Modeling and Simulation 建模与仿真, 2025, 14(6), 370-382
Published Online June 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.146504

Hans汉斯

基于YOLOv8改进的牙齿X光医学图像分割模型

杜一族，张学典*

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2025年5月26日；录用日期：2025年6月19日；发布日期：2025年6月26日

摘　要

医学牙齿图像对于帮助医生根据图像中的器官、组织和病变视觉表现，来快速进行诊断和临床干预。其中牙齿分割和编号是口腔分析诊断的重要开始。在牙科治疗中，X光片是一种常见的诊断工具，在诊断中被医生用于检查牙齿、牙龈、颚骨和口腔骨骼结构的状态。自动准确地分割牙齿位置和区域，对牙齿分割模型非常重要，由于牙科X射线图像的对比度差，特征提取困难，导致预测掩膜质量较低。本文为了提高牙齿X光图像分割的质量提出了YDBD模型。该模型基于YOLOv8改进的EfficientNet骨干网络中加入了空间注意力机制，增加了牙齿位置的识别的准确性。在其Neck结构中，使用了动态上采样来减少边缘伪影，并加入高层特征图拼接，对边缘特征进行更好的提取，提高预测掩膜贴合度。实验结果表明准确性和平均精度均值方面对比原模型有部分提升。

关键词

牙齿分割，注意力机制，实例分割，YOLOv8

Dental X-Ray Medical Image Segmentation Model Improved Based on YOLOv8

Yizu Du, Xuedian Zhang*

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: May 26th, 2025; accepted: Jun. 19th, 2025; published: Jun. 26th, 2025

Abstract

Medical dental images play a crucial role in helping doctors make rapid diagnoses and clinical

*通讯作者。

文章引用: 杜一族, 张学典. 基于 YOLOv8 改进的牙齿 X 光医学图像分割模型[J]. 建模与仿真, 2025, 14(6): 370-382.
DOI: 10.12677/mos.2025.146504

interventions based on the visual manifestations of organs, tissues, and lesions in the images. Among them, tooth segmentation and numbering are important starting points for oral analysis and diagnosis. In dental treatment, X-rays are a common diagnostic tool, which are used by doctors to examine the condition of teeth, gums, jawbones and oral skeletal structures during diagnosis. Accurate and automatic segmentation of tooth positions and regions is of great significance for the tooth segmentation model. Due to the poor contrast of dental X-ray images and the difficulty in feature extraction, the quality of the prediction mask is relatively low. In this paper, to improve the quality of dental X-ray image segmentation, the YDBD model is proposed. This model is based on the improved EfficientNet backbone network of YOLOv8 and incorporates a spatial attention mechanism, which enhances the accuracy of tooth position recognition. During upsampling, dynamic upsampling is used to reduce edge artifacts and high-level feature maps are concatenated to better extract edge features, thereby improving the fitting degree of the prediction mask. Experimental results show that there is a partial improvement in terms of accuracy and mean average precision compared to the original model.

Keywords

Tooth Segmentation, Attention Mechanism, Instance Segmentation, YOLOv8

1. 引言

根据世界卫生组织 2022 年全球口腔健康状况报告表明，全球受口腔疾病影响的人约为 35 亿人[1]。而在我国进行的四次口腔卫生流行病学调查报告的数据，我国居民受到口腔以及牙齿方面的困扰也不在少数[2]。在牙科治疗中，X 光片是一种常见的诊断工具，在诊断中被医生用于检查牙齿、牙龈、颚骨和口腔骨骼结构的状态。医学图像在协助牙医观察患者口腔状态中起到重要作用，观察医学图像，识别牙科问题，使得他们为患者提供更好的治疗方法[3]。随着深度学习的发展，科技突破技术成熟，就医患者病史和数据信息处理逐渐庞大，因此对于医学图像处理的方法的需求也日渐增多[4]。

注意到这种需求之后，相对牙齿图像分割的研究开始变多变细[5]。其中不乏对传统图像处理在牙齿检测、下颌骨和牙槽骨分割三个方向的研究[6]。在 2017 年以来，深度学习的算法得到广泛关注，卷积神经网络(Convolution Neural Network, CNN) [7]、Mask r-cnn [8]、SOLOv2 [9]等等，使得医学图像分割处理得到快速发展。研究分析牙齿的形态和解剖特征来改进并优化算法，加强算法获取图像特征的能力提升精准度、掩膜贴合度等等[9]。口腔诊断中分为牙齿检测和编号[10]、牙齿分割[11]、龋齿检测[12]等等。其中牙齿分割和编号是口腔分析诊断的基础起点。牙齿分割识别图像中的牙齿区域，编号识别各个牙齿为其分配不同的号码。

图像分割主要分为语义分割、实例分割和全景分割，实例分割相对语义分割多了区分同一类的不同实例的功能。这样就可以对各个牙齿进行编号区分。早期工作主要是集中在检测[13]或者分割出牙齿[14]的单一任务中。也有结合起来做了一个 33 多类分割[15]，但是只是进行研究并且 mAP 相对较低。在其中只对牙齿进行了分割但是没有独立编号[16]和独立掩膜[17]。以及相邻牙齿位移并重叠，导致在图像中非常难以准确定位目标并区分。并且深度学习预测牙齿掩膜分割面临着重要挑战，由于 X 光图像在牙齿边缘部分对比度低，和边缘区域高度相似。但是在医学领域，分割边界对于医生分析诊断和治疗都是尤为

重要的。提高分割分割边界的质量是一项极为重要的挑战任务。图像边界的独特性和网络丢失捕获的特征信息都是影响预测掩膜贴合度的重要因素。因此为了解决在模型图像分割边界的质量，我们提出了YDBD。

YDBD 基于 YOLOv8 [18]但是针对牙齿医学影像特性进行了修改。X 光图像包括牙齿和脸部骨骼，牙齿是分割目标，分割结果是预测掩膜。原本 YOLOv8 的颈部结构是不利于细节特征的提取的，在低像素的特征图来说边缘细节各种特征都会模糊，并且牙齿数据集相对来说目标密集识别不同实例的定位能力相对困难。因此对于原本的 YOLOv8 的结构，1) 加入对高像素特征图采集并拼接用于检测和预测掩膜，减少信息丢失。2) 对于密集目标实例类别定位，增加了空间可变形注意力机制 DAT，动态调整空间采样位置聚焦特征图中目标区域。

2. 模型介绍

YOLO [18]算是较为广泛使用的模型，主要根据 UNet [19]改进得到，在分割和检测上准确性和速度方面具有尖端性能，由于其跟新较快，这里就采用性能和精度平衡的 YOLOv8 了。YOLOv8 其系列各个变体对不同的任务进行优化，保证高性能和高精度，便于在部署和开发的不同阶段使用。YOLOv8 分割模型是在 YOLOv8 检测模型的基础上改进而来的，主要是在 Head 层增加输出了三个预测 Box 的特征图和三个用来生成 Mask 系数的特征图。

2.1. YOLOv8 网络结构

实例分割主要是结合目标检测和语义分割的一种高级任务，使用目标检测区分出不同类别后用锚框进行目标定位，语义分割是区分出不同类别后用掩膜(mask)进行标记[14]，实例分割就是区分不同实例用 mask 来进行标记。其中 YOLOv8 就是一种目标检测的模型，YOLOv8 团队对 Head 层进行改动，由此得到 YOLOv8-seg 的模型。YOLOv8-seg 实例分割主要分为 Backbone 层、Neck 层和 Head 层，主要操作是对输入图片进行特征提取，进行下采样逐步降低图片分辨率和上升通道数。主干网络包括 CBS、C2f 和 SPPF 模块，再由 UpSample 模块扩大特征图再与上一层特征图进行融合。将不同特征池化结果拼接后再卷积融合生成特征表示，起到多尺度信息融合的作用。如图 1 所示。

2.2. 改进思路

在牙齿实例分割由于现有方法的局限性，口腔组织分割任然是热门话题。建模精度与图像分割质量直接相关[20]。由于 X 射线牙齿图像中，牙齿边缘模糊、牙齿重叠，相邻牙周组织、牙齿颌骨位置像素强度相似。此外还有牙齿根部轮廓会分裂变化成分支状，种种[21]原因导致牙齿实例分割复杂或者边缘模糊等。还有各种牙齿正畸和嵌体导致图像中经常出现金属伪影。在此之前主要是由有经验的牙医手动区分标记，消除相邻齿间噪点，但是效率低下[22]。

因此在本文中改进 YOLOv8 实例分割的基本模型，对其进行牙齿分割的特化，解决掩膜边缘不贴合等问题。主要的改进方法，引入可变形注意力卷积(Deformable Attention, DAT) [23]集成到其主干网络中，使得模型聚焦特征区域，提升分割的精度，使得牙齿的位置预测更加精准，减少实例识别错误。对于牙齿边缘分割预测掩膜不够贴合的问题，在 Neck 网络中插入 BiFPN [24]的多尺度融合网络结构，收集浅层信息高像素完整牙齿特征图，形成双向跨尺度连接结构，将深层特征和浅层特征跨尺度融合，使得高像素的完整牙齿特征得以保留。最后在上采样部分，部分样本参数不足，导致一些牙齿没法被很好地分割出来，引入 DySample [25]动态采样进行上采样，对低像素的特征图插值更加平衡，减少语义分割的伪影，使得上采样输出特征图更能提取牙齿特征。改进后的网络如图 2 所示。

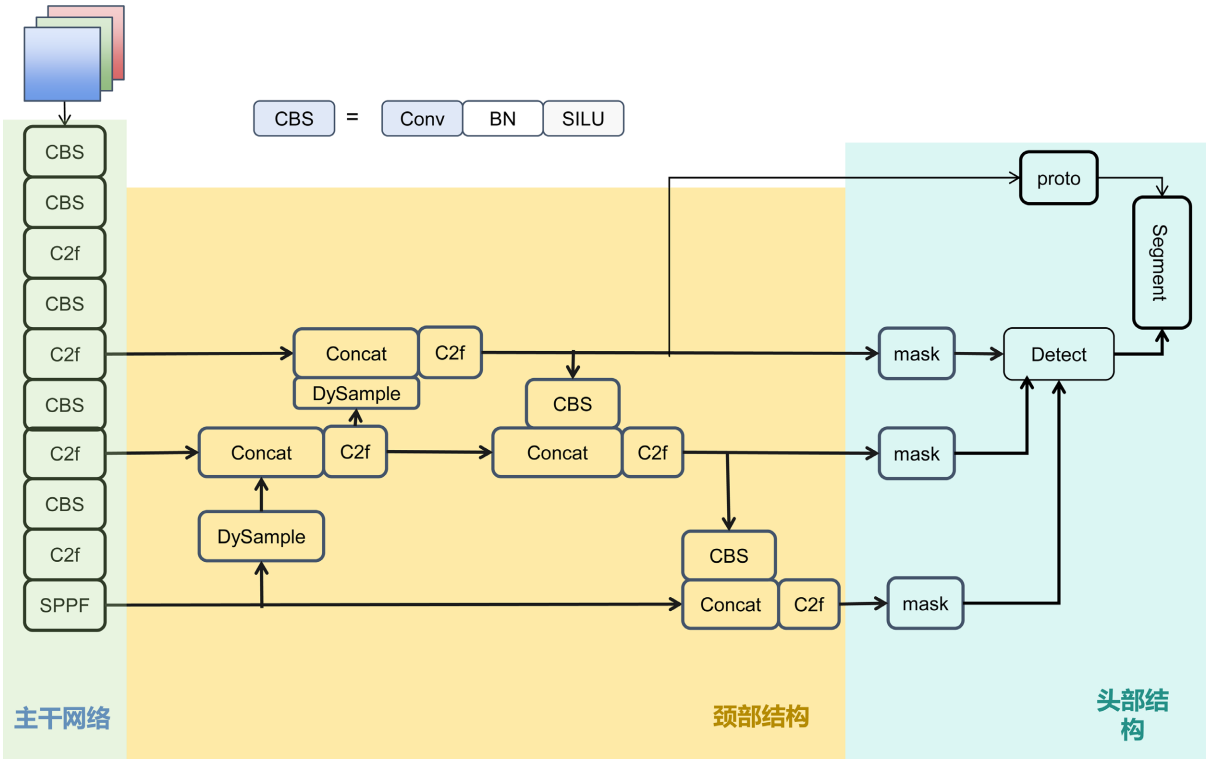

Figure 1. YOLOv8 architecture
图 1. YOLOv8 网络结构

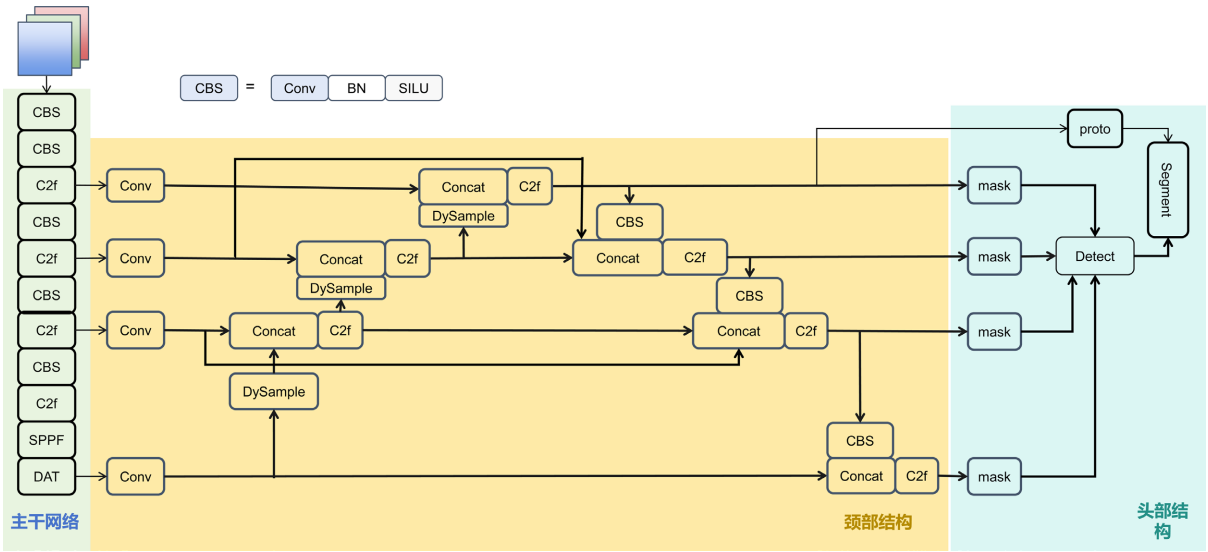

Figure 2. Improved YOLOv8 architecture
图 2. 改进后的 YOLOv8 网络结构

2.3. 可变形注意力机制

YOLOv8 原本的主干网络(Backbone)主要是在 EfficientNet [26]改进得到，在实例图像分割任务中，越高像素的特征图往往特征更多，识别特征的多少决定了一个模型的性能强弱。于是[23]在原本的 Transformer 的基础上提出了空间可变形自注意力机制(Deformable Attention Transformer, DAT)，形成了 DAT。

核心设计是模块中用数据依赖来选择自注意力中键值对应的位置，使得训练过程中模型能够聚焦于牙齿特征区域。空间可变形注意力机制示意图如图 3 所示，在传统注意力机制中，权重是通过对位置固定的注意力模型进行计算得到的，而在可变形注意力中，可以动态地调整注意力模型的形状和大小，以更好地适应不同任务和输入数据的特点。这样使得 DAT 可以在原本感受野的基础上，通过偏移网络算出到特征位置的偏移量，改变原始感受野的范围，在叠加偏移量的过程中，模拟出目标移动旋转缩放，是一种适合牙齿定位能力的空间注意力机制。DAT 流程图如图 3 所示。

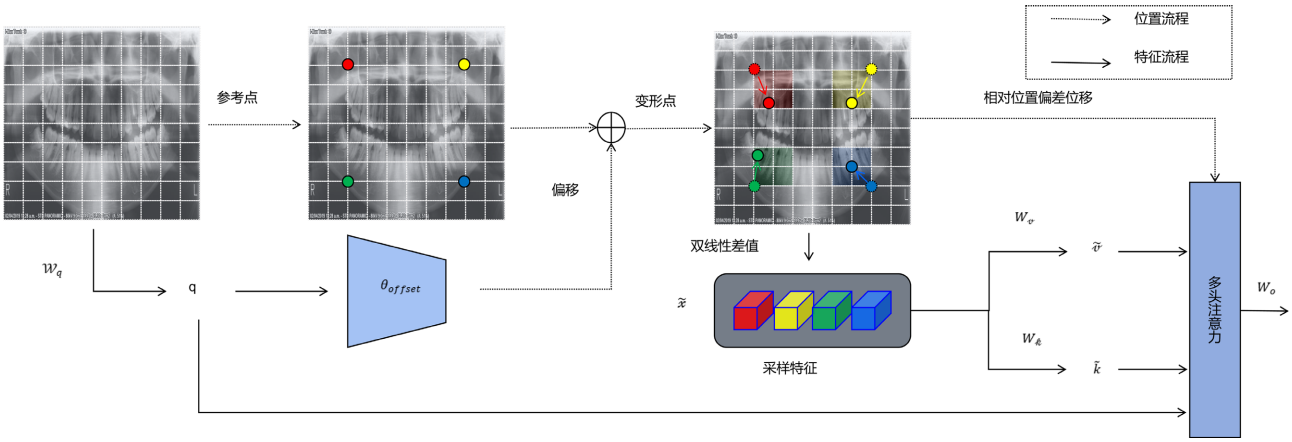

Figure 3. DAT architecture
图 3. DAT 网络结构

输入特征图表示为 $x\in\mathbb{R}^{H\times W\times C}$，生成一个点 $p\in\mathbb{R}^{H_G\times W_G\times 2}$ 的网格作为参考。将特征映射线性投射到 query token q 然后输入到轻量子网络 θ_{offset} 后生成需要的偏移量 $\Delta\theta_{offset}(q)$ 产生变形点，在变形点的位置采样得到 key, k 和 value, v 并与 q 一起传入多头注意力机制，得到 W_o 投影得到输出 z。

M 头注意力机制(Multi-Head Self-Attention, MHSA)可公式为：

$$q=xW_q,k=xW_k,v=xW_v, \tag{1}$$

$$z^{(m)}=\sigma\left(q^{(m)}k^{(m)\mathrm{T}}/\sqrt{d}\right)v^{(m)},m=1,\cdots,M \tag{2}$$

$$z=\mathrm{Concat}\left(z^{(1)},\cdots,z^{(M)}\right)W_o \tag{3}$$

其中的 $\sigma(\cdot)$ 为 softmax 函数，$d=C/M$ 为每个头部维度。$z^{(m)}$ 表示来自第 m 个注意力头部的传入输出，$q^{(m)}$、$k^{(m)}$ 和 $v^{(m)}\in\mathbb{R}^{N\times d}$ 分别表示查询、键和值的插入。为了得到每个参考点的偏移量 offset，特征映射线性投影到 query token，q 输入 θ_{offset} 生成偏移量 Δp。

$$q=xW_q,\tilde{k}=\tilde{x}W_k,\tilde{v}=\tilde{x}W_v \tag{4}$$

$$\Delta p=\theta_{offset}(q),\tilde{x}=\phi(x;p+\Delta p) \tag{5}$$

其中 $\tilde{k}$ 和 $\tilde{v}$ 表示变形的键和值，Δp 表示偏移量。采样函数 ϕ 设置成双线性插值，并使其可微：

$$\phi\left(z;\left(p_x,p_y\right)\right)=\sum_{(r_x,r_y)}g\left(p_x,r_x\right)g\left(p_y,r_y\right)z\left[r_y,r_x,:\right] \tag{6}$$

其中 $g(a,b)=\max(0,1-|a-b|)$，且 (r_x,r_y) 可以索引到输出 $z\in\mathbb{R}^{H\times W\times C}$ 的所有位置。g 将仅在最靠近 $(p_x;p_y)$ 的四个积分点上为非零，将公式(6)简化为四个位置上的加权平均，在对 $q;k;v$ 执行多层注意，并采用相对位置偏移 R。注意头的输出公式为：

$$z^{(m)}=\sigma\left(q^{(m)}\tilde{k}^{(m)\mathrm{T}}/\sqrt{d}+\phi\left(\hat{B};R\right)\right)\hat{v}^{(m)} \tag{7}$$

其中 $\phi\left(\hat{B};R\right)\in\mathbb{R}^{H\times W\times C}$ 对应了先前处理的位置嵌入并调整，最终每个头部的特征被连接在一起并通过 W_o 投影到最终输出 z 公式(3)上。

图 4 为 offset 网络。具有两个非线性激活模块，输入特征首先通过一个 5 × 5 的深度卷积来捕获低像素的局部特征。采用 GELU 激活和 1 × 1 卷积得到二维偏移量，在卷积 Conv 1 × 1 中偏差会被限制，防止特征图提取位置的强迫偏移。

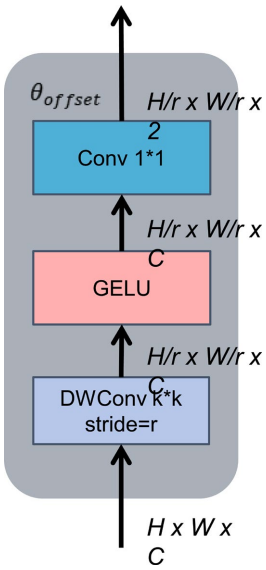

Figure 4. Offset network
图 4. Offset 网络架构

在本文中，主要采用部分的 DAT 注意力机制公式缩写如下：

$$\Delta\rho_m=\mathrm{GELU}\left(\mathrm{Conv1}D\left(\mathrm{DepthwiseConv}_{5\times5}\left(q_m\right)\right)\right) \tag{8}$$

$$K_m^{def},V_m^{def}=\mathrm{BilinearInterp}\left(x,p_r+\gamma\Delta\rho_m\right)W_m^K,\mathrm{BilinearInterp}\left(x,p_r+\gamma\Delta\rho_m\right)W_m^V \tag{9}$$

其中 $q_m=xW_m^Q$ 表示第 m 个头的查询投影，$\Delta\rho_m$ 表示偏移量，γ 为缩放因子，p_r 表示参考点网格，BilinearInterp 是双线性插值采样；W_m^K 和 W_m^V 是 Key 和 Value 的投影矩阵。多头注意力输出 z 与原文的相同可缩写为公式。R 表示相对位置编码，$d=C/M$ 为头维度，W_o 是输出投影矩阵。公式如下。

$$z_m=\mathrm{softmax}\left(\frac{q_m\left(K_m^{def}\right)^{\mathrm{T}}}{\sqrt{d}}+R\right)V_m^{def},m=1,\cdots,M \tag{10}$$

$$z=\mathrm{Concat}\left(z^{(1)},\cdots,z^{(M)}\right)W_o \tag{11}$$

这样模块能够聚焦于特征区域，从而捕获想要的信息特征精确定位牙齿位置。同时减少在需要特征外的区域无关部分的影响，从而减少计算量的同时增加牙齿定位精准度。

2.4. 改进 Neck 结构

YOLOv8 [18]原本的 Neck 函数缩略图，如图 5 所示。YOLOv8 的 Neck 部分是其架构中的关键组件，

主要用于多尺度特征融合，在对接收的 Backbone 部分的进行上采样，整体是 PAN-FPN (Path Aggregation Network + Feature Pyramid Network)结合了自顶向下和自低向上的双向特征融合路径，通过主干网络下采样输出 p_3、p_4 和 p_5 分辨率逐渐降低。再通过上采样与底层特征且大小相同的特征图拼接融合，利用下采样与之前拼接的融合生成多尺度融合后的特征提供个 Head 进行预测。YOLOv8 原本的 Neck 结构，如图 6。这样的方式会使得丢失一定的特征信息，且原本的 Concat 模块只进行低像素的拼接融合不同尺度特征，不能很好地将高像素的更多特征融合提取处理。由于原本 X 光医学图像牙齿边缘本就与背景区别不大，其图像灰度近乎相同，有时候肉眼都难以区分。高像素特征图特征无法很好的提取，导致分割效果精准度不高。

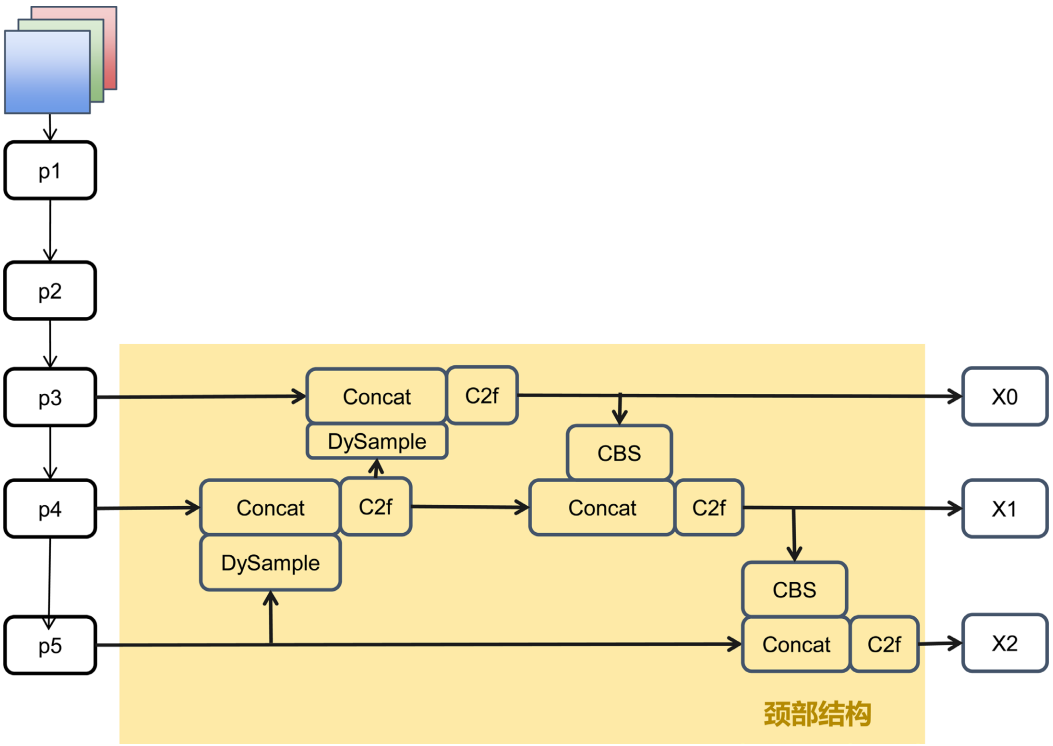

Figure 5. YOLOv8 Neck architecture
图 5. YOLOv8 原本 Neck 网络结构

为了解决在原本模型中 Neck 没有引进高像素特征图，导致预测牙齿掩膜贴合度不高的问题。越大的特征图分辨率越高，所持有的特征越多。因此在改进过程中采用了 BiFPN [24] (Bidirectional Feature Pyramid Network, BiFPN)也是采用了双层连接，允许信息在不同分辨率级别之间双向传播，有助于更好地将底层和高层之间的特征，加大了特征的上下文传播，提升了模型的分割准确性。

BiFPN 网络结构如图 7(a)所示，在图中，蓝色传递低分辨率的语义信息，红色传递高分辨的位置信息。紫色线段则是在同层的输入和输出新加的一条边，这样做使得计算成本不大的同时也可以融合更多特征。BiFPN 使得每个双向路径无论是自上而下还是自下而上都作为一个要素网络层，并重复同一层的融合，以实现更高像素的特征图信息被提取出来。在本文中使用的都是原本模型的模块，只是在图像融合的时候加入了同层级的特征信息，且新加入浅层的高像素特征图的线路，保证训练速度的同时效率高。调整不同层级之间的特征更好地匹配分割的任务需求，可以改进特征融合的效果。加入后的结构图如图 7(b)所示。先对骨干输出的特征图进行卷积操作后，模拟 BiFPN 结构对同一层的特征进行融合，最大地

保留高像素特征图的信息，使得牙齿边缘信息被保留，从而提高分割精度。

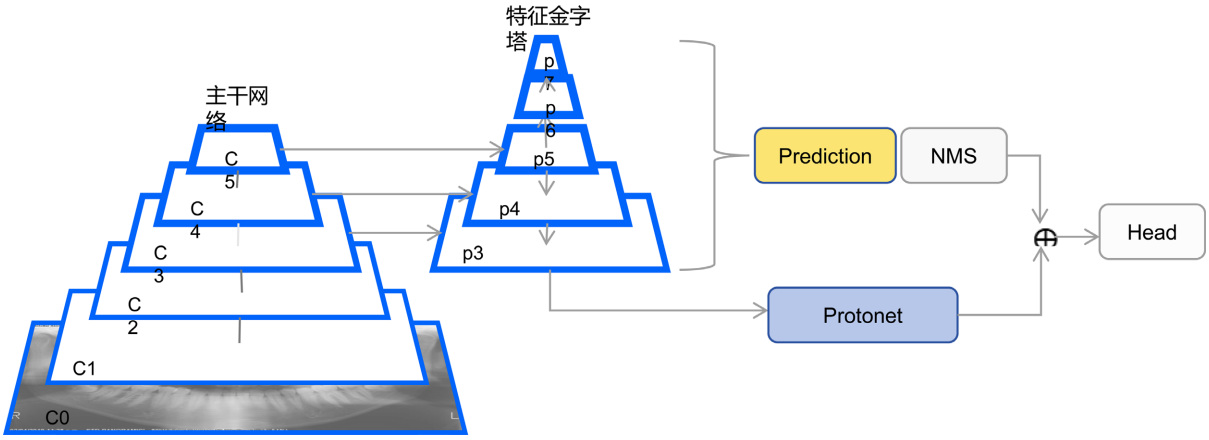

Figure 6. Neck architecture
图 6. 原本 Neck 结构

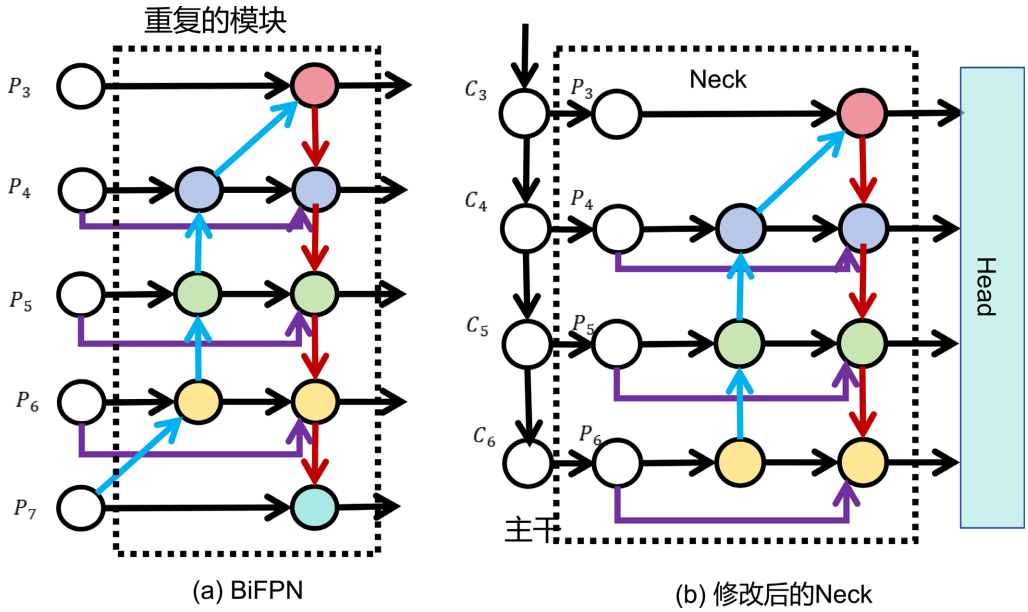

Figure 7. (a) BiFPN architecture; (b) Improved Neck architecture
图 7. (a) BiFPN 结构；(b) 改进的 Neck 结构

2.5. 上采样改进 DySample

在 yolov8 中上采样模块使用的是双线性差值算法，输入是经过主干网络后得到的低分辨率的特征图，使用双线性插值算法，将特征图的宽和高放大两倍，使得尺寸与要融合的上层特征图大小相同。主要原理是对目标位置的四个最近相邻像素进行加权平均，权重和距离相关，公式如下：

$$P(x,y)=\sum_{i,j}W_{i,j}\cdot Q(i,j) \tag{12}$$

在这里 $Q(i,j)$ 表示低像素的特征图，在点 (x,y) 上下左右邻近的的像素值，在 $W_{i,j}$ 表示双线性权重，表示距离填充点的斜度。虽然快速且没有什么其他参数，但是细节模糊。上采样特征的质量对分类、定位和分割精度有很大的影响[18]。这样就会导致分割的时候很多牙齿边缘的细节以及牙冠的细节被模糊

掉，于是为了实例分割的时候，特征图的细节可以较大程度的保留，且计算量不明显增加。就选择了DySample [25]最大的保留特征图上的特征。且 DySample 不仅在性能提升，由于在上采样时候放大特征图更精选，也可以解决参数量少的样本分割不准确的问题。动态上采样中同样使用低分辨率引导的特征作为输入，其主要是通过在采样时候控制初始采样位置，调整偏移的移动范围，将上采样分为独立的几个组。这里给定特征图 $\mathcal{X}$ 大小为 $C\times H_1\times W_1$ 和采样集 $\mathcal{S}$ 大小为 $2\times H_2\times W_2$，其中这个 2 表示 x 坐标和 y 坐标，网格采样(grid_sample)函数使用 $\mathcal{S}$ 里面的点位，将特征图 $\mathcal{X}$ 重新采样为大小是 $C\times H_2\times W_2$ 的特征图 $\mathcal{X}'$，函数如下。

$$\mathcal{X}'=\text{grid\_sample}\left(\mathcal{X},\mathcal{S}\right) \tag{13}$$

普通的实现就是使用给定上采样比例因子 s 和大小为 $C\times H\times W$ 的特征图 $\mathcal{X}$，输入通道 C 输出为 $2s^2$ 的线性层来生成大小为 $2s^2\times H\times W$ 的偏移 $\mathcal{O}$，通过像素重排整形为 $2\times sH\times sW$。偏移 $\mathcal{O}$ 和采样网格 $\mathcal{G}$ 的和为采样集 $\mathcal{S}$。

$$\mathcal{S}=\mathcal{G}+0.25\text{linear}\left(\mathcal{X}\right) \tag{14}$$

再利用网格样本设置的采样来生成上采样特征图 $\mathcal{X}'$，流程如下图 8 所示。

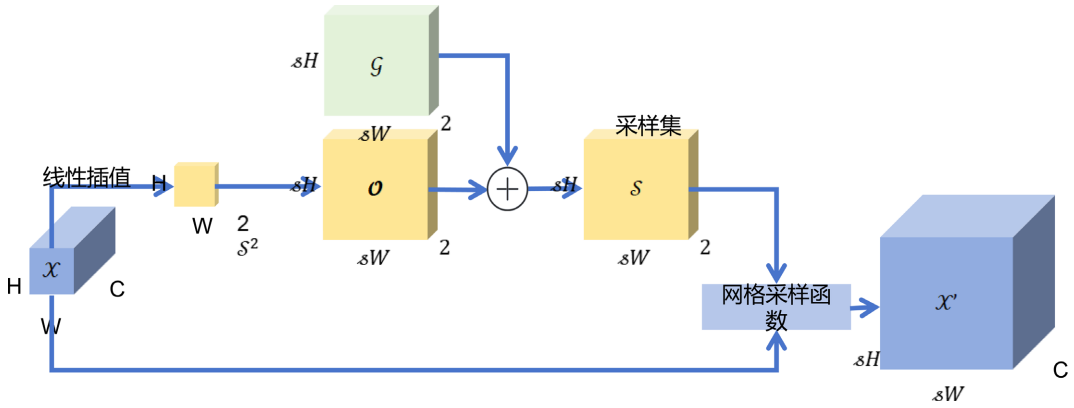

Figure 8. Sampling based dynamic upsampling
图 8. 动态采样流程

图 9 左边是原本的采样方法，而动态采样的时候规定了初始采样位置如图 9 中，尽可能与原本 YOLOv8 的插值采集信息位置不同。并规定了一定的偏移范围，设置静态因子 0.25，采样位置的移动范围受到局部约束如图 9 右。减少预测边缘的伪影，从而使得预测掩膜的牙齿边缘伪影减少，进一步提升预测掩膜贴合度。

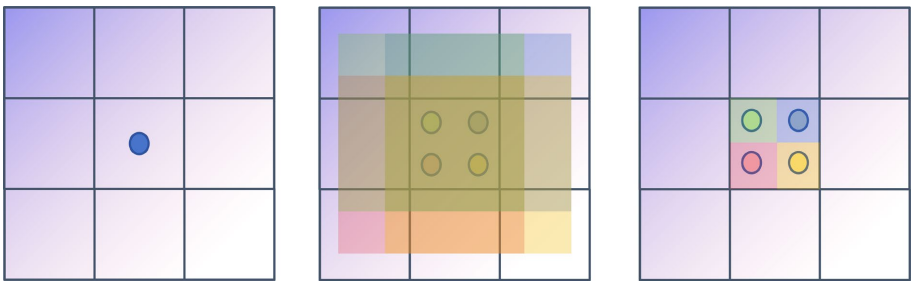

Figure 9. Set sampling positions and offset scopes
图 9. 设定取样位置和偏移位

3. 模型训练

本文采用了与 Roboflow 网站的公开数据集 Tooth Types。总共数据集图像 8329 张，训练集 7428 张，验证集 547 张，测试集 354 张。全部为牙齿的 x 光图像，已经做好了标签。根据牙齿的位置进行了编号与划分为 32 类，这样对于分割后的结果也能根据位置进行划分。已经对数据集进行过预处理调整到大小为 640×640。并在 Tooth Types 数据集上对训练样本。这个数据集对牙齿 X 光图像进行一定程度的旋转、剪切、模糊和曝光处理，旨在测试模型抗干扰能力。

3.1. 实验环境

为确保实验合理性，所有实验都是在相同的实验环境和配置参数下进行。实验环境包括 Ubuntu11.4.0 操作系统、NVIDIA RTX 4090 显卡、python3.10.12、CUDA 12.4 和 torch 2.3.0，使用 Ultralytics YOLOv8.3.96 作为基础网络模型。训练参数初始输入图像分辨率为 640×640 像素，使用上述方法对模型改进，总训练轮数(epoch)设置为 200，批次大小(batch)设置为 8，其余参数均为默认参数。

3.2. 评估指标

常见的评估指标由 TP、FP、TN、FN 组成，是机器学习模型预测的基本指标。TP (True Positive)预测为 i 类，真值也是 $i\left(i=1,\cdots,N\right)$；FP (False Positive)预测为 i 类，真值不是 i 类；TN (True Negative)预测不是 i 类，真值也不是 i 类；FN (False Negative)预测不是 i 类，真值是 i 类。其中精准率(Precision)表示正确预测的正例占所有预测为正例的比例，代表模型预测的准确率，公式(15)。召回率(Recall)表示正确预测正例占所有真实正例的比例。正确预测为某一类，占全部真值为某一类的比例。但是由于类别为 N 需要求其均值。

$$\mathrm{mP}=\frac{1}{N}\sum_{i-1}^{N}\frac{\mathrm{TP}_i}{\mathrm{TP}_i+\mathrm{FP}_i} \tag{15}$$

$$\mathrm{mR}=\frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{TP}_i}{\mathrm{TP}_i+\mathrm{FN}_i} \tag{16}$$

mIoU (mean Intersection over Union)表示预测掩膜和真实掩膜之间的交并比的平均值，由于分类有多个类别 N，因此需要先算出每个类别的 IoU 再取所有类别的平均值。公式如下：

$$\mathrm{mIoU}=\frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{TP}_i}{\mathrm{TP}_i+\mathrm{FP}_i+\mathrm{FN}_i} \tag{17}$$

mAP (mean Average Precision)在分割任务的时候结合了检测框的精度和分割掩码的准确性。用于衡量模型在不同类别下的平均检测精度和分割掩码的准确性。公式如下：

$$\mathrm{mAP}=\frac{1}{N}\sum_{i=1}^{N}\mathrm{precision}\left(i\right)\times\left[\mathrm{recall}\left(i\right)-\mathrm{recall}\left(i-1\right)\right] \tag{18}$$

mAP50 表示 IoU 大于 0.5，$\mathrm{mAP}\left[50:5:95\right]$ 指的是 IoU 阈值从 0.5 到 0.95 (步长 0.05)的平均值，是一个更加严格的评估指标。用于衡量掩膜与真值之间的关系。

3.3. 实验结果与分析

在这里对数据集进行了研究，以展示 YDBD 的每个模块如何对整体模型性能做出的贡献，验证 DAT 注意力机制、改进 BiFPN 和 DySample 这些模块，每次添加一个模块再次训练修改后的网络后，评估指

标在测试数据集上计算。在相同配置下逐个添加模块，来评估这些模块对于整体模型性能的影响。得出以下表 1 不同模型下的评价指标。这里做的是分割任务主要采用的就是 mask 相关的精准度召回率等。实验结果如表一所示，结果显示模型在该数据集上预测精度为 0.923 比原模型上涨了 1.7%，召回率上涨 1.6%，mAP50 上涨 0.9%，mAP50-95 上涨 0.6%。结果如图 10 所示，左图是 YDBD 的预测结果，右图是 YOLOv8s 预测结果。

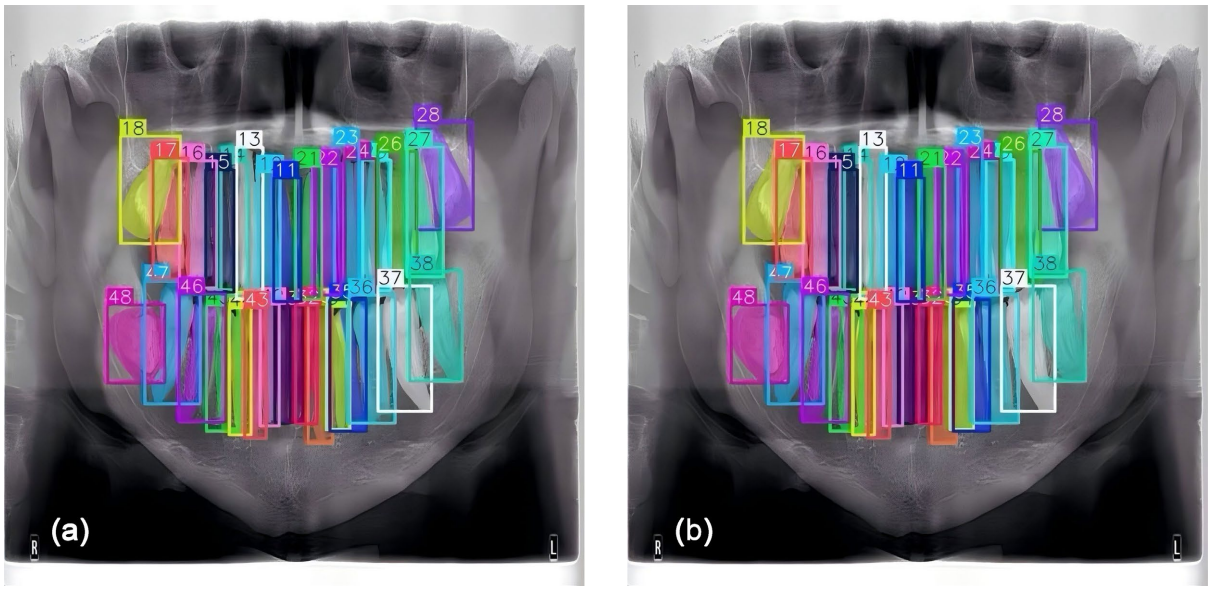

Figure 10. (a) YDBD; (b) YOLOv8
图 10. (a) YDBD; (b) YOLOv8

加入 DAT 模块到主干网络中得到的精准度、召回率和 mAP50 均有所提升，反应了加入了 DAT 注意力机制之后，解决了牙齿分割实例归类错误的问题。但是 mAP50-95 的反而下降了，这里体现了预测掩膜不够贴合真实掩膜。分析以下原因可能是数据集中于小目标或者密集目标占比高的问题，使得模型在低 IoU 上表现良好，确实牙齿数据集都是相对密集的数据，并且提高置信度阈值会过滤到定位精准但是预测掩膜不够贴合的分割结果，从而导致高 IoU 下召回率下降的问题。DAT 是将注意力分配在特征点位置的方式来提高分割精度，因此在识别位置的时候提升了准确率。

加入改进 Neck 结构后得到的 mAP 提升了，表明其先前设想正确，使用改进的 BiFPN 结构确实提升了边界感知能力，mAP 提升表明预测掩膜和真实掩膜更贴合。召回率部分降低可能表明 BiFPN 在提取高像素特征图后，预测掩膜更贴合了，但是定位能力下降牺牲了对 IoU 的预测，导致 FN 增加之后，精度不变 FP 同步减少，所以召回率下降了。

引入 DySample 上采样以后精度提升，mAP50-95 也有少量提升，表明 DySample 具有一定的边界分割提升确实与原本分析一致。而将两者同时架构进 YOLOv8 之后，他们完美地解决对方的问题，且 DySample 具有一定的少样本精度提升，所以两者同时用在 Neck 网络就使得边界伪影减少的同时还融合了更多高像素的特征信息，使得准确率，召回率 mAP 都有提升。在总改进网络上，三者将自身特点结合，使得相对原本模型的预测掩膜来说都有一定的提升。

下面是对比试验的结果如表 2 所示。主要对比了 YOLOv11s 和 Mask R-CNN 模型，v11 的分割网络可以看到虽然在原本的精度有提升，但是高 IoU 下仍然会比 YOLOv8 低。而 Mask R-CNN 也同样因为漏检变多使得召回率下降。

Table 1. Model ablation experiment
表 1. 模型消融实验

模型算法	Precision	Recall	mAP50	mAP50-95
YOLOv8s	0.906	0.909	0.909	0.525
YDBD-DySample	0.911	0.914	0.913	0.523
YDBD-DAT-DySAmple	0.907	0.905	0.909	0.530
YDBD-DAT-BiFPN	0.910	0.909	0.909	0.527
YDBD-DAT	0.919	0.915	0.917	0.536
YDBD	0.923	0.925	0.918	0.531

Table 2. Comparative experimental results
表 2. 对比实验结果

模型	Precision	Recall	mAP50	mAP50-95
YOLOv8s	0.906	0.909	0.909	0.525
YOLOv11s	0.913	0.912	0.912	0.518
Mask R-CNN	0.907	0.910	0.909	0.527
YDBD	0.923	0.926	0.918	0.531

4. 结论

本文对牙齿图像分割技术进行了探讨，提出了一种基于 YOLOv8 的模型 YDBD，用来对牙齿医学图像分割处理进行了改进，该方法先对牙齿图像进行检测定位同时生成预测掩膜。再对生成的数张预测图像进行评估得到的分割后预测掩膜和定位框。最后通过实验结果及分析来验证，改进后的算法相较于原本算法精确度、召回率和 mAP 都有提升，可克服牙齿预测掩膜不贴合，定位不精准的问题。该模型在用于检测 X 光牙齿位置和像素分割提供了一个有效的方法，为实例分割牙齿工作提供一定帮助和参考。

参考文献

[1] World Health Organization (2022) Global Oral Health Status Report: Towards Universal Health Coverage for Oral Health by 2030. World Health Organization.

[2] Zhang, T., Zhang, Q., Peng, Y., Zheng, K., Yang, L., Xiao, Q., *et al*. (2022) Retrospective Summary Analysis on the Results of Oral Health Epidemiological Investigations in China. *Journal of Oral Biology and Craniofacial Research*, **12**, 809-817. https://doi.org/10.1016/j.jobcr.2022.09.008

[3] Vimala, R. and Preethi, D.M.D. (2024) Maxillary Sinus Disease Detection and Analysis Approaches in Deep Learning: Survey. 2024 *International Conference on Cognitive Robotics and Intelligent Systems* (*ICC—ROBINS*), Coimbatore, 17-19 April 2024, 173-181. https://doi.org/10.1109/icc-robins60238.2024.10534006

[4] Tandon, D., Rajawat, J. and Banerjee, M. (2020) Present and Future of Artificial Intelligence in Dentistry. *Journal of Oral Biology and Craniofacial Research*, **10**, 391-396. https://doi.org/10.1016/j.jobcr.2020.07.015

[5] Wu, D., Jiang, J., Wang, J., Bi, Z. and Yu, G. (2025) Segmentation Algorithms of Dental CT Images: A Comprehensive Review from Classical to Deep Learning Trend. *Expert Systems with Applications*, **275**, Article ID: 126853. https://doi.org/10.1016/j.eswa.2025.126853

[6] Zhang, L., Li, W., Lv, J., Xu, J., Zhou, H., Li, G., *et al*. (2023) Advancements in Oral and Maxillofacial Surgery Medical Images Segmentation Techniques: An Overview. *Journal of Dentistry*, **138**, Article ID: 104727. https://doi.org/10.1016/j.jdent.2023.104727

[7] Suganyadevi, S., Seethalakshmi, V. and Balasamy, K. (2021) A Review on Deep Learning in Medical Image Analysis. *International Journal of Multimedia Information Retrieval*, **11**, 19-38. https://doi.org/10.1007/s13735-021-00218-1

[8] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2980-2988. https://doi.org/10.1109/iccv.2017.322

[9] Wang, X., Zhang, R., Kong, T., *et al*. (2020) Solov2: Dynamic and Fast Instance Segmentation. *Advances in Neural Information Processing Systems*, **33**, 17721-17732.

[10] Tuzoff, D.V., Tuzova, L.N., Bornstein, M.M., Krasnov, A.S., Kharchenko, M.A., Nikolenko, S.I., *et al*. (2019) Tooth Detection and Numbering in Panoramic Radiographs Using Convolutional Neural Networks. *Dentomaxillofacial Radiology*, **48**, Article ID: 20180051. https://doi.org/10.1259/dmfr.20180051

[11] Koch, T.L., Perslev, M., Igel, C. and Brandt, S.S. (2019) Accurate Segmentation of Dental Panoramic Radiographs with U-NETS. 2019 *IEEE* 16*th International Symposium on Biomedical Imaging* (*ISBI* 2019), Venice, 8-11 April 2019, 15-19. https://doi.org/10.1109/isbi.2019.8759563

[12] Lee, J., Kim, D., Jeong, S. and Choi, S. (2018) Detection and Diagnosis of Dental Caries Using a Deep Learning-Based Convolutional Neural Network Algorithm. *Journal of Dentistry*, **77**, 106-111. https://doi.org/10.1016/j.jdent.2018.07.015

[13] Chung, M., Lee, J., Park, S., Lee, M., Lee, C.E., Lee, J., *et al*. (2021) Individual Tooth Detection and Identification from Dental Panoramic X-Ray Images via Point-Wise Localization and Distance Regularization. *Artificial Intelligence in Medicine*, **111**, Article ID: 101996. https://doi.org/10.1016/j.artmed.2020.101996

[14] Jader, G., Fontineli, J., Ruiz, M., Abdalla, K., Pithon, M. and Oliveira, L. (2018) Deep Instance Segmentation of Teeth in Panoramic X-Ray Images. 2018 31*st SIBGRAPI Conference on Graphics*, *Patterns and Images* (*SIBGRAPI*), Parana, 29 October-1 November 2018, 400-407. https://doi.org/10.1109/sibgrapi.2018.00058

[15] Silva, B., Pinheiro, L., Oliveira, L. and Pithon, M. (2020) A Study on Tooth Segmentation and Numbering Using End-To-End Deep Neural Networks. 2020 33*rd SIBGRAPI Conference on Graphics*, *Patterns and Images* (*SIBGRAPI*), Porto de Galinhas, 7-10 November 2020, 164-171. https://doi.org/10.1109/sibgrapi51738.2020.00030

[16] Zhao, Y., Li, P., Gao, C., Liu, Y., Chen, Q., Yang, F., *et al*. (2020) TSASNet: Tooth Segmentation on Dental Panoramic X-Ray Images by Two-Stage Attention Segmentation Network. *Knowledge-Based Systems*, **206**, Article ID: 106338. https://doi.org/10.1016/j.knosys.2020.106338

[17] 丁状状, 侯俊, 梁善淇, 等. 牙齿图像分割算法研究[J]. 软件工程与应用, 2022, 11(6): 1282-1287.

[18] Varghese, R. and M., S. (2024) YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. 2024 *International Conference on Advances in Data Engineering and Intelligent Computing Systems* (*ADICS*), Chennai, 18-19 April 2024, 1-6. https://doi.org/10.1109/adics58448.2024.10533619

[19] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2015, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[20] Kamencay, P., Zachariasova, M., Hudec, R., Benco, M. and Radil, R. (2014) 3D Image Reconstruction from 2D CT Slices. 2014 3*DTV-Conference*: *The True Vision—Capture*, *Transmission and Display of* 3*D Video* (3*DTV-CON*), Budapest, 2-4 July 2014, 1-4. https://doi.org/10.1109/3dtv.2014.6874742

[21] Li, Q., Chen, K., Han, L., Zhuang, Y., Li, J. and Lin, J. (2020) Automatic Tooth Roots Segmentation of Cone Beam Computed Tomography Image Sequences Using U-Net and RNN. *Journal of X-Ray Science and Technology*: *Clinical Applications of Diagnosis and Therapeutics*, **28**, 905-922. https://doi.org/10.3233/xst-200678

[22] Christ, P.F., Elshaer, M.E.A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., *et al*. (2016) Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G. and Wells, W., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2016, Springer, 415-423. https://doi.org/10.1007/978-3-319-46723-8_48

[23] Xia, Z., Pan, X., Song, S., Li, L.E. and Huang, G. (2022) Vision Transformer with Deformable Attention. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 4784-4793. https://doi.org/10.1109/cvpr52688.2022.00475

[24] Tan, M., Pang, R. and Le, Q.V. (2020) EfficientDet: Scalable and Efficient Object Detection. 2020 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 10778-10787. https://doi.org/10.1109/cvpr42600.2020.01079

[25] Liu, W., Lu, H., Fu, H. and Cao, Z. (2023) Learning to Upsample by Learning to Sample. 2023 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Paris, 1-6 October 2023, 6004-6014. https://doi.org/10.1109/iccv51070.2023.00554

[26] Mehla, N., Ishita, Talukdar, R. and Sharma, D.K. (2023) Object Detection in Autonomous Maritime Vehicles: Comparison between YOLO V8 and EfficientDet. In: Namasudra, S., Trivedi, M.C., Crespo, R.G. and Lorenz, P., Eds., *Data Science and Network Engineering*, Springer, 125-141. https://doi.org/10.1007/978-981-99-6755-1_10