

文章编号: 2095-2163(2023)11-0049-09

中图分类号: TP183

文献标志码: A

# 一种面向口腔移植骨分割的 SA-UNet 网络

徐常鹏, 赵宇, 丁德锐

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

**摘要:** 在牙齿种植治疗中, 口腔移植骨分割对于辅助医生诊断有重要参考价值。口腔 CBCT 图像具有对比度较低, 移植骨边缘模糊等诸多特征, 严重制约着现有深度分割网络的应用。研究以 U-Net 为基准网络, 通过设计一种新颖的轻量级 Sharp-Attention 模块, 提出了一种改进的 SA-UNet 网络模型。具体地, 在 Sharp 模块中, 带锐化卷积核的 Depthwise 卷积操作通过锐化浅层特征以加强特征细节。CBAM 模块提升模型对于图像低层信息的关注度。进而, 采用新型联合损失函数, 缓解样本比例失衡带来的影响。最后, 在口腔移植骨数据集上验证了模型的有效性。在模型复杂度方面, 与基准网络 U-Net 相比, 在几乎没有增加计算开销的情况下, 图像分割精度得到了有效的提升; 在分割精度方面, 与现有的主流医学分割模型对比, 在 *IoU*、*Dice* 系数、Hausdorff 距离三个评价指标上的表现最佳, 得分达到了 0.866 5、0.926 2、0.509 2。

**关键词:** 口腔 CBCT 图像; 移植骨分割; 深度学习; SA-UNet; 注意力机制

## SA-UNet neural network for oral grafted bone segmentation

XU Changpeng, ZHAO Yu, DING Derui

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract:** In dental implant treatment, oral grafted bone segmentation has important value for assisting doctors in diagnosis. Oral CBCT images have many features such as low contrast and blurred edges of bone and restrict the application of the segmentation networks. Taking U-Net as the benchmark network, an improved SA-UNet model is proposed by designing a novel lightweight Sharp-Attention module. Specifically, in the Sharp module, the depthwise convolution with a sharpening convolution kernel enhances feature details by sharpening shallow features. The CBAM improves the model's attention to the low-level information. Furthermore, a novel joint loss function is adopted to alleviate the imbalance of the samples. Finally, the effectiveness of the model is validated on the dataset. In terms of complexity, compared with U-Net, the segmentation accuracy has been effectively improved with almost no increase in computational overhead; compared with state-of-the-arts, the best performance is achieved on *IoU*, *Dice* coefficient and Hausdorff distance, with scores reaching 0.866 5, 0.926 2, and 0.509 2.

**Key words:** oral CBCT images; grafted bone segmentation; deep learning; SA-UNet; attention mechanism

## 0 引言

在口腔种植领域, 新骨体积量是判断能否进行分期种植的主要依据。近年来, 由于低剂量、成像质量高、价格低廉等优点, 锥束计算机断层扫描 (Convolutional Block Attention Module, CBCT) 几乎占据牙科领域的市场<sup>[1-2]</sup>。在口腔医学图像分割中, 移植骨区域的体积分割面临较大考验<sup>[3]</sup>。究其原因, 是临床使用的骨移植材料有个体差异, 同时在新骨生成改建过程中伴随有移植材料的吸收, 使得

CBCT 图像中移植骨区边缘与周围组织很难清晰辨识<sup>[4-5]</sup>, 从而影响医生对移植骨区域准确性的判断。对口腔移植区的勾画通常要求临床医生手工勾勒靶区的轮廓边界。一个标准的 CBCT 图像通常包含几百张切片, 传统的手动勾勒需要耗费大量的时间和人力。不同医生对同一植体区域的勾画也有可能存在一定的差异, 这些差异会影响治疗计划的质量与优化。另一方面, CBCT 图像的质量也会受到多种因素影响<sup>[6]</sup>, 如 CBCT 成像对移植体区域的图像分辨率不高、图像对比度低等。因此, 要想获得比较精

**基金项目:** 国家自然科学基金(61973219)。

**作者简介:** 徐常鹏(1997-), 男, 硕士研究生, 主要研究方向: 医学图像处理、深度学习。

**通讯作者:** 丁德锐(1981-), 男, 博士, 教授, 主要研究方向: 图像处理、智能优化算法。Email: deruiding2010@usst.edu.cn

**收稿日期:** 2022-11-21

确的分割结果,仍然面对巨大的挑战。

基于深度学习技术的图像分割方法具有操作方便、准确性高等诸多优势,在口腔医学领域有着广泛的应用前景和很高的经济价值<sup>[7-8]</sup>。目前,卷积神经网络被广泛用于医学图像分割任务中。例如,Long等学者<sup>[9]</sup>提出全卷积(Fully Convolutional Networks, FCN)网络结构。FCN采用反卷积层对特征图上采样,使其恢复到输入图像相同的尺寸,实现了像素级的分类。Ronneberger等学者<sup>[10]</sup>提出了U-Net网络结构。该网络采用U型编码解码结构,通过跳跃连接实现不同尺度特征图低级细节与高级语义的结合,成为了医学图像分割的基准网络。Oktay等学者<sup>[11]</sup>提出Attention U-Net。该模型以U-Net为基础,在解码器部分通过注意力门(Attention Gates)控制特征的重要性,抑制不相关背景区域的影响。Zhou等学者<sup>[12]</sup>提出U-Net++网络,通过不同深度的U-Net的有效集成来缓解未知的网络深度,通过深度监督进行共同学习。Alexey等学者<sup>[13]</sup>提出Vision Transformer(ViT)网络。这是首次将自然语言处理中的Transformer模型用于计算机视觉中,并且在大规模数据集中取得较好效果。Chen等学者<sup>[14]</sup>提出TransUnet,通过Transformer的全局注意力弥补了U-Net远程建模依赖关系的局限性。

针对口腔CBCT图像,尤其是对于口腔骨移植区分割任务<sup>[15]</sup>,传统的实施方案主要依赖于医生手动的线性勾勒,目前并无成熟的口腔移植骨分割网络模型。同时,移植区位于牙槽骨内部或表面,体积相对较小,因此口腔移植骨图像具有样本不均衡,边缘模糊等特点。虽然U-Net网络通过多尺度特征提取和跳跃连接层保留低层信息的特性,在网络医学图像中有许多成功的应用,但由于低级特征层和高级特征层的语义差距较大,U-Net应用跳跃连接来合并低级别和高级别的特征层,不仅容易导致特征映射模糊,而且还易导致过度和欠分割的目标区域。此外,由于口腔移植骨CBCT图像数据集具有标记的数据量较少,加之样本不均衡、边界模糊等问题,U-Net网络分割方法会出现模型特征提取能力较差、边界分割效果差等问题,导致网络分割结果精度有限。

针对上述难点,本文提出一种基于U-Net和注意力机制的新型网络(SA-Unet),以提高口腔移植骨分割的准确率。具体地,使用U-Net网络作为主体框架,将跳跃连接层简单地复制操作改进为

Depthwise卷积,以锐化低级特征层并减轻特征拼接引起的语义差距。其次,通过在解码器部分特征拼接之后嵌入CBAM<sup>[16]</sup>模块,增强模型对于特征的提取能力,使得模型更加关注目标区域的细节信息(如边缘、纹理等),并抑制其它无用信息。最后为解决数据正负样本不均衡以及简单困难样本不均衡的问题,论文设计了新型的联合损失函数。

本文主要贡献如下:

(1)提出基于U-Net和注意力机制的编码解码网络(SA-UNet),以实现口腔移植骨的精准分割。在跳跃连接层设计的锐化操作不会引入任何额外的可学习参数。

(2)设计一种轻量级的Sharp-Attention模块。通过引入带锐化卷积核的Depthwise卷积和CBAM模块,增强了图像的细节信息,平衡了不同层间语义差距。本模块也适用于其他语义简单的医学图像,也可集成到其他U型网络结构中用以提升模型性能。

(3)采用新型联合损失函数,缓解正负样本比例失衡带来的影响。在制作的口腔CBCT图像数据集上验证本文提出模型的有效性。

实验结果表明,相比其他方法,本方法在IoU、Dice系数、HD距离三个指标上均取得了最佳评分。

## 1 相关工作

由于U-Net网络结构在生物医学图像分割中的良好表现,通常被作为医学图像分割的基准网络,并且陆续推出了基于U-Net改进的系列成果。受到残差连接的启发,Xiao等学者<sup>[17]</sup>提出了Res-UNet用于视网膜图像的分割。Guan等学者<sup>[18]</sup>提出了Dense-UNet,将UNet的每一个子模块分别替换为密集连接的形式,并设计Fully Dense UNet模型用于去除图像中的伪影。Alom等学者<sup>[19]</sup>提出了循环卷积网络(RU-Net)和循环残差卷积网络(R2U-Net)。这2种网络分别用循环卷积层(RCLs)和带有残差的循环卷积层(RCLs)代替正向卷积层。Çiçek等学者<sup>[20]</sup>基于U-Net提出了3DU-Net,实现了对于3D图像的医学分割。Milletari等学者<sup>[21]</sup>在3DU-Net的基础上提出了V-Net,用卷积代替池化,通过转置卷积上采样,实现了基于体积的分割方法。

注意力机制的基本方式是通过启发式搜索对卷积特征进行选择,即通过学习要强调或抑制的特征来有效帮助信息在网络内的流动。Jaderberg等学者<sup>[22]</sup>提出了Spatial Transformer Network,空间注意

力主要关注图像的空间位置信息,生成空间特征图保存关键信息。Hu 等学者<sup>[23]</sup>提出了 SENet,通道注意力可以有选择性地关注具有更多信息的特征通道,并对无用特征进行抑制。Zhao 等学者<sup>[24]</sup>提出了 PSANet。该方法只计算每个像素与其同行同列、即十字上的像素的相似性,通过进行循环间接计算得到各像素间的相似性,有效降低计算复杂度。Wang 等学者<sup>[25]</sup>提出一种可堆叠的残差注意力网络(Residual Attention Network)。Vaswani 等学者<sup>[26]</sup>提出自注意力(Self-Attention)机制,不使用 RNN 或 CNN 等复杂的模型,仅仅依赖于 Attention 模型就可以实现训练并行化且拥有全局信息。Woo 等学

者<sup>[16]</sup>提出了轻量级的卷积注意力机制模块(CBAM),CBAM 模块会依次沿着通道和空间维度推断注意力图,而后将注意力图与输入特征图相乘以进行自适应特征优化。

## 2 方法

### 2.1 SA-UNet 网络整体结构

本文以 U-Net 为基准网络,通过设计一种 Sharp-Attention 机制,同时融入 CBAM(Convolutional Block Attention Module)注意力模块,提出了一种改进的 SA-UNet 模型。该模型是一种多尺度的对称 U 型结构网络,如图 1 所示。

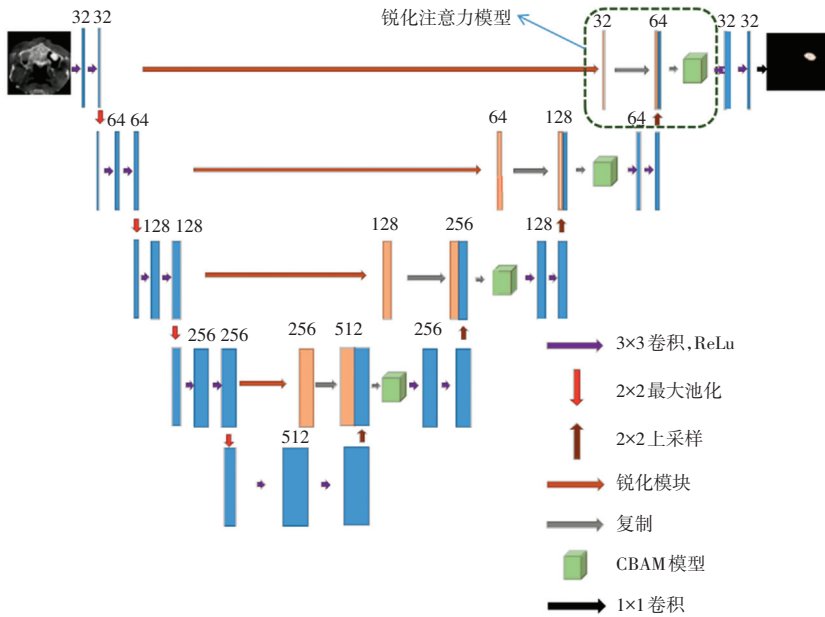


图 1 SA-UNet 网络结构

Fig. 1 The structure of SA-UNet

设计的 SA-UNet 网络由编码器、跳跃连接层、解码器三个部分组成。具体地,编码器部分包含了卷积、池化、下采样等模块;跳越连接层部分在浅层引入了 Sharp 模块<sup>[27]</sup>;解码器部分包含双线性上采样、CBAM 模块、卷积和池化模块。此外,针对口腔移植骨图像边界模糊、样本不均衡的特点,网络的训练采用了一个新设计的联合损失函数,用以提升图像分割质量。

在细节上,编码器部分与传统的 U-Net 类似,由 5 个模块组成,每个模块包含 2 个  $3 \times 3$  卷积层和一个 ReLU 激活层,随后连接一个  $2 \times 2$  的最大池化层,且第 5 个模块不包含池化层。分别使用 32、64、128、128、256 和 512 个卷积核,即在每个模块之后,特征图通道的数量都增加了一倍。

在 U-Net 型网络结构中,编解码器网络中的跳跃连接层对于恢复预测的细粒度细节方面起着至关重要的作用。传统的对浅层信息直接复制的跳跃连接方式在融合低层和高层不同的语义特征时,由于较大的语义差异容易导致模糊的特征映射,从而降低了分割精度。为此,本文改进了这一连接方式,在前 2 个模块的跳跃连接层中引入了 Sharp 模块,使用带锐化卷积核的 Depthwise 卷积,从而在每个特征通道上对特征图进行沿通道的卷积操作。Depthwise 卷积操作锐化了浅层特征以加强特征细节,使得在特征拼接时平衡了不同层间语义差距,同时也有助于减少早期阶段在整个网络层中传播的高频噪声成分。值得一提的是,Sharp 模块不会引入任何额外的可学习参数。

解码器部分通过嵌入 CBAM 模块,结合空间和通道的注意力机制,提高了模型对于感兴趣区域的关注程度。解码器部分同样由 5 个模块组成,每个模块包含一个  $2 \times 2$  的反卷积(即对特征进行上采样)、特征拼接、CBAM 模块,然后是 2 个带有 ReLU 激活的  $3 \times 3$  卷积层,且第 5 个模块有一个额外的  $1 \times 1$  卷积层实现特征降维。解码器模块的卷积层分别用 256、128、64 和 32 个卷积核,即每个模块之后特征通道数减半,最后输出与输入图像同等分辨率的输出图像。输出是对于每个像素类别的预测。接下来将进一步阐述所设计网络核心模块。

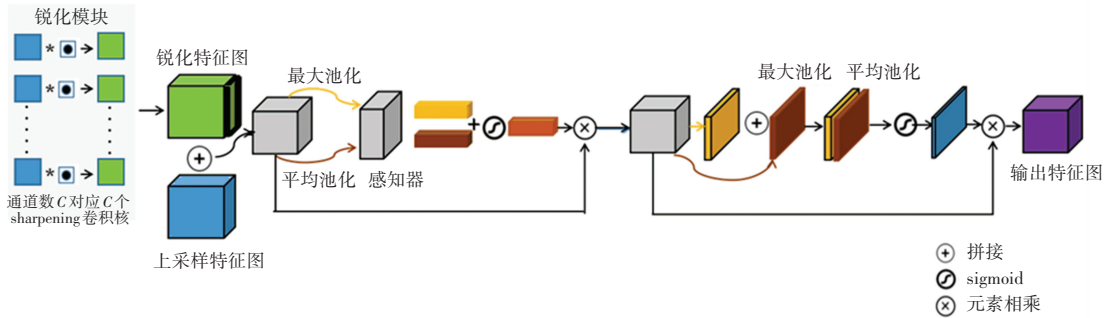


图 2 Sharp-Attention 模块图

Fig. 2 Sharp-Attention block

(1) Sharp 模块。模块中的 Depthwise 卷积操作是在特征融合之前使用锐化卷积核每个通道的特征图独立地进行卷积。锐化卷积核是图像拉普拉斯算子<sup>[28]</sup>的近似,是一个二阶导数算子,能够在任何方向上响应强度跃迁。例如,带有如下卷积核的一个拉普拉斯高通滤波器,考虑了输入图像中参考像素的所有 8 个临近值,如式(1)所示:

$$K = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (1)$$

显而易见,该卷积核的功能是增加了中心像素相对于相邻像素的强度。进而,设  $I$  是输入图像,则锐化图像  $S$  为:

$$S = I + K * I \quad (2)$$

其中,“ $*$ ”表示卷积。

由于编码器的特征层是多维的,一般大小为  $W \times H \times C$ ,其中  $W$ 、 $H$  和  $C$  分别表示编码器特征映射的宽度、高度和通道数。因此,本文使用基于拉普拉斯滤波器核  $K$  的锐化空间核沿通道对每一个特征层进行卷积,即 Depthwise 卷积操作。具体地,使用  $C$  个滤波器,输入特征图的每个通道分别与核  $K$  进行卷积,步幅为 1。每一个卷积操作都产生一个

## 2.2 Sharp-Attention 模块设计

如前所述,本文的创新性工作之一是将 Sharp 模块与 CBAM 模块结合,提出了 Sharp-Attention 模块,并应用于 U-Net 网络跳跃连接与解码器部分。该模块主要由 2 个部分组成,如图 2 所示。一部分是 Sharp 模块,通过带锐化卷积核的 Depthwise 卷积操作,对不同尺度的特征图实现锐化操作;另一部分是 CBAM 模块,该模块包含一个通道注意力单元和空间注意力单元,用来对不同特征图赋予不同的关注程度。对此拟做阐释分述如下。

大小为  $W \times H \times 1$  的特征图。为了保持输出维度与输入的维度相同,在特征拼接的过程中执行填充操作。稍后,将这些特征图堆叠在一起获得了尺寸为  $W \times H \times C$  的输出。由于锐化卷积核  $K$  没有可调参数,因此在模型优化过程中没有参数更新,不会产生额外的计算成本。

(2) 卷积注意力模块(CBAM)。是一种简单而有效的注意力模块,如图 3 所示。

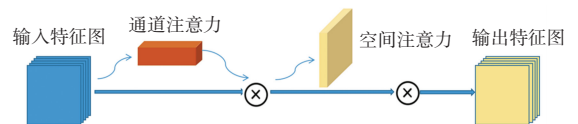


图 3 CBAM 模块图

Fig. 3 CBAM block

在解码器部分,把拼接之后的特征图输入到 CBAM 模块,此后沿着通道和空间维度依次推断注意力图,然后将注意力图与特征图相乘后的结果进行自适应优化。给定大小为  $W \times H \times C$  的特征图, CBAM 依次推断出大小为  $1 \times 1 \times C$  的一维通道注意力图  $M_c$  和大小为  $H \times W \times 1$  的空间注意力图  $M_s$ 。具体地,通道注意力将输入的特征图  $F_{input}$  分别经过基于高度和宽度的全局池化和平均池化,而后再分别经

过多层感知器、element-wise 加和操作、Sigmoid 激活, 最终生成通道注意力特征图  $M_c$ 。通道注意力特征图与输入特征图做元素相乘生成空间注意力模块所需的输入  $F_{out1}$ , 数学定义如下:

$$F_{out1} = M_c(F_{input}) \otimes F_{input} \quad (3)$$

空间注意力模块包括全局池化和平均池化、拼接、卷积操作、降维以及 Sigmoid 激活而生成空间注意力特征图  $M_s$ 。同样地, 与输入特征图  $F_{out1}$  相乘得到 CBAM 模块的输出特征图  $F_{out2}$ , 其数学表达式为:

$$F_{out2} = M_s(F_{out1}) \otimes F_{out1} \quad (4)$$

其中, “ $\otimes$ ” 为对应像素值相乘。

对于一个特征图来说, 通道注意力模块压缩了输入特征图的空间维数, 用于聚集空间信息。空间注意力模块是利用特征的空间关系来生成一个空间注意图。空间特征更多关注目标的位置信息, 与通道注意力互为补充。本文认为 Sharp 模块对编码器特征的每个通道独立地进行卷积操作, 能够增强浅层特征图的细节信息, 同时有助于减少早期阶段整个网络层中传播的高频噪声成分。进而, 在解码器部分高层与低层特征图拼接之后嵌入的 CBAM 模块, 使得通道注意力模块能够更加有效地关注经过 Sharp 模块增强的浅层特征通道的信息, 从而加强对于浅层信息的提取, 更好地平衡高层和浅层的语义差距。

### 2.3 损失函数

加权交叉熵常用来解决医学图像中的类别不平衡问题。通过对每个类别加上适当的权重, 从而抵消了数据集中存在的类不平衡。进而, 为了降低易分样本对损失函数的贡献, Lin 等学者<sup>[29]</sup> 通过引入难易区分权重, 使模型更加专注难分样本的区分, 得到了 Focal Loss (FL):

$$Loss_{FL} = - \frac{1}{N} \sum_{i=1}^N (\beta y_i (1 - p_i)^\gamma \log p_i + (1 - \beta) (1 - y_i) p_i^\gamma \log(1 - p_i)) \quad (5)$$

其中,  $\beta$  表示类别平衡因子,  $\gamma$  是难易样本平衡因子。通过对  $\beta$  参数的调整, 可解决正负样本不平衡的问题; 通过  $(1 - p_i)^\gamma$  参数的调整, 可改善难易样本不平衡问题。

Hausdorff Distance (HD) Loss 可用于优化分割的最大距离误差<sup>[30]</sup>:

$$Loss_{HD} = \frac{1}{|\Omega|} \sum_{\Omega} ((p - q)^2 \cdot (d_p^\alpha + d_q^\alpha)) \quad (6)$$

其中,  $\Omega$  表示图像定义的网格;  $q$ ,  $p$  分别表示预测图与 ground-truth;  $d_q$ ,  $d_p$  分别表示预测图与 ground-truth 的距离变换图; “ $\cdot$ ” 表示对应元素相乘。HD Loss 不是只关注最大的分割误差, 而是使用  $\alpha$  作为惩罚因子, 平稳地对于较大的分割误差给予惩罚。

在医学图像分割任务中, 类别不平衡问题很常见, 即正负样本比例失衡。口腔移植骨图像也是如此, 背景像素约占 95% 以上, 数据集具有显著的正负样本比例不平衡特征。同时, 由于二维图像中移植骨区域较小, 对于较小的样本容易产生误判。考虑到基于边界的损失函数能够通过最大距离的约束有效地减小误判的产生, 本文设计一种联合 FL 与 HD Loss 的损失函数, 其定义如下:

$$L = Loss_{FL} + \lambda Loss_{HD} \quad (7)$$

这里, 参数  $\lambda$  平衡了 HD Loss 在训练中的权重, 以实现更好的分割性能。经过实验发现,  $\gamma$  参数设置为 2,  $\beta$  设置为 0.8,  $\alpha$  值在 1.0 和 3.0 之间实验结果较好。本文的实验中,  $\alpha$  设置为 1.5,  $\lambda$  设置为 0.2。

## 3 实验及结果分析

### 3.1 实验环境与参数设置

本文实验使用 Python 编程语言, Pytorch 框架, 硬件配置如下: 处理器为 Inter (R) Core (TM) i9 - 10900X CPU @ 3.7 GHz, 内存 (RAM) 为 64.0 GB, GPU 为 NVIDIA GeForce RTX 3090 24 GB 显存, 计算机系统为 Linux 操作系统 Ubuntu 18.04。

在训练过程中, 采用自适应动量估计 (adaptive moment estimation, Adam) 优化器和反向传播算法对网络进行优化与梯度更新。训练批次 (batchsize) 设为 16, 初始学习率 (learning rate) 设为 0.001, 每 100 个周期学习率下降 0.2 倍, 动量值设为 0.9, 训练次数设置为 500。

### 3.2 实验数据集建立

本实验的 CBCT 数据集来自上海交通大学医学院附属第九人民医院, 由不同年龄阶段的 10 例健康患者在术后经同一台 CBCT 扫描仪扫描得到 (伦理批号: SH9H-2022-TK53-1), 剔除无关部位的扫描切片后, 共计 505 张图片。扫描时的参数设置: 电压 120 kV、电流 5 mA、扫描时间 16~20 s、voxel size: 0.25 mm; FOV: 25 cm (D)  $\times$  17 cm (H) 分辨率, 图像矩阵为 651 $\times$ 651 $\times$ 651 体素。考虑到人工标注的骨移植区域作为标注金标准的可信度和深度学习的

可解释性,对于每张原始 CBCT 图像,口腔移植骨区域由经验丰富的口腔专科医生手工标注,并将最终标注结果以 NIFTI 格式进行存储。为了提升模型泛化能力,采取旋转、偏移、裁剪、水平翻转、随机旋转等操作进行数据集扩充,最终得到尺寸为  $224 \times 224$  的图片 1 830 张。进而可得,训练数据集为 1 464 张图片,测试数据集为 366 张图片。

### 3.3 评价指标

为了定量评估分割效果,本文使用 Jaccard 指数 (Intersection-over-Union, *IoU*)、Dice 相似系数 (Dice Similarity Coefficient, *DSC*)、Hausdorff 距离 (Hausdorff Distance, *HD*)、召回率 (*Recall*)、精确率 (*Precision*) 作为评估指标。对于各评估指标的数学定义及表述详见如下。

(1) *IoU*。给定 2 个集合  $G$  和  $P$ , 对应表示真实标签和预测标签, *IoU* 定义为:

$$J(G, P) = \frac{|G \cap P|}{|G \cup P|} \quad (8)$$

其中,  $G \cap P$  为真实标注区域与预测输出区域的交集,  $G \cup P$  为真实标注区域与预测输出区域并集。Jaccard 系数的范围从 0 到 1, 这里 1 表示真实标签和预测标签之间的完全匹配, 而 0 表示真空标签和预测标签之间的完全不匹配。

(2) *DSC*。定义为:

$$DSC(G, P) = \frac{2|G \cap P|}{|G| + |P|} \quad (9)$$

同样地, *DSC* 的值在  $[0, 1]$  范围内。其值越大表示网络的预测输出与真实标注之间的重合率越高。

(3) 召回率 (*Recall*)。定义为所有预测输出像素中被正确预测出来的比例。数学定义公式为:

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

其中,  $TP + FN$  表示数据集中的所有正例。

(4) 精确率 (*Precision*)。表示在所预测的正样本中, 预测正确的正样本所占的比例, 其计算公式为:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

(5) Hausdorff 距离。是用来度量 2 组点集的相似程度。假设有 2 组集合  $A = \{a_1, \dots, a_p\}$  和  $B = \{b_1, \dots, b_p\}$ , 则这 2 个点集合之间的 *HD* 为:

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (12)$$

其中,

$$h(A, B) = \max_{a \in A} \{ \min_{b \in B} \|a - b\| \} \quad (13)$$

$$h(B, A) = \max_{b \in B} \{ \min_{a \in A} \|b - a\| \} \quad (14)$$

其中,  $\|\cdot\|$  表示点集  $A$  和点集  $B$  间的距离范式。*HD* 值越小, 表示  $A, B$  之间的重叠度越高, 分割性能越好。

### 3.4 消融性实验

为了评估的 SA-UNet 模型中不同模块 (即 Sharp 模块和 CBAM 模块) 和不同损失函数 (即  $Loss_{FL}$  和  $Loss_{HD}$ , 记作 *FL* 和 *HD*) 在口腔移植骨分割任务中的性能, 本节设计了 7 组不同实验来进行对比。表 1 给出了在不同模块组合下的不同模型对移植骨的平均分割性能。从表 1 不难看出, 不同模块的组合都能在一定程度上提升网络模型的性能, 且 SA-UNet 网络在性能上达到了最佳, *DSC* 值达到 0.923 6, Hausdorff 距离为 0.566 3。当同时考虑提出的联合损失函数时, *DSC* 值达到 0.926 2, Hausdorff 距离为 0.509 2, 在性能上进一步提升了 0.28% 和 10.08%。

表 1 不同改进方案下的网络性能

Tab. 1 The performance of network with different improved strategies

Model	Sharp	CBAM	<i>FL</i>	<i>HD</i>	<i>DSC</i>	<i>HD/mm</i>
U-Net					0.905 6	0.697 4
U-Net w/ CBAM		✓			0.918 8	0.611 4
U-Net w/ Sharp	✓				0.920 1	0.602 0
U-Net w/ <i>FL</i> + <i>HD</i> Loss			✓	✓	0.911 0	0.542 5
SA-UNet w/ o Loss	✓	✓			0.923 6	0.566 3
SA-UNet w/ o <i>HD</i> Loss	✓	✓	✓		0.926 2	0.524 7
SA-UNet	✓	✓	✓	✓	0.926 2	0.509 2

### 3.5 网络模型对比实验

本节将通过与 U-Net、UNet++、DeepLabV3、Attention U-Net、ResUNet 等当前主流分割模型进行对比分析, 进一步验证提出模型的准确性和有效性。表 2 给出了不同网络模型在测试集上指标 *IoU*、*DSC*、*Recall*、*Precision* 与 Hausdorff 的得分情况。

根据表 2 不难看出, 本文提出的 SA-UNet 在 *IoU*、*DSC*、Hausdorff 距离三个性能指标皆达到最优, 其值分别为 0.866 5、0.926 2、0.509 2。分别领先位于其后的 Attention U-Net 网络 0.41%、0.22% 和 8.0%。ResNet 网络的 *Recall* 指标最优为 0.962 2,

Attention U-Net 网络的 *Precision* 指标在所有模型结果中最优为 0.897 0。本文的 SA-UNet 性能皆位居第二, 略低 0.08% 和 0.19%。DeepLabv3 模型由于网络层数较深, 模型参数量较大, 难以有效关注到图像低层的细节信息, 导致分割效果较差, 从而反映了 U 型网络中跳跃连接的重要性, 侧面验证本文通过跳越连接层的改进与注意力机制的结合, 加强图像细节信息而提高分割结果的合理性和优越性。

表 2 不同网络模型对比

Tab. 2 Performance comparison of different network modules

Model	<i>IoU</i>	<i>DSC</i>	<i>Recall</i>	<i>Precision</i>	<i>HD/ mm</i>
U-Net	0.831 0	0.905 6	0.935 7	0.882 4	0.694 7
U-Net++	0.848 0	0.910 4	0.948 1	0.882 8	0.641 9
DeepLabv3	0.764 7	0.852 2	0.887 8	0.820 2	0.854 2
Attention U-Net	0.863 0	0.924 2	0.959 3	<b>0.897 0</b>	0.553 5
ResUNet	0.861 3	0.921 8	<b>0.962 2</b>	0.893 0	0.597 9
SA-UNet	<b>0.866 5</b>	<b>0.926 2</b>	0.961 4	0.895 3	0.509 2

### 3.6 模型复杂度实验

为了进一步检验模型的复杂度, 本节将 SA-UNet 与 U-Net、U-Net++、Attention U-Net、DeepLabv3<sup>[31]</sup>、ResUNet 等 5 种经典分割网络模型的参数量和浮点运算量进行了分析, 结果见表 3。结合表 2 可以看出, 与基准网络 U-Net 相比, 本文提出的 SA-UNet 参数量只增加了 0.03 M, 计算量只增加了 0.4 G, 而 *IoU* 指标提升了 4.27%, *DSC* 指标提升了 2.27%。也就是说, 在几乎没有增加开销的情况下, 图像分割精度得到了有效的提升。进而, 相比于 Attention U-Net, 本文研究的模型参数量减少了 0.06 M, 计算量降低了 1.53 G, 而 *IoU* 指标提升了 0.41%, *DSC* 指标提升了 0.22%, 网络的分割精度和计算效率也得到了有效的改善。这进一步验证了 Sharp 模块没有额外可学习参数, CBAM 模块只有极少数的额外参数。因此, SA-UNet 作为轻量级的模型在小样本医学图像数据集中更加适用。

表 3 模型参数

Tab. 3 Model parameters

Model name	<i>Params/ M</i>	<i>Flops/ G</i>
U-Net	7.76	8.80
U-Net++	9.04	26.85
Attention U-Net	7.85	10.73
DeepLabv3	26.00	20.80
ResUNet	13.04	50.10
SA-UNet	7.79	9.20

### 3.7 可视化分割结果分析

本节将给出不同网络模型对口腔 CBCT 图像移植骨分割的可视化结果, 如图 4 所示。图 4(a) 中第 1~5 列分别表示来自 5 位不同病人的口腔 CBCT 图像, (b) 为移植骨的真实标签, (c)~(h) 分别为 U-Net、U-Net++、DeepLabv3、Attention U-Net、ResUNet、SA-UNet 的分割结果。

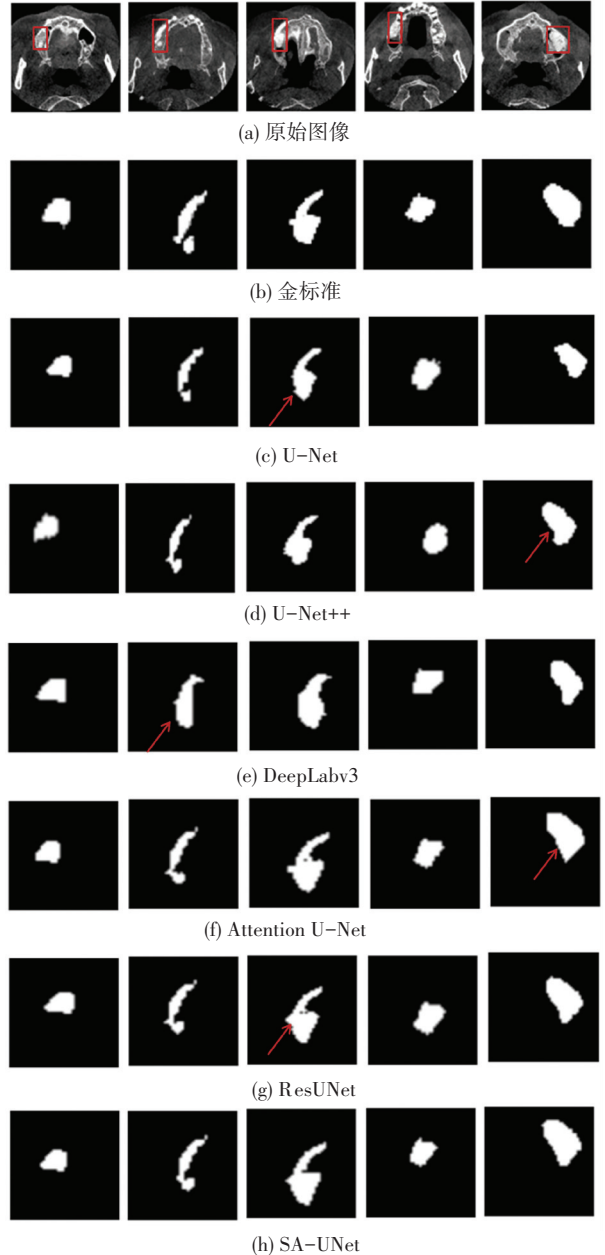


图 4 不同患者口腔 CBCT 图像的分割结果

Fig. 4 Segmentation results of CBCT images from different patients

U-Net 和 U-Net++ 网络能够基本实现对移植骨的分割。但是 U-Net 只是通过跳跃连接实现低层特征信息和高层特征信息的拼接, 忽略了二者较大的语义差距带来特征信息的丢失; U-Net++ 网络通

过大量的短连接实现多尺度特征信息的整合,但并没有加强低层特征的重要性。由图4(c)、(d)可以发现,在U-Net与U-Net++网络模型的分割结果中,容易发生欠分割,并且对于移植骨凸出的边缘分割效果不佳。DeepLabv3模型由于网络层数较深,且在解码部分使用多次的双线性采样。由图4(e)可以看出,该模型难以捕捉到移植骨图像的细节信息(如图像的边缘,纹理信息),分割边缘过于平滑,整体分割效果较差。ResUNet通过残差连接的方式保留低层特征信息,Attention U-Net通过注意力门的方式抑制不相关的区域,但仍然有特征学习不充分,在分割区域内有少量噪声产生。由图4(f)、(g)可知,2个网络对于移植骨的整体轮廓分割效果较好,但依然有噪声的残留(见第三列),并且对于移植骨凸出部分分割不够平滑,过于尖锐(见第五列)。与这些网络模型相比,本文提出的SA-UNet可以获得更优的分割性能,在分割中能够更准确地捕捉边界信息过滤分割区域中的细小噪声(见图4(h)中的第二列、第三列);此外,本文网络还能够平滑地处理模糊边界的情况,获得更为精细的分割结果(见图4(h)中的第四列、第五列)。

## 4 结束语

本文提出了一种基于U-Net和注意力机制的新型编码解码网络模型,该模型通过构建Sharp模块改进跳跃连接简单的复制操作,增强低层图像细节信息,平衡拼接操作产生的语义差距,通过在解码器部分的每一层嵌入CBAM模块,通过注意力机制分配不同权重使得模型更加关注重要信息,抑制无用信息。通过实验表明,本文提出的SA-UNet网络结构能够通过Sharp模块的锐化操作增强低层特征,并通过注意力机制提高对于图像细节信息的关注度,同时设计了联合损失函数对不平衡数据集进行优化,在移植骨图像分割结果中有较好的性能。在模型复杂度方面,与U-Net相比,几乎没有增加计算开销,并且参数量远小于其他医学模型,同时分割精度与U-Net相比在IoU、Dice系数两个指标上分别提高了0.035 5、0.020 6;在分割精度方面,与现有的主流分割模型对比,在IoU、Dice系数、Hausdorff距离三个评价指标上的表现最佳,得分达到了0.866 5、0.926 2、0.509 2。由于本模型能够有效提升图像细节信息的特征提取能力,且模型参数相对较少,因此同样适用于语义简单、结构固定的其他小样本医学图像数据集的分割以及辅助诊断

应用。

## 参考文献

- [1] 鞠昊,朱红华,段涛,等. CBCT的基本原理及在口腔各科的应用进展[J]. 医学影像学杂志,2015,25(5):907-909,942.
- [2] 王艳,李均,陈兆学,等. CBCT在口腔三维成像中的应用研究[J]. 中国医学物理学杂志,2013,30(2):4008-4011.
- [3] GUPTA J, ALI S P. Cone beam computed tomography in oral implants[J]. National Journal of Maxillofacial Surgery, 2013, 4(1):2-6.
- [4] JACOBS R, QUIRYNEN M. Dental cone beam computed tomography: justification for use in planning oral implant placement[J]. Periodontology 2000, 2014, 66(1): 203-213.
- [5] WIESLI M G, MUTLU Ö. High-performance polymers and their potential application as medical and oral implant materials: A review[J]. Implant dentistry, 2015, 24(4):448-57.
- [6] LIANG X, JACOBS R, HASSAN B, et al. A comparative evaluation of cone beam computed tomography (CBCT) and multi-slice CT (MSCT): Part I. On subjective image quality[J]. European Journal of Radiology, 2010, 75(2): 265-269.
- [7] 田启川,孟颖. 卷积神经网络图像语义分割技术[J]. 小型微型计算机系统,2020,41(6):1302-1313.
- [8] 张继凯,赵君,张然,等. 深度学习的图像实例分割方法综述[J]. 小型微型计算机系统,2021,42(1):161-171.
- [9] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Xi'an, China;IEEE,2015: 3431-3440.
- [10] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention. Cham; Springer, 2015: 234-241.
- [11] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention u-net: Learning where look for the pancreas[J]. arXiv preprint arXiv:1804.03999, 2018.
- [12] ZHOU Z, RAHMAN S M M, TAJBARKHSH N, et al. Unet++: A nested u-net architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham;Springer, 2018: 3-11.
- [13] ALEXEY D, BEYER L, KOLESNIKOV A, et al. Animage is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [14] CHEN Jieneng, LU Yongyi, YU Qihang, et al. Transunet: Transformers make strong encoders for medical image segmentation[J]. arXiv preprint arXiv:2102.04306, 2021.
- [15] 王俊成,时权,刘洪臣. 人工智能在口腔种植中的应用[J]. 口腔颌面修复学杂志,2022,23(2):81-85.
- [16] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: dblp, 2018: 3-19.
- [17] XIAO Xiao, LIAN Shen, LUO Zhiming, et al. Weighted res-unet for high-quality retina vessel segmentation[C]//Proceedings of the 2018 9th International Conference on Information Technology in Medicine and Education (ITME). Hangzhou, China; IEEE,



- 2018: 327–331.
- [18] GUAN S, KHAN A A, SIKDAR S, et al. Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal [J]. *IEEE Journal of Biomedical and Health Informatics*, 2019, 24 (2): 568–576.
- [19] ALOM M Z, HASAN M, YAKOPCIC C, et al. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation [J]. *arXiv preprint arXiv:1802.06955*, 2018.
- [20] ÇIÇEK Ö, ABDULKADIR A, LIENKANP S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation [C]//*Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention*. Cham: Springer, 2016: 424–432.
- [21] MILLETARI F, NAVAB N, AHMADI S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation [C]//*Proceedings of the 2016 fourth International Conference on 3D Vision (3DV)*. Stanford, CA, USA: IEEE, 2016: 565–571.
- [22] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks [J]. *Advances in Neural Information Processing Systems*, 2015, 28: 2017–2025.
- [23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2018: 7132–7141.
- [24] ZHAO Hengshuang, ZHANG Yi, LIU Shu, et al. PSANet: Point-wise spatial attention network for scene parsing [C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. Munich, Germany: dblp, 2018: 267–283.
- [25] WANG Fei, JIANG Mengqing, QIAN Chen, et al. Residual attention network for image classification [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE, 2017: 3156–3164.
- [26] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5998–6008.
- [27] ZUNAIR H, HAMZA A B. Sharp U-Net: Depthwise convolutional network for biomedical image segmentation [J]. *Computers in Biology and Medicine*, 2021, 136: 104699.
- [28] BURT P J, ADELSON E H. The Laplacian pyramid as a compact image code [J]. *Reading in Computer Vision*, 1987, 31 (4): 671–679.
- [29] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//*Proceedings of the IEEE International Conference on Computer Vision*. Los Angeles, California, USA: IEEE, 2017: 2980–2988.
- [30] KARIMI D, SALCUDEAN S E. Reducing the Hausdorff distance in medical image segmentation with Convolutional Neural Networks [J]. *arXiv preprint arXiv:1806.07564*, 2018.
- [31] CHEN L C, PAPANIREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [J]. *arXiv preprint arXiv:1706.05587*, 2017.