

基于改进 YOLOv8 的异常牙齿和修复体 X 射线影像检测

梁洪¹, 邱定乾², 丁世宇¹, 栾宽^{1*}

¹哈尔滨工程大学智能科学与工程学院, 黑龙江 哈尔滨 150001;

²哈尔滨工程大学材料科学与化学工程学院, 黑龙江 哈尔滨 150001

摘要 目前, 牙科疾病检测研究多集中于单一病症或修复体, 而同一张全景片中经常同时存在牙科疾病和修复体, 这对于现有的计算机辅助诊断方法是一个挑战。针对上述问题, 提出了一种基于 YOLOv8 框架的综合性智能辅助诊断网络——YOLOv8-Dental 模型, 其是专为异常牙齿和修复体检测设计的。首先, 为了提升骨干网络的特征提取能力, 集成了“空间分组增强”注意力机制, 强化了模型捕获复杂口腔特征的能力; 其次, 针对口腔影像中龋齿和填充物等小目标对象难以识别的问题, 添加了小目标检测层, 通过融合多层特征并维护细节信息, 提高了模型对小目标对象的识别精度; 然后, 模型采用广义交并比(GIoU)的损失函数优化边界框预测的准确性, 提升了定位性能; 最后, 采用分层自适应幅度剪枝(LAMP), 对 LAMP 评分低的连接进行剪枝, 以降低模型的计算量和参数量, 从而提高检测速度。实验结果显示, YOLOv8-Dental 模型在检测龋齿、阻生牙、种植体、根管治疗牙、充填物、牙冠及牙桥方面达到了 89.8% 的平均精度(mAP), 帧率(FPS)达到了 409 frame/s, 在辅助牙科医生进行智能诊断方面表现出潜力与优势。

关键词 深度学习; 牙科疾病; 口腔全景片; 智能诊断

中图分类号 TP391

文献标志码 A

DOI: 10.3788/CJL241265

1 引言

口腔健康是公共卫生关注的重点领域之一, 而牙科疾病如龋齿、牙周病等的高发病率凸显了其对公众健康的严重影响。中国第四次口腔健康流行病学调查报告显示, 口腔疾病广泛影响各年龄段人群, 其中 3~5 岁、35~44 岁和 55~64 岁年龄组的龋病率分别为 62.9%、89% 和 95.6%, 凸显了这一问题的普遍性和严峻性^[1]。

在临床中, 牙齿疾病导致的牙齿损伤常常需要通过修复体治疗, 例如牙冠、桥接或者种植体等。然而, 修复体并非一劳永逸, 同样可能面临并发症或失败的风险, 比如边缘适合性差、继发龋齿以及生物相容性问题等。因此, 在牙齿疾病与修复体并存的情况下, 定期的专业检查对于早期发现潜在问题至关重要, 不仅可以帮助预防进一步的损害, 还能确保修复体的效果。

在牙科诊断中, 口腔全景片是一种广泛应用的诊断工具^[2]。对于医生而言, 在日常繁重的临床工作中, 手动分析复杂的影像数据既耗时又费力, 且易受多种人为因素的干扰, 包括长时间工作导致的疲劳、情绪波动以及医生专业技能的差异。这些因素均可能对诊断

准确性产生负面影响, 甚至延误治疗, 从而对患者的健康造成不良影响^[3]。鉴于此, 迫切地需要构建一套高效的异常牙齿和修复体自动检测模型, 旨在通过智能化手段辅助识别口腔问题, 以此来减轻牙科医生的工作负担, 在加快诊断流程的同时, 提升医疗服务的质量和效率。

近年来, 卷积神经网络(CNN)^[4]在医学诊断领域中成为主流网络。根据研究目标的不同, 深度学习在牙科检测方面的应用可分为两类: 异常牙齿检测和修复体检测。在疾病检测方面, 研究人员聚焦于提高特定病症的识别精度。Jiang 等^[5]提出了 RDFNet, 即一种专为口腔全景片设计的龋齿检测网络。通过在 YOLOv5 骨干网中融入变压器结构与 FReLU 激活函数, RDFNet 不仅增强了对复杂信息的捕捉能力, 还提升了模型的运行速度, 最终实现了 56.9% 的检测精度。刘丰等^[6]使用 YOLOv5 算法对 X 光片中的龋齿和牙根尖周病变进行了检测, 检测准确率超过 95%, 这一成果证明了牙齿病变智能化诊断的可行性。李彦甫等^[7]则将研究焦点转向了利用手机等移动设备拍摄的口腔图像, 提出了 CR-PANet 龋齿检测模型, 在其测试数据集上的平均精度达到了 88.2%, 不仅为患者自

收稿日期: 2024-10-12; 修回日期: 2024-11-20; 录用日期: 2024-11-28; 网络首发日期: 2024-12-06

基金项目: 国家自然科学基金(81301297)、黑龙江省应用技术研究与开发计划(GC13A308)

通信作者: *luankuan@hrbeu.edu.cn

提供了可能,也为医生的辅助诊断提供了有力支持。鉴于患者通常携带多种口腔病变,孙召飞等^[8]提出了基于改进 YOLOv5s 的口腔识别网络 YOLO-Teeth,可识别龋齿、牙根尖周炎、根分叉病变和阻生齿,检测的平均精度达到 86.6%,充分验证了多种口腔疾病并行检测的有效性。而在修复体检测领域中,单一目标的检测精度较高。例如, Park 等^[9]利用深度学习在全景片中检测了种植体,检测准确率为 85.20%。当检测目标增多时,会出现检测精度下降的情况。例如, Oztekin 等^[10]采用 ResNext50 检测汞合金和复合填充物的平均精度为 76.7%。Bonfanti-Gris 等^[11]利用 YOLOv5s 检测牙齿、种植体和填充物,达到了 74% 的平均精度。Yüksel 等^[12]提出了基于 YOLO 模型的 DENTECT 系统,能够检测包括根尖病变、填充物、根管以及拔牙位置在内的多种口腔状况,尽管检测平均精度为 59%,但这标志着牙科检测技术向前迈出了坚实的一步。

目前大多数研究聚焦于牙科疾病和修复体各自的检测,这导致了检测类型的局限性和效率的低下。相比之下,当目标检测范围较小时,模型展现出较高的精确率。随着检测范围的扩大,检测精度下降,难以满足临床实际应用的需求。本文针对性地提出了一种改进的 YOLOv8 模型——YOLOv8-Dental,旨在全面实现龋齿、阻生牙、种植体、根管治疗牙、填充物、牙冠和牙桥等的自动检测。我们在该模型骨干网络中融入了空间分组增强(SGE)注意力机制,增设了专用于微小病灶识别的层,通过结合广义交并比(GIoU)的边界损失函数,利用分层自适应幅度剪枝(LAMP)方法去除多余的通道,有效增强了模型在复杂口腔影像中定位与区分异常牙齿和修复体的能力。

2 Yolov8-Dental 算法原理

YOLOv8 模型^[13]以其检测速度快、精度高的优势在目标检测领域中得到了广泛的关注,但在牙科诊断中仍存在检测精度低的问题。为了使 YOLOv8 适用于牙科疾病检测,我们对其主干(backbone)、颈部(neck)和头部(head)进行了改进,改进方法如图 1 所示,其中 Concat 表示对特征图进行维度叠加。首先,基于自适应图像缩放策略将输入图像尺寸统一调整为 640 pixel×640 pixel。然后,图像进入主干网络,通过卷积(Conv)层、CSP(cross stage partial bottleneck with two convolutions)和快速空间金字塔池化(SPPF)层进行卷积和特征提取。主干网络包含多层卷积操作,可能导致信息丢失,且口腔全景片中可能存在干扰和噪声,导致检测精度受限。因此,在主干网络中,在 SPPF 层尾部添加了 SGE 注意力机制,使得模型能够更高效地聚焦于口腔影像中的关键部位,减少背景噪声的影响,提高特征提取的效率。基于 SGE 注意力机制输出的信息传入颈部网络,颈部网络负责多尺度融

合,最后输出至检测头。YOLOv8 网络只有 20 pixel×20 pixel、40 pixel×40 pixel 和 80 pixel×80 pixel 尺寸的检测头,对小目标的检测效果较差。而在本文中,龋齿、根管治疗牙及填充物通常被视为小目标,因此我们在颈部网络中引入了小目标检测层,其包含上采样、下采样和拼接操作,使得模型能够利用更多的特征尺度进行检测,生成新的特征图。同时,在头部网络中添加了一个尺寸为 160 pixel×160 pixel 的检测头,用于接收新生成的特征图。四个检测头生成的特征图都需要在头部进行后处理,而 YOLOv8 头部采用解耦头设计,分为分类分支和回归分支。为了进一步提升模型在病灶区域定位上的精确度与稳定性,采用 GIoU 损失函数替代原模型中的边界框回归损失函数。通过非极大值抑制(NMS)处理最终检测结果,去除重复的检测框,保留置信度最高的边界框,确保每个目标只被检测一次。利用 LAMP 方法,删除对以上策略得出的算法模型性能贡献小的通道,提高模型的检测速度。通过上述改进,最终得到的网络模型就是 YOLOv8-Dental。整体的检测流程如下:基于自适应缩放策略将输入的图像尺寸调整至 640 pixel×640 pixel;该图像被送入网络进行检测处理,四个检测头输出不同尺寸的特征图;经过 NMS 后输出检测结果,从而实现了端到端的检测。

2.1 优化主干网络

口腔全景片在成像过程中由于技术限制,容易受到多种干扰因素的影响。首先,金属牙冠、牙桥等金属物体会产生金属伪影,而骨骼的不均匀厚度或密度也可能导致伪影的产生。此外,射线散射会引入背景噪声。值得注意的是,口腔全景片还可能包含颅骨、颌骨等头骨结构,这些结构可能与病变区域重叠,从而影响图像质量。所有这些干扰因素共同作用,增加了识别和定位病变区域的难度。此外,患者口腔内常常同时存在多种牙齿病变,这些病变的共存易导致信息混淆和特征重叠,进一步增加了检测的复杂性。

在图像检测任务中,卷积神经网络负责挖掘图像内的特征信息。其中,某些局部细节特征相较于其他部分可能蕴含更多的关键信息。为使网络能够专注于这些关键特征,避免对所有特征平均分配注意力,引入注意力机制成为必要。目前,多种注意力机制如空间、通道和卷积注意力机制已被提出,尽管这些注意力机制能够增强特征提取的针对性,但它们也往往导致模型复杂度增加和运行效率降低。为确保影像识别的实时性,引入轻量级注意力机制 SGE^[14],以提高网络对检测目标区域的关注度和识别能力,并在不增加额外计算负担的前提下,增强特征提取能力,有效抑制干扰信息。该注意力机制基于全局与局部特征间的相似性,通过分组处理特征图,生成相应的注意掩码,进而得到空间分布更优的语义特征。其结构图如图 2 所示,流程如下:

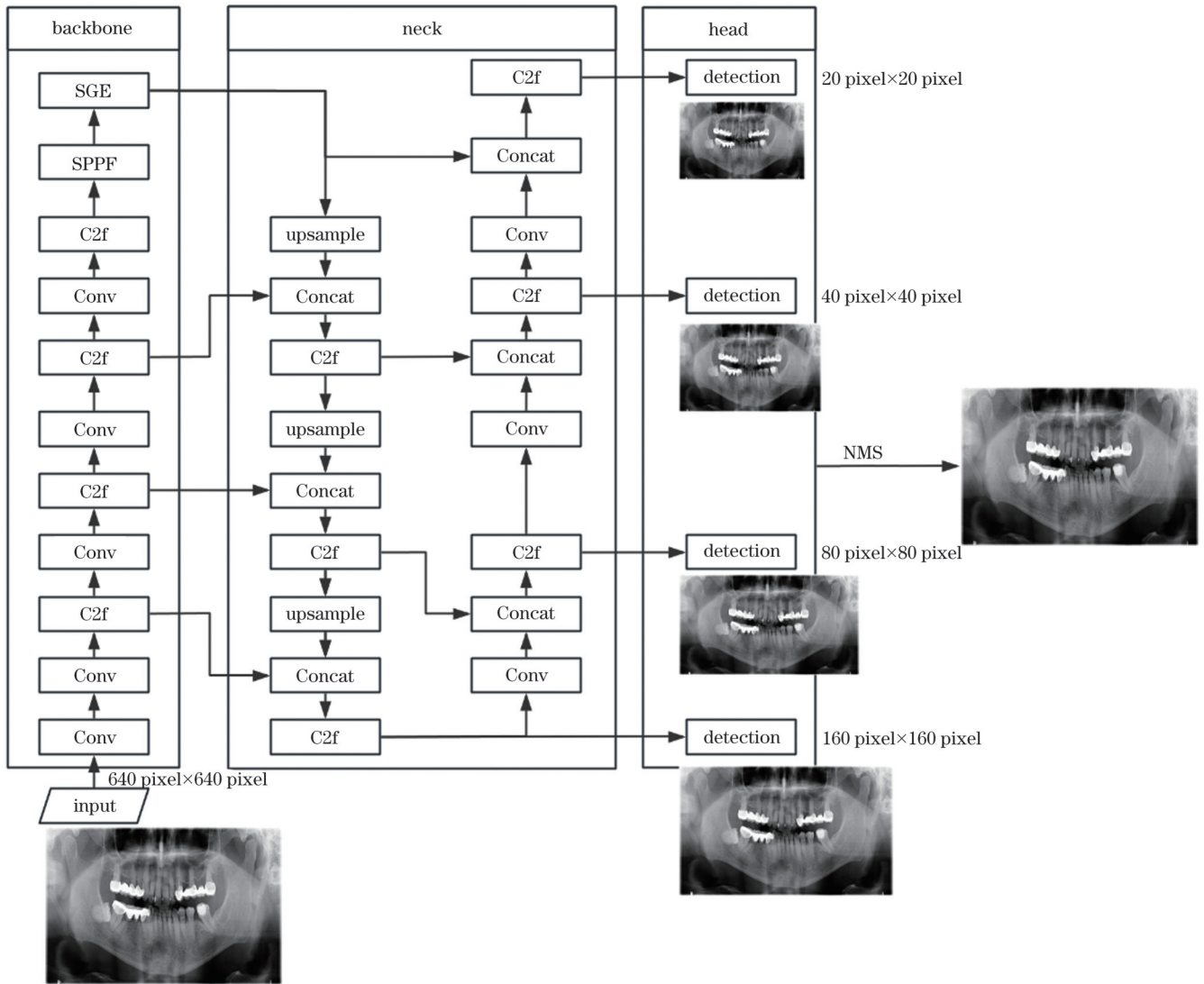


图1 YOLOv8-Dental 结构图
Fig. 1 Frame diagram of YOLOv8-Dental

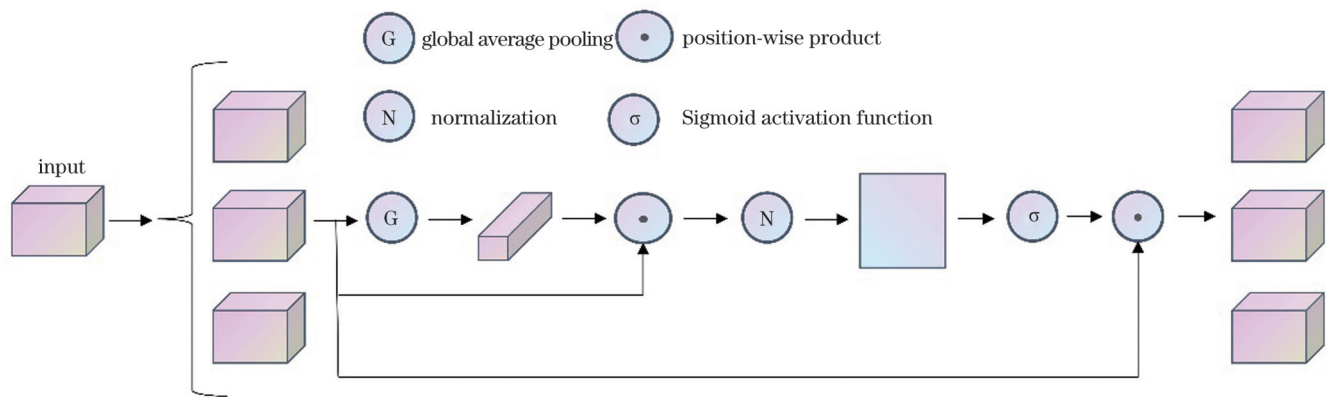


图2 轻量化SGE模块示意图
Fig. 2 Illustration of lightweight SGE module

1) 将输入特征分组,每一组并行进行操作,允许每一组在学习过程中捕获不同的语义特征。
2) 将每一组特征与其全局平均的池化特征矩阵的相应元素相乘并进行归一化处理,从而利用整个空间的无噪特性显著增强关键区域的语义学习。归一化

的目的是防止组内不同样本产生的注意力掩码在组内分布上出现差异。
3) 根据 Sigmoid 激活函数得到最终的输出特征,将其与原始特征图相乘,从而实现从原始特征到新增特征的映射。

2.2 优化检测头

在 YOLOv8 模型中,原始的检测层输出三种不同尺寸的特征图:20 pixel×20 pixel,40 pixel×40 pixel 和 80 pixel×80 pixel。尺寸较小的特征图具有较大的感受野,富含丰富的语义信息,更适用于检测大尺寸对象。而尺寸较大的特征图具有较小的感受野,包含了更多关于目标位置和局部特征的细节,因此,更适于检测小尺寸对象。在牙齿病变检测领域中,相较于口腔全景片,龋齿、根管治疗牙及填充物等通常被视为小目标。

然而,YOLOv8 模型设计中较高的下采样率导致深层特征图可能丢失小目标的细节信息,进而使得模型难以捕捉这些小目标的特征,最终影响检测性能。为解决下采样过程中小目标信息丢失的问题,以及深层特征图中小目标定位和识别的不准确性问题,本文在模型的头部增加了一个尺寸为 160 pixel×160 pixel 的小目标检测层专门用于小目标的检测。这一新增的小目标检测层能够实现更细致的目标划分,并对小目标给予更细致的关注。该层结合了浅层的位置信息与深层的语义信息,从而能够更准确地定位和识别小目标,进而有效减少了漏检现象。增加的小目标检测层如图 3 所示。

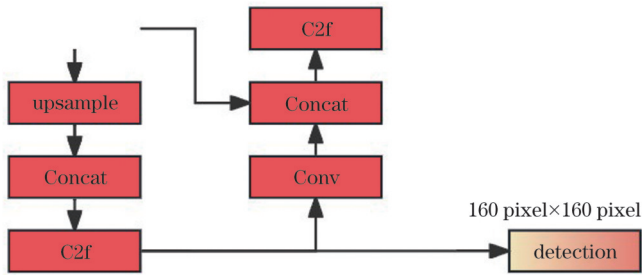


图 3 小目标检测层

Fig. 3 Small object detection layer

2.3 优化损失函数

YOLOv8 模型采用的边界框损失函数是完全交并比(CIoU)的损失函数,不过 CIoU 描述的纵横比是相对值,并且没有考虑难易样本的平衡问题^[15]。为了解决这些问题,研究中引入 GIoU 损失函数来取代 CIoU。作为交并比(IoU)的一个改良版,GIoU 保持了尺度不变性这一优势。此外,与 IoU 相比,GIoU 还额外关注了目标和预测框之间的非重叠区域,这一改进解决了 IoU 无法对真实框与预测框之间非重叠部分进行量化的问题。IoU 和 GIoU 的表达式分别为

$$f_{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

$$f_{GIoU} = f_{IoU} - \frac{|C \setminus (A \cup B)|}{|C|} \quad (2)$$

式中: A 和 B 为两个检测框; C 为包围 A 、 B 的最小外接

矩形;符号“\”表示去掉。 f_{IoU} 的取值区间为 $[0, 1]$,而 f_{GIoU} 的取值区间为 $[-1, 1]$ 。GIoU 示例图如图 4 所示,其中 D 为矩形 C 中不含 A 、 B 的部分。

GIoU 损失函数的表达式为

$$L_{GIoU} = 1 - f_{GIoU} \quad (3)$$

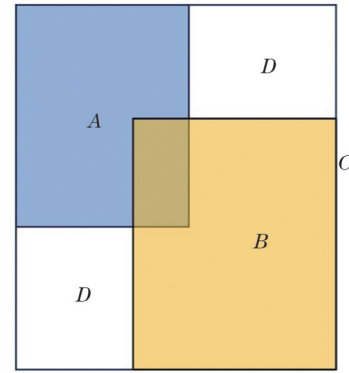


图 4 GIoU 示例图

Fig. 4 GIoU example diagram

2.4 模型剪枝

深度学习技术尽管已展现出卓越能力,但其对计算资源和内存的高需求成为部署的一大障碍。因此,剪枝算法逐渐受到研究者的关注与青睐。剪枝算法的主要目的是精简神经网络结构,去除资源占用高但对性能贡献小的部分。LAMP 剪枝算法是一种基于分层自适应幅度的剪枝方法^[16]。幅度是权重的大小或绝对值^[17]。在神经网络中每个连接都分配有一个权重,该权重决定了输入信号的重要性。评分是剪枝过程中决定是否保留通道结构的依据。目前广泛使用的评价标准是权重的绝对值大小,需要修剪的部分通常具有较小的权重,这些部分被认为对函数性能没有显著贡献^[18-19]。具体的得分公式为

$$S(u; W) = \frac{[W(u)]^2}{\sum_{v \geq u} [W(v)]^2} \quad (4)$$

式中: W 为卷积核展开的一维向量的权重; $W(u)$ 和 $W(v)$ 分别为索引 u 和 v 映射的权重项。

LAMP 算法的核心步骤概括如下:1) 获取模型训练产生的权重参数文件,并执行初始化配置;2) 对各连接权重的幅值进行平方运算,并通过归一化处理,计算得出 LAMP 评分;3) 基于所得评分,识别出评分较低的连接,并依据预定的全局稀疏性指标,挑选相应数量的连接实施剪枝操作;4) 执行剪枝动作,即把选定的连接权重设为零,在模型架构中进行实质移除;5) 对经过剪枝的模型实施回炉训练,旨在弥补剪枝可能导致的性能损耗;6) 对回炉后的模型性能进行综合评估。剪枝方式如图 5 所示,其中 ϕ 为空集,表示该通道被剪枝, u_1, u_2, u_3 为一层的通道权重(降序排列), v_1, v_2, v_3 为另一层的通道权重(降序排列)。

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ \frac{u_2^2}{u_1^2 + u_2^2} \\ \frac{u_3^2}{u_1^2 + u_2^2 + u_3^2} \\ \vdots \end{bmatrix} \begin{bmatrix} 1 \\ \frac{v_2^2}{v_1^2 + v_2^2} \\ \frac{v_3^2}{v_1^2 + v_2^2 + v_3^2} \\ \vdots \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 \\ \frac{u_2^2}{u_1^2 + u_2^2} & \phi \\ \phi & \phi \\ \vdots & \vdots \end{bmatrix}$$

图 5 LAMP 得分流程

Fig. 5 LAMP scoring process

3 实验结果分析

3.1 口腔全景片病症数据集的制作

本文的数据集来源于 DENTEX2023 挑战赛^[20], 选用 1000 张无变形、显示完全且符合研究需求的高质量影像, 并在专业医师的指导下使用 LabelImg 图像标注工具对其进行了重新标注, 确保每张影像的标签准确且符合口腔诊断标准。示例图如图 6 所示。将数据集的 1000 张图像按 8:2 的比例划分为训练集和测试集。训练集和测试集中每种标签的数量如表 1 所示。在处理数据集时, 未使用数据增强技术。由于部分图像同时包含多种研究对象, 且某些研究对象在一张图像中可能存在多个实例, 因此会出现训练集和测试集中标签数量不同的情况。

本文选取龋齿、阻生牙、种植体、根管治疗牙、填充物、牙冠和牙桥作为研究对象, 并确定了它们对应的标签: caries(龋齿)、impacted(阻生牙)、implant(种植体)、RCTT(根管治疗牙)、filling(填充物)、crown(牙冠)和 bridge(牙桥)。为了直观展示这些研究对象, 图 7 给出了它们的示例图。具体而言, 龋齿是牙齿表面细菌产生的酸性物质溶解牙齿表面的矿物质而引起

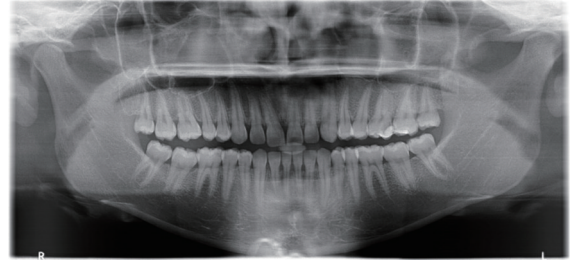


图 6 口腔全景片示意图

Fig. 6 Example of panoramic oral film

表 1 标签数量
Table 1 Numbers of tags

Label type	Train set	Test set
Caries	477	107
Impacted	584	166
Implant	42	20
RCTT	1336	183
Filling	1515	235
Crown	189	47
Bridge	395	152

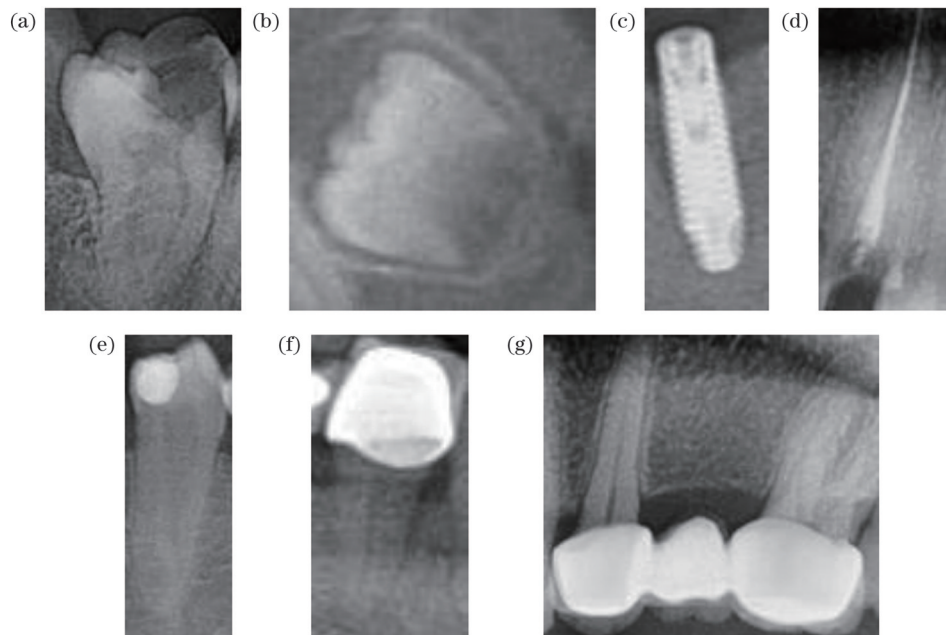


图 7 异常牙齿和修复体。(a)龋齿;(b)阻生齿;(c)种植体;(d)根管治疗牙;(e)填充物;(f)牙冠;(g)牙桥

Fig. 7 Abnormal teeth and restorations. (a) Dental caries; (b) impacted teeth; (c) implant; (d) root canal treated teeth; (e) filling; (f) dental crown; (g) dental bridge

的,在口腔全景片中表现为牙齿上有不规则深色块;阻生齿是指因空间不足或位置异常而无法完全正常生长的牙齿,在口腔全景片中呈现为斜向或横向生长的牙齿;种植体是通过外科手术植入颌骨中的人工牙根,用于替代缺失的牙齿,在口腔全景片中表现为螺丝状的牙根;根管治疗是一种治疗牙髓感染或根尖周炎等炎症的方法,即清除受感染的牙髓并用填充物密封根管,在口腔全景片中,根管治疗牙通常表现为牙齿根部有较明显的填充影像;填充物是用于修复牙齿表面受损或龋坏部位的材料,旨在恢复牙齿的形态和功能,在口腔全景片中表现为牙齿表面不规则的白色物料;牙冠是覆盖在牙齿表面的人工修复体,旨在恢复牙齿的外观和功能,在口腔全景片中表现为覆盖牙齿表面的规则结构;牙桥是用于连接多个缺失牙齿的人工修复体,在口腔全景片中表现为多个牙冠连接在一起的结构。

3.2 实验配置与评估标准

本文采用 Ubuntu 20.04 作为实验的操作系统,中央处理器为 Intel (R) Core (TM) i7-7800X,规格为 3.50 GHz。图形处理单元选用 NVIDIA GeForce RTX 2080 Ti,配备 11 GB 显存。编程语言环境为 Python 3.9.1 版本。此外,实验采用 PyTorch 深度学习框架,版本号为 1.13.1,并搭配 CUDA 11.7 版本以优化计算性能。在训练过程中,优化器选择了 AdamW,学习率设定为 0.001,学习率衰减权重为 0.01,批次大小 (batch size) 为 16,训练总轮次为 300,耐心值为 50,网络输入图像尺寸为 640 pixel × 640 pixel,未加载预训练权重。

模型的检测精度采用准确率 (P)、召回率 (R)、单类标签精度 (AP) 和所有类标签的平均精度 (mAP) 来衡量。这些指标是通过使用真阳性 (TP, 个数为 N_{TP})、假阳性 (FP, 个数为 N_{FP})、假阴性 (FN, 个数为 N_{FN}) 和真阴性 (TN, 个数为 N_{TN}) 四个参数进行计算。而模型的计算能力和推理速度采用参数量 (Params)、计算量 (FLOPs)、帧率 (FPS) 来评价。其中参数量和计算量是通过网络可视化工具 TorchSummary 进行计算的。

P 表示模型判断为阳性样本时预测框中正确预测的比例,其计算公式为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (5)$$

R 表示模型在样本中成功检测到的所有对象的比例,其计算公式为

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (6)$$

单类标签精度是用于度量模型在某个类别上的精度,其值是 P - R 曲线与坐标轴围成的面积,计算公式为

$$f_{AP} = \int_0^1 P(R) dR \quad (7)$$

mAP 是用于度量所有类别的平均精度,其值是所有类别 AP 的平均值,计算公式为

$$f_{mAP} = \frac{1}{n} \sum_{i=0}^n f_{AP}(i) \quad (8)$$

式中: i 为类别序号; n 为总类别数。

FPS 代表 1 s 内所能处理的图片数量,其计算公式为

$$f_{FPS} = \frac{1000}{t_p + t_i + t_{NMS}} \quad (9)$$

式中: t_p 为图像预处理时间; t_i 为推理时间; t_{NMS} 为后处理时间。

3.3 注意力机制对比实验

SGE 注意力机制通过特征分组来提高网络对检测区域的关注度,在不增加计算量的情况下增强特征提取能力并抑制干扰信息。为了验证 SGE 注意力机制的有效性,增加了不同注意力机制的模型训练,实验结果如表 2 所示。

表 2 注意力机制对比实验结果

Table 2 Comparative experiment results of attention mechanisms

Model	P / %	R / %	mAP / %	Params / 10^6	FLOPs / 10^9
Baseline	78.7	83.2	84.6	3.00	8.1
Baseline+DA	81.2	63.6	74.3	3.13	8.2
Baseline+EMA	82.5	77.6	84.1	3.02	8.2
Baseline+CA	79.4	82.4	84.0	3.01	8.1
Baseline+Triplet	80.9	80.6	85.0	3.01	8.1
Baseline+SGE	81.1	85.3	86.5	3.01	8.1

本文在保证其他条件相同的情况下,进行了不同注意力机制的对比实验。从表 2 可以看出,通过引入双重注意力 (DA)^[21]、指数移动平均 (EMA)^[22]、坐标注意力 (CA)^[23]、三重注意力 (Triplet)^[24] 和 SGE 等注意力机制,与基线算法模型对比,检测精度有升有降,证明了实验的有效性。与 DA、CA 和 Triplet 注意力机制相比,SGE 注意力机制在准确率、召回率和 mAP 方面均有所提升,且其参数量和计算量并未显著增加。与 EMA 注意力机制相比,SGE 注意力机制的准确率虽然低了 1.4 个百分点,但其召回率和 mAP 分别提升了 7.7 个百分点和 2.4 个百分点。综上所述,SGE 注意力机制对模型性能的提升效果最为显著,将模型的准确度提升至 81.1%,召回率提升至 85.3%,mAP 提升至 86.5%,同时并未增加模型的参数量和计算量。因此选择引入 SGE 注意力机制。

3.4 剪枝实验

通过网络结构的优化,提升了模型检测异常牙齿

和修复体的精度,但是增加了模型的参数量和计算量,降低了模型的运行速度。为加快检测的速度,对优化后的网络实施 LAMP。在启动剪枝步骤之前,需要先进行跳层操作,以规避网络中不可削减的结构,确保剪枝工作的顺利进行。

主干网络的输出层主要负责对已检测到的目标进行分类。如果直接进行剪枝,可能会导致分类信息的丢失。因此,我们选择跳过这一层,以保留完整的分类能力。对于 SGE 注意力模块来说,它的参数量相对较小,而且在其内部结构中存在分组操作,这种操作在模型训练中扮演着关键的角色。由于其复杂性和高度关联性,SGE 注意力模块不适合进行剪枝处理。因此,我们同样选择跳过这一层,以确保模型性能的稳定性。

在完成前期准备工作后,通过调整加速比($f_{\text{Speed-up}}$)参数来执行不同程度的网络剪枝操作。表 3 详细展示了剪枝操作对模型性能的具体影响。 $f_{\text{Speed-up}}$ 定义为原始模型计算量与剪枝后模型计算量的比,当其值为 1 时,表明模型未经过任何剪枝处理。

根据表 3,我们可以观察到以下情况:当 $f_{\text{Speed-up}}$ 设置为 1.3 时,相较于 $f_{\text{Speed-up}}$ 为 1.0 时,虽然 mAP 有所提高,但参数量和计算量都有所减少,这符合我们的需求。

表 3 不同加速比下的剪枝效果

Table 3 Effects of pruning under different acceleration ratios

$f_{\text{Speed-up}}$	mAP / %	Params / 10^6	FLOPs / 10^9
1.0	88.8	2.93	12.3
1.3	88.9	1.32	9.4
1.5	87.7	1.08	8.2
1.7	89.8	0.91	7.2
2.0	87.6	0.72	6.1

然而,当 $f_{\text{Speed-up}}$ 提高到 1.7 时,相较于 $f_{\text{Speed-up}}$ 为 1.0 时,mAP 提升了 1 百分点,同时参数量和计算量的下降更为显著。相比之下,当 $f_{\text{Speed-up}}$ 为 1.5 和 2.0 时,相较于 $f_{\text{Speed-up}}$ 为 1.0 时,虽然参数量和计算量有所降低,但 mAP 的下降更为严重。综合考虑,最终选择 $f_{\text{Speed-up}}$ 为 1.7 时的剪枝操作。图 8 是卷积层通道对比图,横轴表示模型的不同卷积层,纵轴表示每层的通道数量。其中,base 表示剪枝前,prune 表示剪枝后,M 代表模型,M-1 代表模型的第一个模块。某些模块含有多个卷积层,例如 M-3-1 代表模型的第三个模块的第一个卷积层。由图 8 可以看出,剪枝后模型的通道数量明显少于剪枝前,模型的参数量和计算量降低,说明该剪枝算法是有效的。

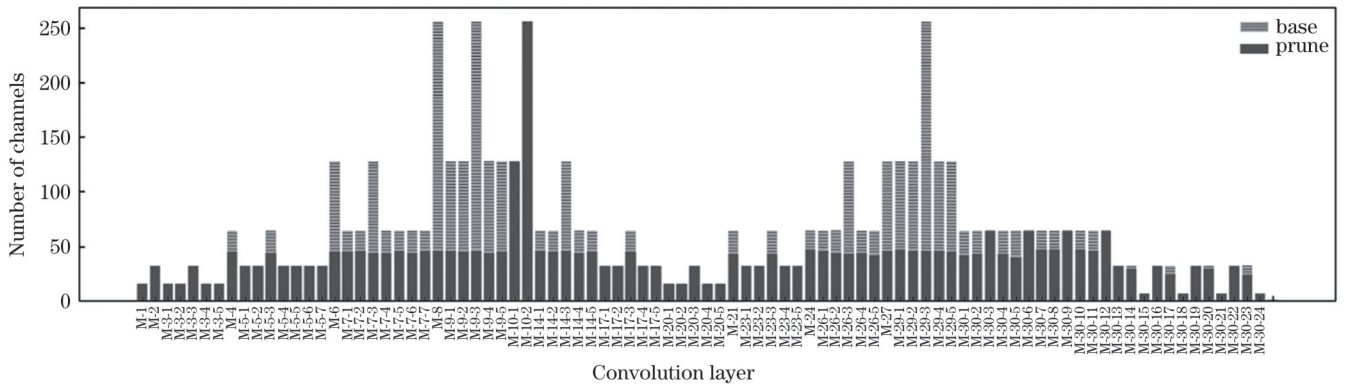


图 8 通道对比图

Fig. 8 Channel contrast diagram

3.5 YOLOv8-Dental 模型消融实验

为了分析 SGE、小目标检测层、GIoU 和 LAMP 改进模块的有效性,以 YOLOv8 模型为基线,在保证其

他条件相同的情况下,进行消融实验。实验结果如表 4 所示。

可以看出,在基线模型中引入 SGE 注意力机制

表 4 消融实验结果

Table 4 Ablation experiment results

Model	P / %	R / %	mAP / %	Params / 10^6	FLOPs / 10^9	FPS / (frame/s)
Baseline	78.7	83.2	84.6	3.01	8.1	501
Baseline + SGE	81.1	85.8	85.6	3.01	8.1	320
Baseline + P2	81.7	85.6	86.7	2.93	12.3	384
Baseline + GIoU	82.3	82.5	85.8	3.01	8.1	330
Baseline + SGE + P2	82.5	83.1	87.7	2.93	12.3	278
Baseline + SGE + P2 + GIoU	85.0	83.4	88.8	2.93	12.3	240
Baseline + SGE + P2 + GIoU + LAMP (YOLOv8-Dental)	83.9	87.8	89.8	0.91	7.2	409

后,模型的准确率、召回率和 mAP 分别提升了 2.4 百分点、2.6 百分点和 1.0 百分点。这充分证明了 SGE 注意力机制具有高效的特征分组能力,能够显著提升模型对不同特征的识别精度,进而有效实现特征的高效提取,并有效抑制无关信息的干扰。在基线模型中增加小目标检测层,模型的准确率、召回率和 mAP 分别提升了 3.0 百分点、2.4 百分点和 2.1 百分点。这表明小目标检测层能够有效地提升模型检测小目标的能力。将损失函数从 CIoU 损失函数更改为 GIoU 损失函数后,尽管召回率下降了 0.7 百分点,但模型的准确率和 mAP 却分别提升了 3.6 百分点和 1.2 百分点。这表明 GIoU 损失函数在提升模型定位精度和识别准确度方面更具优势。将 SGE 注意力机制与小目标检测层同时集成到基线模型中,尽管召回率下降了 0.1 百分点,但模型的准确率和 mAP 却分别实现了 3.8 百分点和 3.1 百分点的显著提升。更重要的是,与单独引入 SGE 注意力机制或小目标检测层相比,这种组合方式在提升模型准确率和 mAP 方面表现更佳。随后,将

SGE 注意力机制、小目标检测层以及 GIoU 损失函数同时融入基线模型,模型的准确率、召回率和 mAP 分别实现了 6.3 百分点、0.2 百分点和 4.2 百分点的显著提升,然而计算量却增加了 4.2×10^9 。为了优化模型性能,我们采用了 LAMP 技术以降低计算量并提高运行速度。本文最终提出的模型 YOLOv8-Dental 的准确率、召回率和 mAP 分别提升了 5.2 百分点、4.6 百分点和 5.2 百分点,参数量和计算量分别下降了 2.1×10^6 和 0.9×10^9 ,FPS 也达到了 409 frame/s。这表明 YOLOv8-Dental 模型具有更好的异常牙齿和修复体检测精度和较快的运行速度,能够帮助医疗人员在短时间内进行更专业的判断。

3.6 YOLOv8-Dental 模型对比实验

为了验证本文所提 YOLOv8-Dental 模型的检测性能,选择与目前主流的目标检测模型 Faster R-CNN^[25]、YOLOv3^[26]、YOLOv5、YOLOvX^[27]、YOLOv6^[28]和 YOLOv7^[29]在相同的实验环境和参数的情况下进行训练和测试。实验结果如表 5 所示。

表 5 不同模型对比实验结果
Table 5 Comparative experiment results of different models

Model	AP / %							mAP / %	Params / 10^6	FLOPs / 10^9
	Caries	Impacted	Implant	RCTT	Filling	Crown	Bridge			
Faster R-CNN	17.3	78.8	85.6	73.2	90.0	81.8	78.1	72.1	41.38	184.32
YOLOv3	33.0	83.5	95.9	57.3	85.0	90.3	97.4	77.5	12.13	18.90
YOLOv5	56.3	86.9	96.2	73.7	88.9	95.6	97.3	85.0	2.50	7.10
YOLOv6	38.1	86.3	95.6	73.2	88.3	91.7	94.8	81.1	4.23	11.80
YOLOv7	55.2	87.5	89.2	76.0	90.2	96.0	92.4	83.8	6.02	13.10
YOLOvX	45.5	79.5	86.4	74.5	81.7	83.1	93.4	77.7	5.04	7.58
YOLOv8	62.3	83.5	96.7	72.7	90.7	89.9	96.1	84.6	3.01	8.10
YOLOv8-Dental	81.5	86.4	95.1	80.7	93.0	95.4	96.2	89.8	0.91	7.20

从对比分析可以看出,虽然 YOLOv8-Dental 在种植体和牙桥检测方面的表现略逊于 YOLOv3、YOLOv5、YOLOv6 和 YOLOv8 模型,但其检测精度仍然分别达到了 95.1% 和 96.2%。在龋齿、阻生牙、根管治疗牙、填充物和牙冠的检测方面, YOLOv8-Dental 的检测精度超过了其他模型。此外, YOLOv8-Dental 模型的参数量和计算量都低于其他模型,这说明 YOLOv8-Dental 模型不仅保持了高效性,而且在多种研究对象的检测任务中,仍能保持较高的检测精度。这样的表现不仅确保了整体检测的精度,也充分证明了 YOLOv8-Dental 模型改进的有效性。

3.7 YOLOv8-Dental 模型特征图可视化

为更清晰看到 YOLOv8-Dental 与 YOLOv8 的检测效果对比,随机选取了测试集中的三张图像 A、B、C,将其输入模型进行检测,并输出检测结果图像和热

力图。热力图能够直观地展示模型对输入图像中不同区域的关注程度,从而帮助医生快速定位图像中的关键区域或感兴趣的部分。如图 9 所示,可以看出 YOLOv8-Dental 可以检测到 YOLOv8 未识别出来的异常牙齿和修复体,并且检测精度更高。

4 结 论

为了探索基于深度学习的牙科口腔全景片人工智能辅助诊断技术,有效缓解口腔医学领域的医疗压力,辅助牙医突破主观判断的限制,提高诊断过程的精确度,首先,采用 YOLOv8 作为基础网络,通过在其骨干特征提取网络中引入空间分组增强注意力机制,增强了网络对关键特征的提取和识别能力;其次,针对口腔图像中的小目标检测难题,添加了小目标检测层,这一改进提升了网络对小目标对象的识别精度;然后,为了进一步优化模型的边界框定位准确性,采用了 GIoU

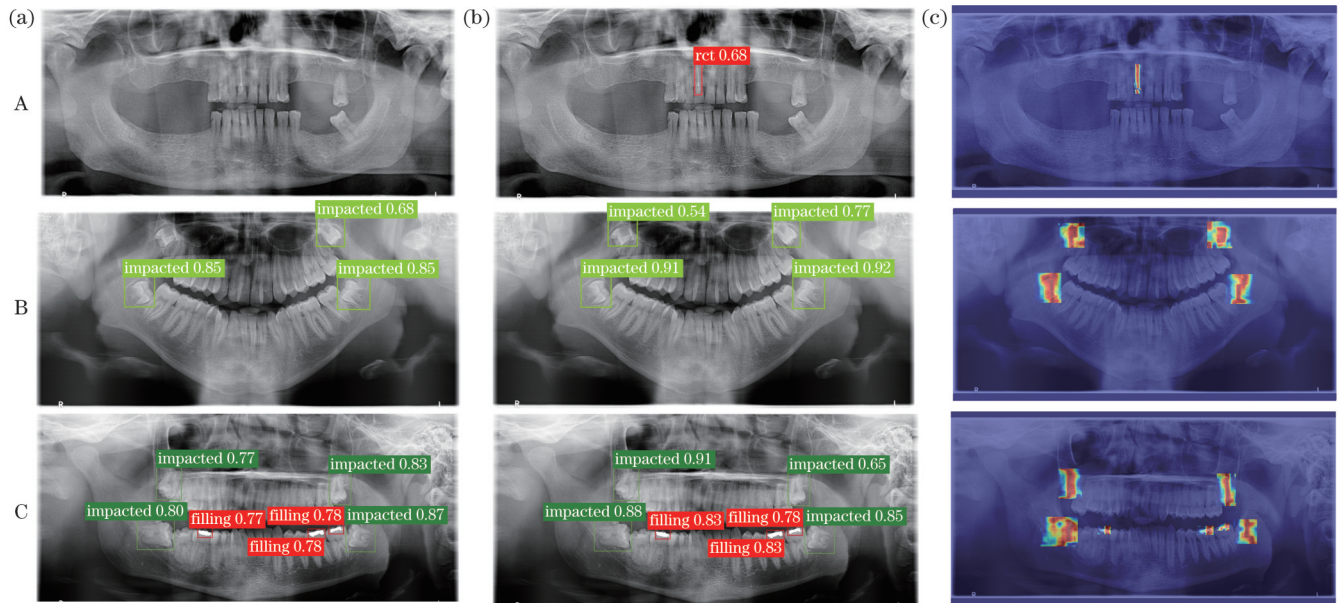


图9 预测结果对比图。(a)YOLOv8检测结果;(b)YOLOv8-Dental检测结果;(c)YOLOv8-Dental检测结果热力图

Fig. 9 Comparison of predicted results. (a) YOLOv8 test results; (b) YOLOv8-Dental test results; (c) heat maps of YOLOv8-Dental test results

边界损失函数来替代原有的损失函数,这一改进使得网络模型的边界框回归性能得到了提升;最后,对模型进行LAMP处理,降低了参数量和计算量,提升了检测速度。将所有优化策略进行整合,建立了YOLOv8-Dental辅助诊断模型。在此基础上进行的对比实验和消融实验表明各优化策略为辅助诊断模型都带来了正增益。实验结果显示,YOLOv8-Dental模型检测龋齿、阻生牙、种植体、根管治疗牙、充填物、牙冠和牙桥的准确率为83.9%,召回率为87.8%,mAP值为89.8%,FPS也达到了409 frame/s。通过检测结果图像和热力图验证了模型的有效性。研究结果可为基于深度学习的口腔医学临床辅助诊断提供理论指导,有效推动了基于深度学习的牙科疾病影像辅助智能诊断的研究。

参 考 文 献

- [1] 王兴,冯希平,李志新.第四次全国口腔健康流行病学调查报告[M].北京:人民卫生出版社,2018:1-34.
Wang X, Feng X P, Li Z X. The fourth national oral health epidemiological survey report[M]. Beijing: People's Medical Publishing House, 2018: 1-34.
- [2] 王曼玲,孙兴龙,王伦昌.数字化全景片在口腔治疗中的应用价值[J].中外医学研究,2013,11(8):61-63.
Wang M L, Sun X L, Wang L C. The clinical application of panoramic radiography(PR) in dental treatment[J]. Chinese and Foreign Medical Research, 2013, 11(8): 61-63.
- [3] Avramova N T. Dental fear, anxiety, and phobia - behavioral management and implications for dentists[J]. Journal of Mind and Medical Sciences, 2023, 10(1): 42-50.
- [4] Zou Z X, Chen K Y, Shi Z W, et al. Object detection in 20 years: a survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.
- [5] Jiang H, Zhang P L, Che C, et al. RDFNet: a fast caries detection method incorporating transformer mechanism[J]. Computational and Mathematical Methods in Medicine, 2021, 2021(1): 6752-6760.
- [6] 刘丰,韩民,万军,等.基于深度学习的牙齿病变自动检测算法[J].中国激光,2022,49(20):2007207.
Liu F, Han M, Wan J, et al. Automatic detection of dental lesions based on deep learning[J]. Chinese Journal of Lasers, 2022, 49(20): 2007207.
- [7] 李彦甫,兰海月,薛婧帆,等.基于多尺度路径聚合的儿童龋齿检测算法[J].中国激光,2024,51(15):1507207.
Li Y F, Lan H Y, Xue J F, et al. Child caries detection algorithm based on multi-scale path aggregation[J]. Chinese Journal of Lasers, 2024, 51(15): 1507207.
- [8] 孙召飞,俞经虎,朱行飞,等.基于改进YOLOv5s的口腔全景片牙齿病症识别算法[J].中国激光,2024,51(15):1507106.
Sun Z F, Yu J H, Zhu X F, et al. Dental-disease-recognition algorithm of panoramic oral radiograph based on improved YOLOv5s[J]. Chinese Journal of Lasers, 2024, 51(15): 1507106.
- [9] Park J H, Moon H S, Jung H I, et al. Deep learning and clustering approaches for dental implant size classification based on periapical radiographs[J]. Scientific Reports, 2023, 13: 16856.
- [10] Oztekin F, Katar O, Sadak F, et al. Automatic semantic segmentation for dental restorations in panoramic radiography images using U-Net model[J]. International Journal of Imaging Systems and Technology, 2022, 32(6): 1990-2001.
- [11] Bonfanti-Gris M, Herrera A, Paraiso-Medina S, et al. Performance evaluation of three versions of a convolutional neural network for object detection and segmentation using a multiclass and reduced panoramic radiograph dataset[J]. Journal of Dentistry, 2024, 144: 104891.
- [12] Yüksel A E, Gültekin S, Simsar E, et al. Dental enumeration and multiple treatment detection on panoramic X-rays using deep learning[J]. Scientific Reports, 2021, 11: 12342.
- [13] Lou H T, Duan X H, Guo J M, et al. DC-YOLOv8: small-size object detection algorithm based on camera sensor[J]. Electronics, 2023, 12(10): 2323.
- [14] Li X, Hu X L, Yang J. Spatial group-wise enhance: improving semantic feature learning in convolutional networks[EB/OL]. (2019-05-23)[2024-08-9]. <https://arxiv.org/abs/1905.09646v2>.
- [15] Zhang H Y, Wang Y, Dayoub F, et al. VarifocalNet: an IoU-aware dense object detector[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 8510-8519.
- [16] Lee J, Park S, Mo S, et al. Layer-adaptive sparsity for the

- magnitude-based pruning[EB/OL]. (2020-10-15) [2024-08-09]. <https://arxiv.org/abs/2010.07611>.
- [17] Li H, Kadav A, Durdanovic I, et al. Pruning filters for efficient ConvNets[EB/OL]. (2016-08-31)[2024-08-08]. <https://arxiv.org/abs/1608.08710>.
- [18] Blalock D, Ortiz J J G, Frankle J, et al. What is the state of neural network pruning? [EB/OL]. (2020-03-06) [2024-08-09]. <https://arxiv.org/abs/2003.03033v1>.
- [19] Molchanov P, Tyree S, Karras T, et al. Pruning convolutional neural networks for resource efficient transfer learning[EB/OL]. (2016-11-19)[2024-08-09]. <https://arxiv.org/abs/1611.06440>.
- [20] Hamamci I E, Er S, Simsar E, et al. DENTEX: an abnormal tooth detection with dental enumeration and diagnosis benchmark for panoramic X-rays[EB/OL]. (2023-05-30)[2024-08-09]. <https://arxiv.org/abs/2305.19112v1>.
- [21] Chen Y P, Li J S, Yan S C, et al. A²-Nets: double attention networks[EB/OL]. (2018-10-27) [2024-08-09]. <https://arxiv.org/abs/1810.11579>.
- [22] Ouyang D L, He S, Zhang G Z, et al. Efficient multi-scale attention module with cross-spatial learning[EB/OL]. (2023-05-23) [2024-08-09]. <https://arxiv.org/abs/2305.13563v2>.
- [23] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20–25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.
- [24] Misra D, Nalamada T, Arasanipalai A U, et al. Rotate to attend: convolutional triplet attention module[EB/OL]. (2020-10-06)[2024-08-09]. <https://arxiv.org/abs/2010.03045v2>.
- [25] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [26] Redmon J, Farhadi A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08) [2024-08-09]. <https://arxiv.org/abs/1804.02767v1>.
- [27] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-07-18) [2024-08-09]. <https://arxiv.org/abs/2107.08430>.
- [28] Li C Y, Li L, Jiang H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07)[2024-08-09]. <https://arxiv.org/abs/2209.02976>.
- [29] Wang C Y, Bochkovskiy A, Liao H Y M. 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), June 17–24, 2023, Vancouver, BC, Canada. New York: IEEE Press, 2023: 7464-7475.

X-Ray Imaging Detection of Abnormal Teeth and Restorations Based on Improved YOLOv8

Liang Hong¹, Qiu Dingqian², Ding Shiyu¹, Luan Kuan^{1*}

¹*School of Intelligent Science and Engineering, Harbin Engineering University, Harbin 150001, Heilongjiang, China;*

²*School of Materials Science and Chemical Engineering, Harbin Engineering University, Harbin 150001, Heilongjiang, China*

Abstract

Objective In contemporary society, dental diseases affect people of all ages, increasing the workload of dentists. Oral panoramic imaging is a widely used diagnostic tool in dentistry, and doctors must process image data from numerous patients amid their heavy daily clinical workload. However, manually analyzing complex image data is time-consuming, laborious, and susceptible to various human factors, such as fatigue, emotional fluctuations, and differences in professional skills. These factors can adversely affect diagnostic accuracy, delay treatment, and damage patient health. Although artificial intelligence (AI) is initially applied in dental disease detection, most current AI research focuses on single disease or restoration. However, when the number of detection targets increases, the decrease in detection accuracy can hinder practical clinical applications. Therefore, this study applies deep learning to identify key image features for efficient and accurate lesion screening in oral panoramic images using a deep learning network architecture. The purpose is to detect abnormal teeth and restorations including dental cavities, blocked teeth, implants, root canal treated teeth, fillings, crowns, and bridges. Intelligent assistance methods can be used to reduce human errors, accelerate diagnoses, and improve medical quality and efficiency.

Methods This study proposes an intelligent assisted diagnostic network based on the YOLOv8 framework, designing a YOLOv8 model specifically for dental imaging. The purpose is to detect abnormal teeth and restorations including cavities, residual teeth, implants, root canal treated teeth, fillings, crowns, and bridges. Intelligent assistance methods can be used to reduce human error, accelerate diagnosis, and improve medical quality and efficiency. First, to enhance feature extraction capability, we integrated a spatial grouping enhancement (SGE) attention mechanism to enhance the model ability to capture complex oral features. In addition, to address the difficulty of identifying small lesions, a small-object detection layer was added. This layer integrates multiple features and maintains detailed information, thereby enhancing the capability of the model in detecting fine lesions. Subsequently, the model

loss function was optimized, adopting the generalized intersection over union (GIoU) loss function to improve the prediction accuracy of bounding box, which further enhanced localization performance. Finally, to reduce the computational burden of improved model, the layer-adaptive magnitude-based pruning (LAMP) method was used. This method eliminates non-contributing channels and improves detection speed.

Results and Discussions The analysis in Table 2 shows that the SGE attention mechanism performs well in target recognition, outperforming other attention mechanisms in all detection results. Table 4 shows the results of the ablation experiment, indicating that integrating the SGE attention mechanism into the baseline model improves accuracy, recall, and mean average precision (mAP) by 2.4, 2.6, and 1.0 percentage points, respectively. This indicates that the SGE attention mechanism can effectively group features, improve recognition rate, enhance feature extraction, and suppress information interference. After the addition of small-object detection layer, accuracy, recall, and mAP increased by 3.0, 2.4, and 2.1 percentage points, respectively, indicating that the small-object detection layer effectively identifies smaller detection targets and enhances the network ability to recognize small objects. After replacing completing intersection over union (CIoU) with GIoU, the accuracy and mAP increased by 3.6 and 1.2 percentage points, respectively; however, the recall rate decreased by 0.7 percentage points. This indicates that GIoU enhances localization performance and improves recognition accuracy. The final model, YOLOv8-Dental, was developed using the LAMP method, which improved accuracy, recall, and mAP by 5.2, 4.6, and 5.2 percentage points, respectively, while reducing parameters and computational complexity by 2.02×10^6 and 0.9×10^9 , respectively. Table 5 shows the comparative experiments, indicating that although YOLOv8-Dental performed slightly worse than some models in terms of implants and dental bridges, it still achieved recognition rates of 95.1% and 96.2% for these, respectively. In detecting the remaining five lesions, the proposed model outperformed the other models in average precision (AP) with fewer parameters and a lower computational workload. This ensures high detection accuracy for multiple lesions and maintains the overall detection rate.

Conclusions This study explored the deep learning-based AI-assisted diagnosis of dental panoramas, aiming to reduce the healthcare burden of dentistry, assist dentists beyond the limitations of subjective judgment, and improve diagnostic accuracy. First, YOLOv8 was used as the base network, which was enhanced by integrating the SGE attention mechanism into its backbone feature extraction network. Second, to detect small target lesions in oral images, a small target detection layer was added to improve recognition accuracy. To further enhance the model bounding box localization accuracy, the GIoU loss function was adopted, which significantly improved the network bounding box regression performance. Finally, the model was pruned using the LAMP method to reduce the number of parameters and computation, thereby improving detection speed. All these optimization strategies were integrated to build the YOLOv8-Dental-assisted diagnosis model. Comparisons and ablation experiments demonstrated the positive impact of each optimization strategy on the diagnosis model. The experimental results showed that the YOLOv8-Dental model achieved a precision rate of 83.9%, recall rate of 87.8%, mAP of 89.8%, and frame rate of 409 frame/s for detecting cavities, blocked teeth, implants, root canal treated teeth, fillings, crowns, and bridges. The validity of the model was verified through physical image detection and heatmap analysis. The results of this study provide theoretical guidance and a methodological reference for deep-learning-based clinical diagnosis, promoting the research on deep-learning-based image-assisted diagnosis of dental diseases.

Key words deep learning; dental diseases; panoramic oral radiograph; intelligent diagnosis