST 临床症状及体征均相似, 鉴别诊断难度较大, 易致延误治疗而引发脊柱畸形、脊髓及神经根受压等严重并发症<sup>[14]</sup>; MRI 有助于鉴别, 但部分表现仍存在重叠<sup>[15]</sup>。何熊等<sup>[16]</sup>构建 4 种机器学习模型以鉴别 BS 与 ST, 以 LR 模型的综合诊断效能较好, 其在训练集及测试集的 AUC 分别为 0.88 及 0.85, 表明 MRI 影像组学结合机器学习鉴别 BS 与 ST 具有可行性。

RT-DETR 为轻量级目标检测框架, 具备高推理速度及并行处理优势。本研究于基线模型 RT-DETR-R18 中引入高效注意力机制 HiLo Attention 及 FEDNet 以开发改良 RT-DETR 模型, 以所获 RT-DETR-FED 模型基于 FS T2WI 自动鉴别 BS 与 ST, 并与近年主流目标检测模型进行对比; 结果显示, YOLO 系列模型存在推理速度慢、无法并行处理、不同场景需更换 NMS 阈值等问题, 其检测精度及速度逊于同等规模的 RT-DETR 模型<sup>[4]</sup>; 其中 YOLOv8-S 模型相对最佳, 但在复杂背景、病灶边界不清或多病灶场景下易出现误检及漏检<sup>[3]</sup>; ViTDeT 及 SwinV2-T 模型存在结构较复杂、推理速度较慢等缺点, 虽在全局特征建模方面具有优势, 但缺乏小目标信息, 使其分割精细解剖结构时存在误差<sup>[17]</sup>。RT-DETR 系列模型

均训练 50 轮次, 诊断效能均优异, 尤以 RT-DETR-FED 模型表现最佳, 其 Pr、Rc、mAP@0.5 及 mAP@0.5 : 0.95 均最高, 且计算效率良好, 表明改进模块后模型效能显著提升, 实现了更强的特征表达, 用于病灶形态不一的脊柱病变表现为更稳定, 且能控制模型计算量, 在检测精度与效率方面兼具优势。

本研究消融实验表明, 以 FEDNet 及 HiLo Attention 模块改良 RT-DETR-R18 模型后, 所获 RT-DETR-FED 模型在测试集效能最佳, 其 mAP@0.5 达 96.39%, mAP@0.5 : 0.95 为 67.18%, 且 Grad-CAM 热图显示 RT-DETR-FED 模型目标检测区域呈现更高激活强度, 提示其能有效提取并利用目标特征信息, 显著提升多尺度特征的代表能力, 有助于提升检测精度和泛化能力, 更适用于鉴别 BS 与 ST。模型损失收敛分析结果表明, RT-DETR-FED 模型的收敛表现优于 RT-DETR-R18 模型, 尤其分类损失和 L1 损失曲线的收敛趋势更稳定, 用于小样本数据集泛化能力更强。

综上, 改良 RT-DETR 模型能基于 FS T2WI 有效自动鉴别 BS 与 ST。但本研究为回顾性分析, 且样本量有限, 难免存在选择偏倚, 有待后续加以完善。

利益冲突: 全体作者声明无利益冲突。

作者贡献: 张昭静研究设计和实施、统计分析、撰写和修改文章; 苏蕾和郝斌研究实施、统计分析; 赵鹏

飞和李宇璞数据分析;王梓怡和张小娟查阅文献;乔鹏飞指导、审阅文章、经费支持。

### [参考文献]

- [1] 贾社星,赵洪增.布氏杆菌性脊柱炎误诊脊柱结核 12 例[J].临床心身疾病杂志,2019,25(4):78-80.
- [2] 胡逸凡,潘佳叶,宫悦,等.基于深度学习影像组学在食管癌的研究进展[J].放射学实践,2024,39(8):1099-1103.
- [3] 王琳毅,白静,李文静,等.YOLO 系列目标检测算法研究进展[J].计算机工程与应用,2023,59(14):15-29.
- [4] ZHAO Y, LV W, XU S, et al. DETRs beat YOLOs on real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024:16965-16974.
- [5] 中华医学会放射学分会传染病学组,中国研究型医院学会感染与炎症放射学专业委员会,中国艾滋病性病防治协会艾滋病放射学专业委员会,等.影像学诊断布鲁氏菌性脊柱炎专家共识[J].中国医学影像技术,2023,39(7):961-965.
- [6] PAN Z, CAI J, ZHUANG B. Fast vision transformers with HiLo Attention [C]. Advances in Neural Information Processing Systems 35, NeurIPS 2022, the Federal Democratic Republic of Ethiopia, 2022:14541-14554.
- [7] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis[J]. Med Image Anal, 2017, 42:60-88.
- [8] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021:2831-2840.
- [9] XU S, WANG C, JIANG B, et al. PP-YOLOE: An evolved version of YOLO [J/OL]. arXiv preprint arXiv: 2203.16250, 2022. <https://doi.org/10.48550/arXiv.2203.16250>.
- [10] LI C, LI L, GENG Y, et al. YOLOv6 v3.0: A full-scale reloading [J/OL]. arXiv preprint arXiv: 2301.05586, 2023. <https://doi.org/10.48550/arXiv.2301.05586>.
- [11] WANG C Y, YE H I H, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable gradient information[C]//Proceedings of the European Conference on Computer Vision, Milan, 2024. Cham:Springer, 2025:1-21.
- [12] LI Y, MAO H, GIRSHICK R, et al. Exploring plain vision transformer backbones for object detection [J/OL]. arXiv preprint arXiv: 2203.16527, 2022. <https://doi.org/10.48550/arXiv.2203.16527>.
- [13] LIU Z, HU H, LIN Y, et al. Swin transformer v2: Scaling up capacity and resolution [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ:IEEE, 2022:12009-12019.
- [14] LIU S, SOONTORNCHAI S, BOCORNKITTI S, et al. Epidemiological characteristics and spatio-temporal clusters of human brucellosis in Inner Mongolia, 2010—2021 [J]. BMC Infect Dis, 2024,24(1):1321.
- [15] 王浩华,赵鹏飞,乔鹏飞.体素内不相干运动及动态对比增强 MRI 评估常规 MRI 无异常的布氏杆菌性脊柱炎病变[J].中国医学影像技术,2024,40(8):1226-1230.
- [16] 何雄,陈艳丽,帕哈提·吐逊江,等.MR T1WI 影像组学对结核性脊柱炎与布鲁菌性脊柱炎的诊断价值[J].分子影像学杂志,2023,46(3):442-447.
- [17] 孙红,莫光萍,徐广辉,等.基于融合视觉 Transformer 与边缘引导编解码网络(RET-Net)算法分割脊柱 MRI[J].中国医学影像技术,2023,39(4):577-581.