

## ❖ 医学物理与工程学

## Automatic detection of mandibular second molar external root resorption based on channel and spatial multi-attention collaborative network model

QIAN Zecheng<sup>1</sup>, DAI Xiubin<sup>2\*</sup>, ZHU Shujin<sup>2</sup>, MAO Tianyi<sup>2</sup>, WANG Dongmiao<sup>3</sup>

(1. School of Chemistry and Life Sciences, 2. College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 3. Department of Oral and Maxillofacial Surgery, Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing 211166, China)

**[Abstract]** **Objective** To explore the value of channel and spatial multi-attention collaborative network model for automatic detection of mandibular second molar (MM2) external root resorption (EER). **Methods** An automatic detection network model integrated channel attention module with spatial-channel transformation fusion attention module was constructed, and its performance for automatically detecting MM2-EER was evaluated according to sensitivity (Sen), specificity (Spe), accuracy (Acc) and precision (Pre). The area under the curve (AUC) and the mean average precision at intersection over union (IoU) threshold 0.5 (mAP@0.5) of the model were compared with those of manual interpretation and other network models including AlexNet, GoogLeNet, VGG-16, ResNet-50 and YOLOv5x. **Results** A channel and spatial multi-attention collaborative network model was successfully constructed, with Sen, Spe, Acc, Pre and AUC in test set of 88.64%, 81.82%, 85.23%, 82.98% and 0.867, respectively, all were the highest ones, and its mAP@0.5 was 0.863. **Conclusion** Spatial multi-attention collaborative network model had good performance for automatically detecting MM2-EER.

**[Keywords]** mandibular second molar; external root resorption; radiography, panoramic; channel and spatial multi-attention; deep learning

DOI:10.13929/j.issn.1003-3289.2026.03.020

## 基于通道和空间多重注意力协同网络模型 自动检测下颌第二磨牙牙根外吸收

钱泽程<sup>1</sup>, 戴修斌<sup>2\*</sup>, 朱书进<sup>2</sup>, 冒添逸<sup>2</sup>, 王东苗<sup>3</sup>

(1. 南京邮电大学化学与生命科学学院, 2. 自动化学院, 江苏 南京 210023;  
3. 南京医科大学附属口腔医院口腔颌面外科, 江苏 南京 211166)

**[摘要]** **目的** 观察基于通道和空间多重注意力协同网络模型自动检测下颌第二磨牙(MM2)牙根外吸收(ERR)的价值。**方法** 构建融合通道注意力模块与空间-通道转换融合注意力模块的自动检测网络模型,以敏感度(Sen)、特异度(Spe)、准确率(Acc)和精确度(Pre)评估其自动检测 MM2 ERR 的效能;根据曲线下面积(AUC)和交并比(IoU)为 0.5 时的平均精度均值(mAP@0.5)等对该模型与人工判读和其他网络模型如 AlexNet、GoogLeNet、VGG-16、ResNet-50 及 YOLOv5x 进行对比。**结果** 所获通道和空间多重注意力协同网络模型在测试集的 Sen、Spe、Acc、Pre 及 AUC 均最高,分别为 88.64%、81.82%、85.23%、82.98%及 0.867,且其 mAP@0.5 为 0.863。**结论** 基于通道和空间多重注意力协同网络模型用于自动检测 MM2-ERR 效能良好。

**[基金项目]** 江苏省社会发展——临床前沿技术项目(BE2023833)。

**[第一作者]** 钱泽程(2001—),男,江苏常州人,在读硕士。研究方向:医学影像处理及分析。E-mail: zecheng\_qian@163.com

**[通信作者]** 戴修斌,南京邮电大学自动化学院,210023。E-mail: daixb@njupt.edu.cn

**[收稿日期]** 2025-08-22 **[修回日期]** 2025-11-06

[关键词] 下颌第二磨牙; 牙根外吸收; 放射摄影术, 全景; 通道和空间多重注意力; 深度学习

[中图分类号] R781.3; R814.42 [文献标识码] A [文章编号] 1003-3289(2026)03-0419-06

下颌第二磨牙(mandibular second molar, MM2)牙根外吸收(external root resorption, ERR)指 MM2 牙根表面进行病理性吸收,影像学可表现为牙根边缘模糊或边界清晰的孤立区域等<sup>[1-2]</sup>;正畸治疗和阻生磨牙挤压等为其常见病因<sup>[1]</sup>,如未能及时诊治,疾病严重性将随年龄增长而加重、甚至需要拔除,及时发现 ERR 具有重要临床意义<sup>[2-3]</sup>。相比锥形束 CT(cone-beam CT, CBCT),全景片(panoramic radiography, PR)辐射剂量和成本均较低且可直接显示口腔全貌,故多首选 PR 以初步诊断 MM2-ERR<sup>[4]</sup>,但缺少三维结构信息使其诊断精度受限;而基于深度学习的 ERR 检测法<sup>[5-7]</sup>的分类效能虽略优于医师,却不能自动定位病灶。本研究观察基于通道和空间多重注意力协同网络自动检测 MM2-ERR 的价值。

### 1 资料与方法

1.1 研究对象 回顾性纳入 2019 年 1 月—2022 年 12 月在南京医科大学附属口腔医院先后接受 PR 与 CBCT 检查的 655 例患者,男 324 例、女 331 例,年龄

18~52 岁、平均(34.2±8.7)岁;2 次检查间隔不超过 3 个月。本研究经医院及大学伦理委员会批准(南医口院伦审 PJ2020-161-001、2025B012)。

1.2 仪器与方法 采用全景 X 光机 OC200D 和 NewTom 5G CBCT 扫描仪采集图像并以 DICOM 格式储存。工作站搭载 13th Gen Intel (R) Core (TM) i7-13700KF 3.40GHz 的 16 核中央处理器,并配有 GeForce RTX 3090 图形处理器。

以取自 PR 图像的 MM2 位置信息和于 CBCT 图像中获得的 ERR 类别信息作为金标准构建训练数据,并对训练所用图像进行数据增强操作;之后构建通道和空间多重注意力协同的 ERR 自动检测网络。将新的 PR 图像输入至已训练完成的网络模型,生成多个候选框及对应的目标置信度、类别置信度,并利用非极大值抑制(non-maximum suppression, NMS)<sup>[8]</sup>算法筛选 MM2 所在区域的最终位置,以及是否发生 ERR 的分类预测结果。其流程图见图 1。

1.3 数据标注及处理 PR 图像尺寸为 2 976×1 536,

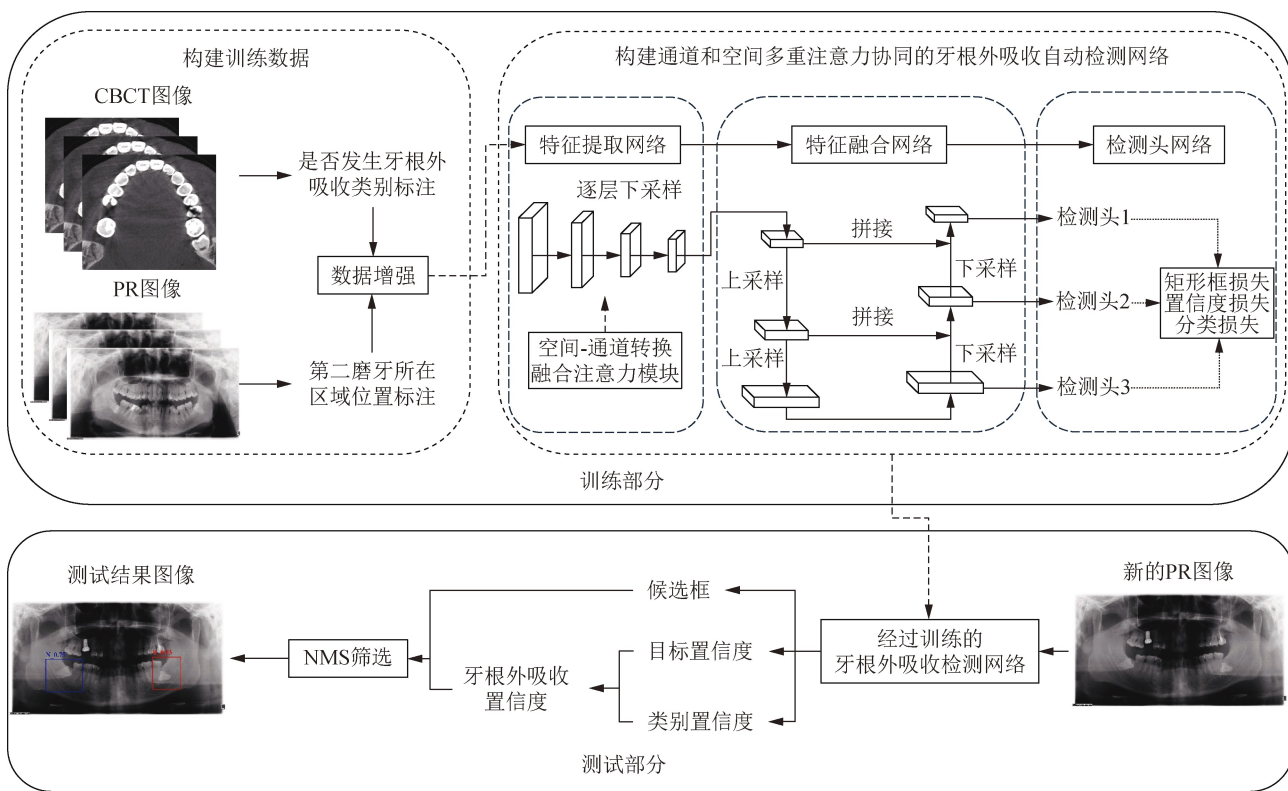


图 1 基于通道和空间多重注意力协同网络模型自动检测 MM2-ERR 流程图

其像素尺寸  $0.07 \text{ mm} \times 0.07 \text{ mm}$ ; CBCT 图像尺寸为  $10 \times 410 \times 273$ , 体素尺寸  $0.3 \text{ mm} \times 0.3 \text{ mm} \times 0.3 \text{ mm}$ 。由 3 名具有 2 年相关经验的口腔科主治医师独立对图像进行判读, 意见不一致时提请由 1 名具有 20 年经验的口腔科主任医师决定; 由此共获得 655 幅图像、870 份 MM2 样本, 含发生与未发生 ERR 各 435 份。按 8 : 1 : 1 比例将其随机划分为训练集 (694 份)、验证集 (88 份) 及测试集 (88 份)。分别由 2 组医师 (其一为 3 名具有 20 年经验的口腔科主任医师, 另一为 3 名具有 2 年相关经验的口腔科主治医师) 独立根据 PR 图像判断测试集有无 MM2-ERR 而不借助对应 CBCT 图像, 取 2 组医师判读结果的平均值作为结果。训练过程中采用 Mosaic、MixUp 和 Copy-Paste 等<sup>[9]</sup> 数据增强方法实时生成多样化数据并输入所获网络, 依次设置其概率为 1.0、0.1 和 0.1; 将 525 幅图用于训练, 每个训练轮次增强均扩充 525 幅新图, 共训练 300 轮次、扩充了 157 500 幅新图。

1.4 构建模型 构建通道和空间多重注意力协同的卷积神经网络, 包括特征提取网络、特征融合网络和检测头网络三部分。其中的特征提取网络包括卷积-批归一化-Mish 激活函数 (convolution-batch normalization-Mish, CBM)<sup>[8]</sup> 组合层、残差多尺度卷积模块一 (ResMultiC1\_n)、压缩与激励 (squeeze-and-

excitation, SE)<sup>[10]</sup> 通道注意力模块、空间-通道转换融合 (spatial-channel transform fusion, SCTR) 注意力模块及快速空间金字塔池化 (spatial pyramid pooling-fast, SPPF) 模块。输入特征并经过 2 个 CBM 后依次经过 ResMultiC1\_4、CBM 和 ResMultiC1\_8 而得到特征 1, 通过 CBM 和 ResMultiC1\_12 得到特征 2; 之后经过 SCTR、CBM、ResMultiC1\_4 和 SPPF 的组合操作得到特征 3; 以 SE 模块处理特征 1 和特征 2 后得到特征 4 和特征 5。特征融合网络包括深度可分离卷积-批归一化-Mish 激活函数 (depthwise separable convolution-batch normalization-Mish, DBM) 组合层、子像素卷积 (sub-pixel convolution, Sub-pixel Conv)<sup>[11]</sup>、特征拼接和残差多尺度卷积模块二 (ResMultiC2)。特征 3 经过 DBM 后得到特征 6, 特征 6 经过 Sub-pixel Conv 后与特征 5 拼接, 随后经过 ResMultiC2 和 DBM 后得到特征 7; 特征 7 经过 Sub-pixel Conv 后与特征 4 拼接, 随后经过 ResMultiC2 得到输出特征 1; 输出特征 1 通过 DBM 后与特征 7 拼接, 再经 ResMultiC2 得到输出特征 2; 输出特征 2 通过 DBM 后与特征 6 拼接, 经过 ResMultiC2 得到输出特征 3。检测头网络接收 3 个不同尺度的输出特征, 每个特征分别经过 CBM, 最终输出多尺度融合后的检测结果。见图 2。

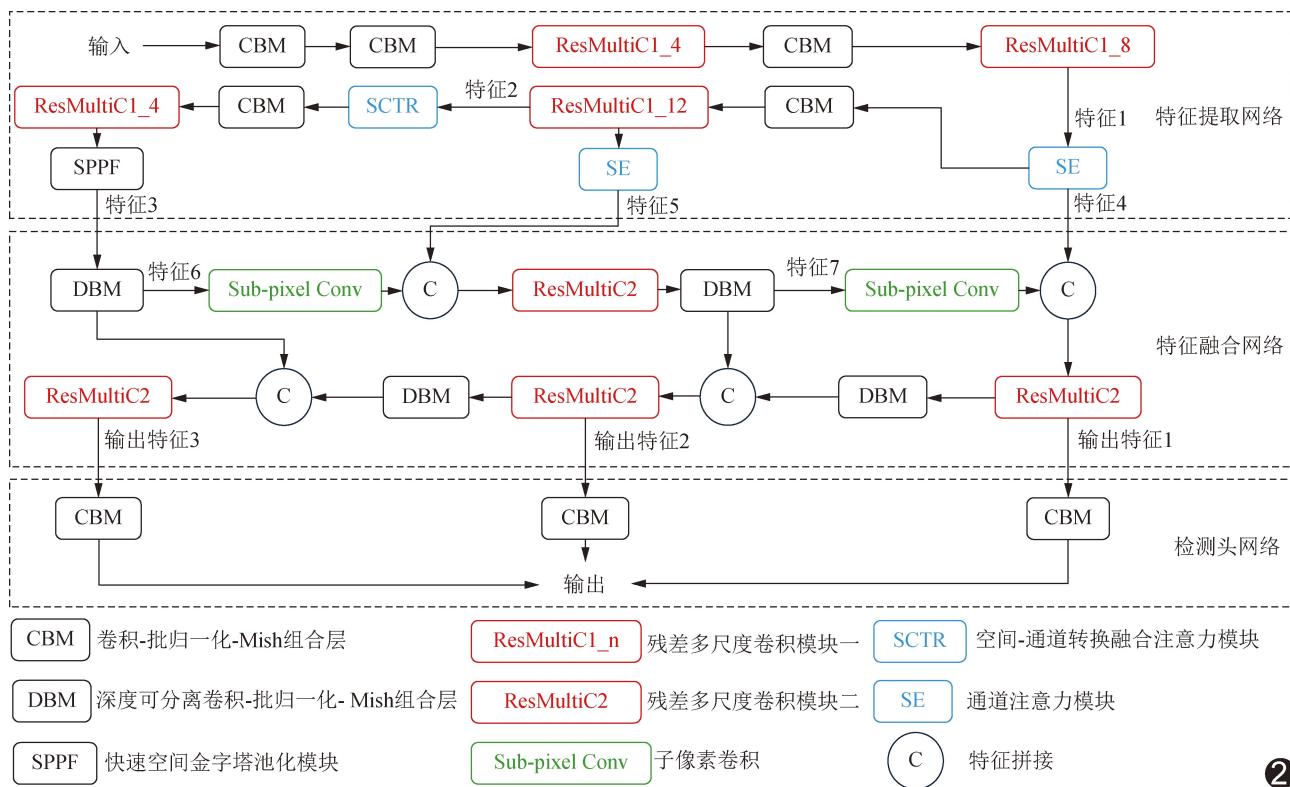


图 2 基于通道和空间多重注意力协同网络模型的结构图

1.4.1 SCTR 首先对输入特征  $X$  进行组归一化 (group normalization, GN) 处理而得到输出  $X_1$ 。见公式(1):

$$X_1 = \text{GN}(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \quad (1)$$

其中,  $\mu$  和  $\sigma$  分别为均值和标准差,  $\epsilon$  为旨在提升稳定性的极小常数。GN 处理利用可学习参数  $\gamma$  和  $\beta$  对特征进行缩放和偏移, 以保持模型表达能力。

通过缩放因子归一化操作得到缩放因子  $W_\gamma$ , 以保证后续通过门控单元时通道权重的相对尺度公平, 见公式(2):

$$W_\gamma = \frac{\gamma}{\sum_{i=1}^c \gamma_i} \quad (2)$$

逐通道将  $W_\gamma$  与特征  $X_1$  相乘; 通过 Sigmoid 激活函数将重加权结果映射至该范围以生成门控权重并将门控掩码约束在 0~1 区间。利用门控单元、依据门控阈值分割重标定后的门控权重, 使特征通道根据其重要性产生差异化响应。分别将阈值以上、以下设为注意力权重  $t_1$ 、 $t_2$ , 逐通道与  $X_1$  相乘, 得到信息丰富且用于强化物体主体的加权特征  $Y_1$  和信息较少且用于保留潜在细节的加权特征  $Y_2$ 。

沿通道维度对  $Y_1$  与  $Y_2$  进行拼接而得到特征  $Y$ ; 对  $Y$  作自适应平均池化, 得到空间-通道信息描述符  $G_1$ 、 $G_2$ , 并通过 SoftMax 函数得到特征权重向量  $s_1$ 、 $s_2$ ; 将  $s_1$ 、 $s_2$  与特征  $Y$  逐元素相乘而得到特征  $Z$ , 再重新沿通道维度对特征进行  $Z$  分割及逐通道相加, 即得整个 SCTR 模块的输出  $X_{out}$ 。见图 3。

1.4.2 ResMultiC1\_n 和 ResMultiC2 在 ResMultiC1\_n 模块中

输入特征, 依次经过 1 个 CBM 与  $n$  个基本单元后, 与首个 CBM 输出特征进行拼接。按照通道数将拼接后的特征均分为 3 路, 分别经过卷积核大小为 1、3、5 的卷积后, 对得到的 3 路特征进行再次拼接。随后再次使拼接后的特征依次经过批归一化和 Mish 激活函数, 得到模块输出。其中的基本单元输入特征经过 2 个 CBM 后与输入特征进行逐元素相加, 产生输出特征。相比 ResMultiC1\_n, ResMultiC2 将其中的  $n$  个基本单元替换为 8 个 CBM。见图 4。

1.5 网络训练 为平衡定位和分类任务学习, 训练网络模型以矩形框损失、置信度损失和分类损失之和作为总损失函数, 以完全交并比 (complete intersection over union, CIoU)<sup>[12]</sup> 损失定义矩形框损失, 以二元交叉熵 (binary cross entropy, BCE)<sup>[12]</sup> 损失定义置信度损失和分类损失。

1.6 评估模型效能 以敏感度 (sensitivity, Sen)、特异度 (specificity, Spe)、准确率 (accuracy, Acc) 和精确度 (precision, Pre) 评估模型效能, 并采用 McNemar 配对  $\chi^2$  检验进行比较。绘制受试者工作特征 (receiver

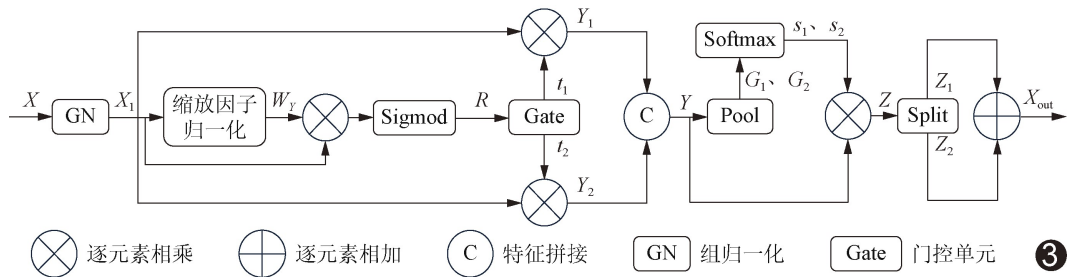


图 3 SCTR 模块结构图

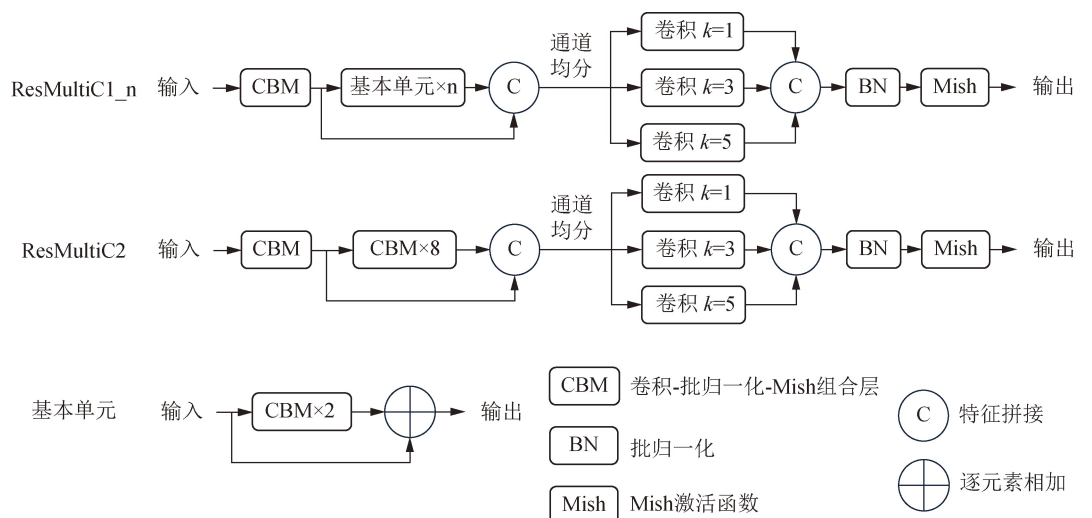


图 4 ResMultiC1\_n 和 ResMultiC2 结构图

operating characteristic, ROC)曲线,计算曲线下面积(area under the curve, AUC),以 DeLong 检验加以比较,并以所获模型与人工判读及其他网络模型(AlexNet、GoogLeNet、VGG-16、ResNet-50、YOLOv5x)在交并比为 0.5 时的平均精度(mAP@0.5)作为补充指标<sup>[13]</sup>。P<0.05 为差异有统计学意义。

## 2 结果

成功构建了基于通道和空间多重注意力协同网络模型。PR 图像中单侧 MM2 缺失时,所获模型仅在保有 MM2 的一侧输出预测框,符合实际临床,体现了模型在不对称口腔结构下定位和分类性能的鲁棒性。见图 5。

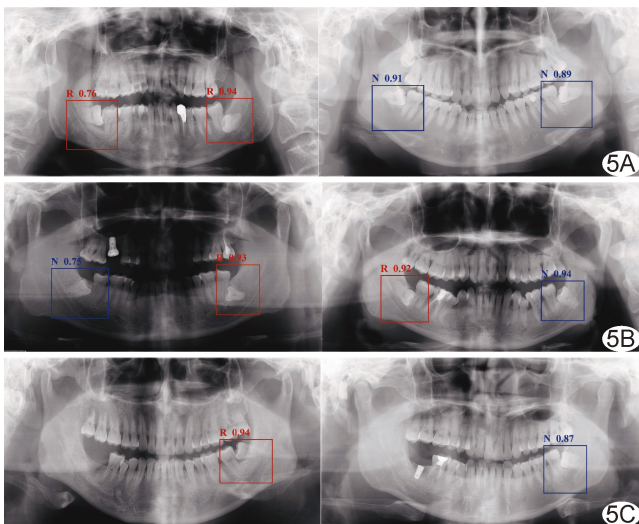


图 5 所获模型预测 MM2-ERR 结果 A. 双侧 MM2 均存在,均发生与未发生 ERR; B. 双侧均见 MM2 且均其中一侧发生 ERR; C. 仅存在单侧 MM2 且发生与未发生 ERR (蓝字母 N 代表 MM2 未发生 ERR,红字母 R 代表 MM2 发生 ERR,数字为预测概率)

2.1 模型效能 训练模型的最优轮次、批大小和学习率分别为 300、16 和 0.007 5。见表 1。

表 1 不同超参数配置对应的分类性能指标对比(%)

参数	Sen	Spe	Acc	Pre
训练轮次	250	93.18	79.55	86.36
	300	88.64	81.82	85.23
	350	90.91	68.18	79.55
批大小	8	79.55	75.00	77.27
	16	88.64	81.82	85.23
	24	75.00	90.91	82.95
学习率	0.005	77.27	81.82	79.55
	0.007 5	88.64	81.82	85.23
	0.01	79.55	75.00	77.27

2.2 对比 所获模型外,仅 YOLOv5x 效能指标与人工判读较接近,而所获模型各项指标均高于 YOLOv5x,见表 2。2 组医师评估结果差异无统计学意义(P=0.839),所获模型与人工判读(主任医师 P=0.076)结果差异无统计学意义;所获模型与人工判读(主治医师 P=0.029)及 YOLOv5x 模型(P=0.041)差异均有统计学意义。

表 2 对比所获模型与人工判读及其他网络模型效能指标(%)

方法	Sen	Spe	Acc	Pre
人工判读	71.59	73.86	72.73	73.26
AlexNet	75.00	61.36	68.18	66.00
GoogLeNet	79.55	61.36	70.45	67.31
VGG-16	75.00	65.91	70.45	68.75
ResNet-50	61.36	65.91	63.64	64.29
YOLOv5x	77.27	70.45	73.86	72.34
本研究模型	88.64	81.82	85.23	82.98

所获模型预测 MM2-ERR 的 AUC 为 0.867,相比 YOLOv5x(0.802)提升 8.10%,但差异无统计学意义(P=0.159)。DeLong 检验显示所获模型判断 MM2-ERR 的 AUC 均高于 AlexNet、GoogLeNet、VGG-16 及 ResNet-50 模型(P 均<0.05),见图 6;其 mAP@0.5 为 0.863。

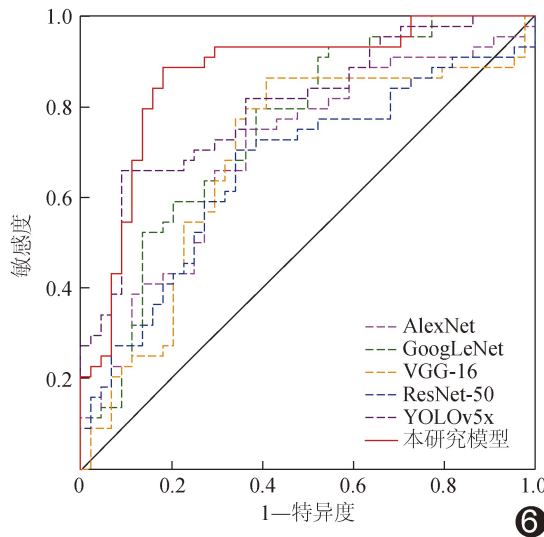


图 6 所获模型与其他网络模型预测 MM2-ERR 的 ROC 曲线

## 3 讨论

本研究构建基于空间和通道多重注意力协同网络模型,以提高基于 PR 识别 MM2-ERR 的精度和效率。针对微小目标任务如检测 ERR,细节和边缘信息尤为关键。本研究所获模型利用 3 项创新突破检测精度:

①向 SCTR 模块引入空间感知门控机制,在通道维度上动态划分并重构特征,使特征提取过程同时具备在空间和通道维度上增强有效特征并抑制冗余干扰的能力,在特征提取网络的中深层部分联合使用 SCTR 模块和 SE 模块,利用空间-通道协同优化精准增强重要特征;②ResMultiC1\_n 和 ResMultiC2 实现了在模块内保留细节的同时融合多尺度特征;③选择 Sub-pixel Conv 进行上采样,在有效保留特征局部细节和低频信息的同时增强其边缘和结构表达能力,进而强化检测微小目标能力。

针对单侧 MM2 缺失、双侧 ERR 不对称等复杂场景,本研究所获模型能自动规避无牙区域,仅于存在 MM2 位置输出检测结果,符合临床实际需求;其与 YOLOv5x 在测试集无显著差异,可能与测试集样本量相对有限、导致检验效能不足有关。本研究所获模型的 Sen、Spe、Acc、Pre、AUC 及 mAP@0.5 均最优,整体判别能力和检测精度均存在一定优势;且直接基于 PR 图像检测 ERR,无需人工裁剪,具有显著效率优势。

综上,通道和空间多重注意力的协同网络模型自动检测 ERR 效能良好,有望辅助临床优化治疗决策,降低漏、误诊风险。

利益冲突:全体作者声明无利益冲突。

作者贡献:钱泽程研究设计和实施、撰写和修改文章;戴修斌指导、研究设计、修改和审阅文章;朱书进和冒添逸指导、研究设计;王东苗数据分析。

### [参考文献]

- [1] HEBOYAN A, AVETISYAN A, KAROBARI M I, et al. Tooth root resorption: A review [J]. Sci Prog, 2022, 105(3):368504221109217.
- [2] 周蕾,王传江.下颌完全埋伏阻生智齿引起邻近第二磨牙牙根外吸收及牙周改变的 CBCT 研究[J].临床口腔医学杂志,2023,39(4):235-238.
- [3] 潘子建,潘剑.阻生第三磨牙引起邻牙牙根外吸收的研究进展[J].国际口腔医学杂志,2024,51(5):616-623.
- [4] MOREIRA-SOUZA L, OLIVEIRA L B, GAËTA-ARAÚJO H, et al. Comparison of CBCT and panoramic radiography for the assessment of bone loss and root resorption on the second molar associated with third molar impaction: A systematic review [J]. Dentomaxillofac Radiol, 2022,51(3): 20210217.
- [5] TAMURA K, KAMIYA T, ODA M, et al. Detection of the root resorption from panoramic X-Ray images using deep metric learning [C]//2021 International Conference on Control, Automation and Systems (ICCAS), Jeju: 2021:1800-1803.
- [6] TAMURA K, KAMIYA T, ODA M, et al. Classification the root resorption from panoramic X-ray image using center loss redefined in angle space [C]//2022 International Conference on Control, Automation and Systems (ICCAS), Busan: 2022: 570-573.
- [7] SAKATA K, KAMIYA T, ODA M, et al. Identification of root resorption on panoramic X-ray image based on efficientNet [C]// International Conference on Control, Automation and Systems (ICCAS), Yeosu: 2023:1749-1752.
- [8] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23)[2025-07-30]. <https://arxiv.org/abs/2004.10934v1>.
- [9] XU M, YOON S, FUENTES A, et al. A comprehensive survey of image augmentation techniques for deep learning [J]. Pattern Recognition, 2023,137:109347.
- [10] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE Trans Pattern Anal Mach Intell, 2020,42(8):2011-2023.
- [11] SHAO G F, SUN Q, GAO Y L, et al. Sub-pixel convolutional neural network for image super-resolution reconstruction [J]. Electronics, 2023,12:3572.
- [12] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. et al. A comprehensive survey of loss functions and metrics in deep learning [J]. Artificial Intelligence Review, 2025,58:195.
- [13] ÇORBACIOĞLU Ş K, AKSEL G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value [J]. Turk J Emerg Med, 2023,23(4):195-198.