文章编号：2095-2163(2023)12-0038-08 中图分类号：TP183 文献标志码：A

面向 CBCT 图像口腔移植骨区域分割的改进 ResUNet 网络

李 辉[1]，丁德锐[1]，王 凤[2]，庄敏杰[2]，朱天佑[1]
（1 上海理工大学 光电信息与计算机工程学院，上海 200093；
2 上海交通大学医学院附属第九人民医院口腔种植科，上海 200011）

摘 要：口腔移植骨区域自动分割在计算机辅助诊断中具有重要的临床意义。针对口腔移植骨区域大小不一，形状相异以及正负样本不平衡等特点，提出一种改进的 ResUNet 深度学习网络，实现对口腔移植骨区域的自动分割。该算法设计了一个新颖的通道敏感注意力，用来捕获所有通道特征图之间的相互依赖关系，进而使用空间注意力关注这些通道特征上感兴趣的区域，提升口腔移植骨区域分割的准确性。实验结果表明，在口腔移植骨区域自动分割任务中，本文所提算法性能均优于目前医学图像分割的主流方法。

关键词：植骨区域分割；ResUNet；通道敏感注意力

An improved ResUNet network for segmentation of oral graft bone regions in CBCT images

LI Hui[1], DING Derui[1], WANG Feng[2], ZHUANG Minjie[2], ZHU Tianyou[1]
(1 School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2 Department of Oral Implantology, Shanghai Ninth People's Hospital, Affiliated to Shanghai JiaoTong University School of Medicine, Shanghai 200011, China)

Abstract: The automatic segmentation of bone regions for oral grafts has important clinical significance in computer-aided diagnosis. Aiming at the characteristics of different sizes, different shapes, and imbalance of positive and negative samples of oral bone grafts, an improved ResUNet is proposed to realize automatic segmentation of oral bone grafts. The algorithm designs a novel channel-sensitive attention to capture the interdependence among all channel feature maps. Furthermore, spatial attention is used to focus on the regions of interest on these channel features to improve the accuracy of bone region segmentation for oral implants. The final experiment shows that the performance of the studied algorithm is better than the current mainstream methods of medical image segmentation in the task of automatic segmentation of oral bone transplantation.

Key words: segmentation of bone graft area; ResUNet; channel sensitive attention

0 引 言

口腔疾病可导致疼痛、咀嚼和吞咽困难以及言语障碍。近年来，牙齿种植修复已经成为缺牙患者的常规修复选择，然而由于骨量不足和骨质较差，在上颌窦后牙区的种植修复常常面临挑战[1]。上颌窦提升术是解决上颌窦后牙区骨量不足的常用方法，是后续牙齿种植的重要辅助手段[2-3]。针对骨量不足，需要将骨材料植入上颌窦底部，一段时间后长成骨移植物[4]。在口腔医学领域，通常使用锥形术计算机断层扫描（Cone Beam Computed Tomography，CBCT）图像评估上颌窦的形态，再基于 CBCT 图像对术后患者的移植骨区域进行测量和分析。因此，在 CBCT 图像中分割出骨移植区域，成为术后检测上颌窦底抬高的关键临床步骤。目前，这一过程主要由经验丰富的口腔医生手动或半自动对口腔移植骨区域分割和测量[5]，然而，手动分割不可避免地遭遇效率低、一致性差和精度低等一系列挑战性问题。Mazzocco 等人[6]采用基于 MIMICS 软件的半自动分割方法获得骨移植区域 3D 模型，其

基金项目：国家自然科学基金(61973219)。

作者简介：李 辉(1995-)，男，硕士研究生，主要研究方向：医学图像处理；王 凤(1979-)，女，博士，副主任医师，硕士生导师，主要研究方向：口腔种植的临床与基础。

通讯作者：丁德锐(1981-)，男，博士，教授，博士生导师，CCF 会员，主要研究方向：医学图像处理。Email：deruiding 2010@ usst.edu.cn

收稿日期：2022-12-01

中区域通过阈值和区域生长相结合的方式提取。在该方法中，仍然需要手动擦除和调整误判区域，以获得准确的口腔骨移植区域，且骨移植区域通常与上颌窦相连，导致其边缘也很模糊，大大增加了分割的难度。毫无疑问，口腔移植骨区域的自动分割，会使口腔医生能更高效、更准确的对植骨区域做出诊断，有着极为重要的临床意义。

医学图像分割作为图像分割领域一个特殊且重要的分支，同样可以使用一些自然图像分割的算法。传统图像分割方法主要有：基于像素阈值的分割算法[7-8]、基于区域生长的分割算法[9]、基于小波变换的分割算法[10]等。阈值分割运用像素点灰度值的特征，并没有考虑空间特性，分割效果对噪声比较敏感，并不适合口腔移植骨区域分割任务。区域生长法是根据像素的相似性质来聚集像素点所形成的区域，该算法对分割具有相同特征的连通区域效果较好，但是由于口腔移植骨 CBCT 图片的噪声和灰度不均的问题，容易产生欠分割和过分割。

近年来，随着计算机视觉技术的飞快发展，基于深度学习的语义分割被广泛应用到医学图像分割领域[11]。现有的基于深度学习的医学图像分割算法，主要依赖于 U 型结构的全卷积神经网络。最为典型的 U 型网络是 Ronneberger 等人[12]提出的用来实现细胞图像分割的 U-Net 网络。在该网络的编码器中，一系列卷积层和连续下采样层用于提取具有大感受野的深层特征；而在解码器中，提取的深度特征通过上采样进行像素级的语义预测，同时将来自编码器的不同尺度的高分辨率特征与跳跃连接进行融合，以减轻由于下采样造成的空间信息丢失。遵循这一技术路线，已经开发出了许多算法。如 3D U-net[13]、Res-UNet[14]、U-Net++[15]、UNet3+[16]等，在各种医学图像分割任务中（如：心脏等器官分割以及息肉等病灶分割）取得了巨大的成功。Chen 等人[17]提出了 DeepLab 网络，利用多个空洞卷积获取图像多尺度特征，创造性地提出了空洞空间金字塔池化（ASPP），在不损失太多图像分辨率的同时，获取更大的感受野。随后，改进的 DeepLabV3 问世，其使用了与 UNet 类似的跳跃连接，分割性能有了显著的提升。Milletari 等人[18]提出用于 3D 图像分割的 V-Net，基于 3D 卷积提取医学图像特征。在该成果中，为解决正负样本类间不平衡问题，采用基于 Dice 系数的损失函数，显著提高了小目标区域图像分割性能。

虽然基于全卷积神经网络的方法在医学图像分割领域已经取得了优异的成绩，但仍不能完全满足医学图像分割任务对分割精度的严格要求。每个卷积核只关注整个图像的局部特征，很难学习图像全局和远程语义信息交互。一些研究试图通过使用金字塔空洞卷积层和自注意力机制来解决这个问题，然而这些方法在建立远程依赖关系方面仍然存在局限性。受 Transformer 在自然语言处理领域取得巨大成功的启发，Dosovitskiy A 等人[19]将 Transformer 带入图像处理领域，提出了用以图像分类任务的 ViT 模型。该模型将图像分为若干小块并加入绝对位置信息作为 Transformer 编码器的输入。与卷积神经网络相比，ViT 不具备卷积操作固有的归纳偏置和平移等变形，且需要在大型数据集上进行预训练，计算成本比较昂贵。Ramachandran 等[20]将自注意力机制从 Transformer 中独立出来，与卷积神经网络组合使用，直接作用于深层网络提取的特征图上，在图像分类和目标检测等密集预测任务中表现出了优异的性能。Wang 等人[21]在自注意力机制基础上将 2D 的自注意力分解为分别沿着高度轴和宽度轴两个 1D 的自注意力，设计出了 Axial-DeepLab 网络，在减少模型复杂度的同时也取得了很好的性能。最近，Valanarasu J 等人[22]在轴向自注意力机制的基础上提出 MedT 模型，使用全局和局部分支实现了对细胞核和婴幼儿脑部病理区域分割。然而，如果在全局分支依然采用 ViT 中的图像分块方案，很有可能把口腔移植骨区域划分到不同的图像快中，在建立全局信息时难免会损失许多空间细节信息。通过文献调研不难发现，在医学图像分割任务中，基于 Transformer 的模型还有很大的应用空间，需要深入探索。

令人遗憾的是，上述算法并不能有效地解决口腔移植骨区域 CBCT 图像的分割任务，其主要的原因存在如下关键性挑战：

（1）由于成像设备、成像原理以及个体自身差异的影响，口腔移植骨 CBCT 图像一般会含有很多噪声；

（2）植骨区域通常与上颌窦相连，植骨区域形状各异，边缘较模糊，没有明显的边界线，大大增加了骨移植区域分割的难度；

（3）口腔移植骨 CBCT 图像中存在严重的类别不平衡问题，相较于大片的背景，口腔移植骨区域小得多。

针对上述问题，本文提出一种面向口腔移植骨 CBCT 图像自动分割的 ResAT-UNet（Axial-

Transformer based ResUNet）算法。该算法在 ResUNet 基础上融合了 Axial-Transformer[21]、细节识别模块（Detail Identification Module，DIM）和多尺度特征融合模块（Multi-Scale Fusion Module，Multi-SFM）（编码部分）。通过与近些年医学分割网络的对比试验以及定量分析的消融实验，验证了本文提出的面向口腔移植骨 CBCT 图像自动分割任务 ResAT-UNet 网络的优越性。

1 本文方法

本文所提出的改进 ResUNet 网络结构如图 1 所示。该网络改进主要由 3 部分组成：细节识别模块、ResATM、多尺度特征融合模块（图 1 中黑色虚线框部分）。细节识别模块用于捕获伪装在低级特征中的细节信息（其中包括纹理、颜色和边缘等），使用注意力机制更多的关注潜在的口腔移植骨区域，从而减少低级语义信息中的有误信息或噪声。ResATM 提取 CBCT 图像的多尺度信息，多尺度特征融合模块渐进式融合 4 个阶段的高级语义特征信息。

编码阶段主要由浅层特征提取的卷积模块、细节识别模块以及不同数量的 Bottleneck 结构组成。首先保留了原始 ResNet 的下采样部分，并使用 3×3 卷积捕获 CBCT 图像的低级特征；进而在每个阶段的 Bottleneck 块中，3×3 卷积被两个多头轴向注意力层（高度轴方向和宽度轴方向）替换，并保留原 ResNet 中的 1×1 卷积，用来调整特征维度。

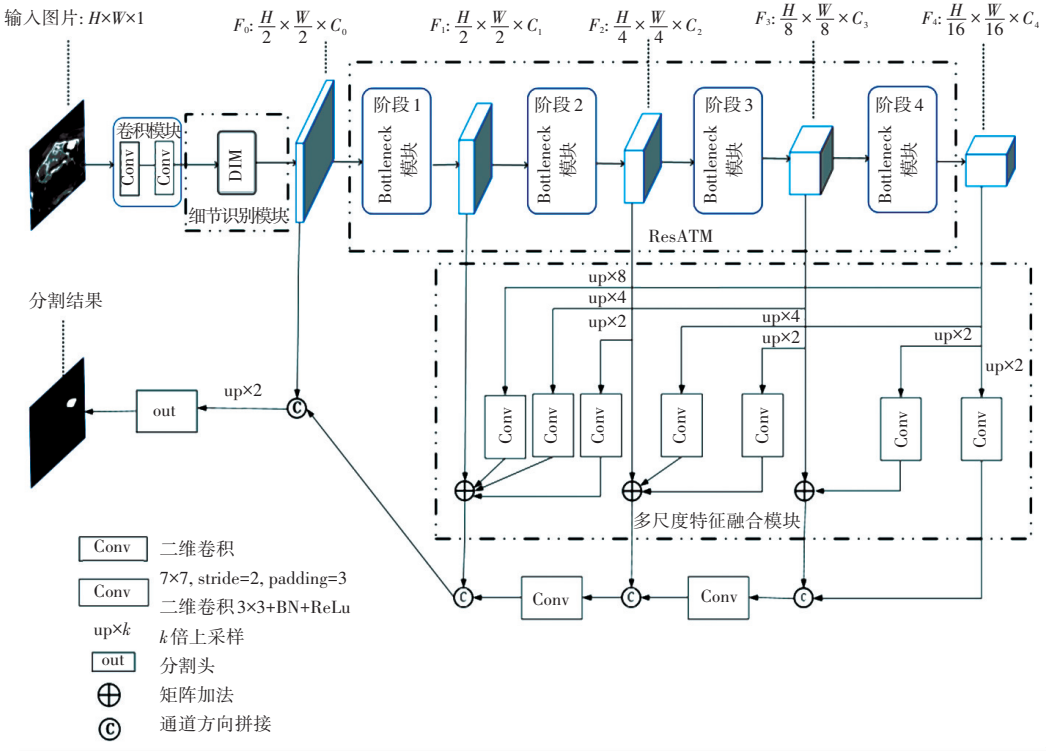


图 1 ResAT-UNet 模型结构图

Fig. 1 Architecture of ResAT-UNet

1.1 细节识别模块（DIM）

低级特征一般包括更多细节信息，如图像的边缘、纹理等特征，对分割区域至关重要。然而，牙齿骨移植区域与周围的背景很相似，特征提取过程中难免会造成反映细节信息的低级特征的损失。为了解决这一问题，本文在低级特征提取阶段添加了细节识别模块 DIM。受混合注意力机制[23]的启发，该模块主要由通道敏感注意力（Channel Sensitive Attention，CSA）和空间注意力 SA 组成。CSA 通过构建所有特征图之间相互依赖关系，有选择的强调更感兴趣的通道特征，以提高特征表达能力；SA 是 CSA 的补充，其在各个通道特征图上增强感兴趣的区域，如分割区域的边缘轮廓等空间细节信息等等。

不同于传统通道注意力机制[24]的 squeeze 操作，本文受自注意力机制的启发，并不对每个特征图做全局平均池化，而是使用两个不同的权重矩阵对 reshape 特征图进行线性变换，然后实施矩阵乘积及其 softmax 运算，从而捕获所有通道图之间的相互依赖关系；在此基础上，为每个通道分配一个可以训练学习的敏感系数 α，进一步提高通道的敏感度。最后通过残差连接，得到所有通道特征与原始特征的加权和，其结构如图 2 所示。

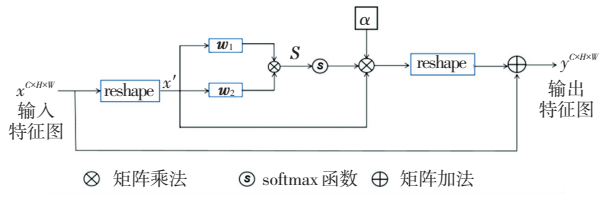


图 2　通道敏感注意力

Fig. 2　Channel sensitive attention

根据上述机理，CSA 可以表示为

$$\mathrm{CSA}(x)=\alpha\times\mathrm{softmax}((W_a x')(W_b x')^{T})x'+x \quad (1)$$

其中，$x\in\mathbb{R}^{C\times H\times W}$ 为输入特征；$x'\in\mathbb{R}^{C\times(H\times W)}$ 为 C 个展平的特征；$\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2\in\mathbb{R}^{C\times(H\times W)/8}$ 是线性映射矩阵；$\boldsymbol{S}\in\mathbb{R}^{C\times C}$ 为通道敏感注意力矩阵；α 是调整通道敏感注意力的可学习参数，该参数初始化为 0。

类似地，本文在空间注意力模块中加入可以训练学习的敏感系数 β，具体数学表达形式为：

$$\begin{cases}M(x)=\delta(\mathrm{Conv}(\mathrm{Cat}(\mathrm{Max}_{\mathrm{channel}}(x),\mathrm{Agv}_{\mathrm{channel}}(x))))\otimes x\\ \mathrm{SA}(x)=\beta\times M(x)+x\end{cases} \quad (2)$$

其中，$\delta(\cdot)$ 是 Sigmoid 函数；$\mathrm{Max}_{\mathrm{channel}}(\cdot)$ 和 $\mathrm{Agv}_{\mathrm{channel}}(\cdot)$ 分别为沿着通道方向的最大池化和平均池化；$\mathrm{Cat}(\cdot)$ 是沿通道方向拼接操作；$\mathrm{Conv}(\cdot)$ 是卷积核为 7×7，步长为 3 的卷积操作；β 是调整空间注意力的可学习参数，初始化为 0。

综上，DIM 的最终输出 F_0 为

$$F_0=\mathrm{SA}(\mathrm{CSA}(x)) \quad (3)$$

1.2　ResATM

ResNet 拥有强大的深度语义特征信息提取能力，已经被广泛用于各类图像处理任务中，但其获取图像的全局信息仍有一定的局限性。受 Transformer 在视觉领域应用的启发，本文在特征图的高度轴和宽度轴上，使用轴向 Transformer（Axial-Transformer）操作，替换 ResNet 中特征提取各阶段的卷积运算，从而构成新的 Bottleneck 结构，如图 3 所示。本文将不同数量的 Bottleneck 结构组成的特征提取模块称为 ResATM 模块，用以捕获图像的全局信息。

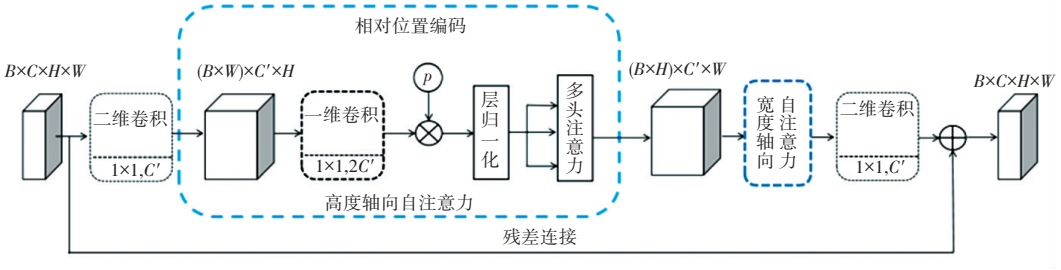


图 3　Bottleneck 结构图

Fig. 3　Structure of Bottleneck

在 Bottleneck 结构中，采用两个二维卷积，调整输入特征的维度；采用高度轴向注意力和宽度轴向注意力（蓝色虚线框部分）代替了传统 ResNet 中的卷积运算。首先，在高度轴上，通过一维卷积运算分别得到查询 $q\in\mathbb{R}^{W\times d_{\mathrm{head}}\times H}$、键 $k\in\mathbb{R}^{W\times d_{\mathrm{head}}\times H}$ 和值 $v\in\mathbb{R}^{W\times 2d_{\mathrm{head}}\times H}$。进而，为了学习到相对位置信息的远程交互能力，将相对位置编码嵌入到查询 q、键 k、值 v 中，以获取更精准的相对位置信息。在此基础上，为了获取更多的空间特征属性，Axial-Transformer 中的两个 Axial-Attention 层皆采用了多头自注意力机制（Multi-Head Self Attention，MHSA），可表示为

$$\mathrm{MHSA}(q,k,v)=\mathrm{Cat}(head_0,\mathrm{head}_1,\cdots,\mathrm{head}_{N_i})W^c \quad (4)$$

$$head_i=\mathrm{softmax}\left[\frac{(q^{\mathrm{T}}k+q^{\mathrm{T}}\boldsymbol{p}^q+k^{\mathrm{T}}\boldsymbol{p}^k)}{\sqrt{d_{\mathrm{head}}}}\right](v+\boldsymbol{p}^v) \quad (5)$$

式中：$W^c\in\mathbb{R}^{d_{\mathrm{head}}N_i\times C_{in}}$ 为神经网络的权重，N_i 表示第 i 个阶段中 Axial-Attention 层的注意力头数，d_{head} 为每个头的维度。$\boldsymbol{p}^q$，$\boldsymbol{p}^k\in\mathbb{R}^{d_{\mathrm{head}}\times H}$、$\boldsymbol{p}^v\in\mathbb{R}^{2d_{\mathrm{head}}\times H}$ 是随机初始化的张量，可以参与训练并不断地更新。并行计算轴向自注意力，最终将得到的输出头拼接在一起。

1.3　多尺度特征融合模块

在网络编码阶段，深层网络特征的语义信息表达能力强，但是特征图的分辨率低，空间几何特征细节表征能力相对较弱。反之，浅层卷积模块提取的特征图分辨率高，且具有较强的空间几何特征表达

能力,而使用注意力机制将使得更多的空间细节信息得到关注。因此,为了更加有效地利用深层语义信息和浅层空间几何信息,本文在解码阶段通过级联(连接方式如图 1)上采样的方式,实现了多尺度特征融合。该模块主要由 3 部分组成,具体说明如下:

(1)将特征图 F_4 进行双线性插值上采样操作,得到与 F_3 相同大小的结果 F_4^{up2};然后,将 F_4^{up2} 经过卷积单元 $\Delta_2(\cdot)$ 的输出与 F_3 相加,其结果再与 F_4^{up2} 经卷积单元 $\Delta_3(\cdot)$ 的输出,沿通道方向进行拼接;最后,使用卷积 $\Delta_1(\cdot)$ 运算,得到特征图 $F_{3,4} \in \mathbb{R}^{(H/8)\times(H/8)\times C_3}$。该过程可以表示为

$$F_{3,4} = \Delta_1(\mathrm{Cat}(\Delta_2(F_4^{up2}) + F_3, \Delta_3(F_4^{up2}))) \quad (6)$$

(2)将特征图 F_3 和 F_4 分别进行双线性插值上采样操作,得到与 F_2 相同大小的结果,分别为 F_4^{up4} 和 F_3^{up2};然后,将 F_4^{up4} 和 F_3^{up2} 分别经过卷积 $\Delta_4(\cdot)$ 和 $\Delta_5(\cdot)$ 运算的输出与 F_2 相加,得到的结果与 $F_{3,4}$ 沿着通道方向拼接;最后,使用卷积单元 $\Delta_6(\cdot)$ 操作得到特征图 $F_{2,3,4} \in \mathbb{R}^{H/4\times W/4\times C_2}$。该过程可以表示为

$$F_{2,3,4} = \Delta_6(\mathrm{Cat}(\Delta_4(F_4^{up4}) + \Delta_5(F_3^{up2}) + F_2, F_{3,4})) \quad (7)$$

(3)将特征图 F_2、F_3 和 F_4 分别进行双线性插值上采样操作,得到与 F_1 相同大小结果,分别为 F_4^{up8}、F_3^{up4}、F_2^{up2};然后,将 F_4^{up8}、F_3^{up4} 和 F_2^{up2} 分别经过卷积单元 $\Delta_7(\cdot)$、$\Delta_8(\cdot)$ 和 $\Delta_9(\cdot)$ 运算的输出与 F_1 相加,得到的结果与 $F_{2,3,4}$ 沿着通道方向拼接;最后,使用卷积单元 $\Delta_{10}(\cdot)$ 操作得到特征图 $F_{1,2,3,4} \in \mathbb{R}^{H/2\times W/2\times C_1}$。该过程可以表示为

$$F_{1,2,3,4} = \Delta_{10}(\mathrm{Cat}(\Delta_7(F_4^{up8}) + \Delta_8(F_3^{up4}) + \Delta_9(F_2^{up2}) + F_1, F_{2,3,4})) \quad (8)$$

在网络末端,将多尺度融合的特征 $F_{1,2,3,4}$ 与 F_0 沿通道方向拼接,再进行双线性插值上采样。由于口腔移植骨区域分割是像素级别的二分类任务,因此在网络的最后一层使用卷积操作将通道数调整为 2,再对结果进行 softmax(·) 操作,得到分类概率,式(9):

$$y_{out} = \mathrm{softmax}(out(up_2(Cat(F_0, F_{1,2,3,4})))) \quad (9)$$

其中,out(·) 定义为 1×1 卷积操作。

2 损失函数

口腔移植骨区域分割属于像素级的二分类任务,因此本文使用二分类任务常用的二元交叉熵损失函数(Binary Cross Entropy, BCE),表达式如式(10)所示:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log_e(p(y_i)) + (1-y_i)\log(1-p(y_i)) \quad (10)$$

其中,N 表示训练时送进模型的最小批次数 Batch size;y_i 表示第 i 个样本的标签;$p(y_i)$ 表示第 i 个样本送入模型预测输出的概率值。

另一方面,由于口腔移植骨分割区域远小于其背景区域,会导致像素正负样本不平衡的问题。为了减少类不平衡的影响,常见的方法是使用 Dice 损失函数和 Tversky 损失函数[25]。其中,Dice 损失被定义为预测结果与真实标签之间的重叠关系,其表达如式(11)所示:

$$L_{\mathrm{Dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i p(y_i) + \varepsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p(y_i) + \varepsilon} \quad (11)$$

式中:参数 ε 是一个特别小的数,用以防止计算损失函数时分母为 0 的情况。

由于 Dice 损失在处理类不平衡数据时有一定的局限性,其会平等的对待预测的假阴性(FN)和假阳性(FP),而牙齿骨移植目标区域较小,且与背景信息高度相似,如果此时 FN 和 FP 被赋予相同的权重,会导致预测结果有较高的精确度但召回率较低。因此,为了平衡 FN 和 FP 的误判,本文使用 Tversky 损失,其表达如式(12)所示:

$$L_{\mathrm{Tversky}} = 1 - \frac{2\sum_{i=1}^{N} y_i p(y_i) + \varepsilon}{\sum_{i=1}^{N} y_i p(y_i) + \alpha\sum_{i=1}^{N} y_i(1-p(y_i)) + \beta\sum_{i=1}^{N}(1-y_i)p(y_i) + \varepsilon} \quad (12)$$

其中,α、β 分别是控制样本预测的假阴性(FN)和假阳性(FP)的超参数,且 α 应该大于 β,用以提高召回率。当 $\alpha=\beta=0.5$ 时,Tversky 损失就变成了 Dice 损失,因此 Tversky 损失更具一般性的概括,可以更加灵活的平衡 FN 和 FP。

为了进一步说明 Tversky 损失比 Dice 损失更适合于口腔移植骨分割任务,本文做了一组对比实验,其中超参数分别设为 $\alpha=0.7$,$\beta=0.3$,实验结果见表 1。由此可见,当仅仅使用 BCE 损失时,无法很好处理口腔移植骨区域类间不平衡问题;在 BCE 损失和 Dice 损失组合使用时,类间不平衡问题得到明显改善,但召回率比精确度小;当使用 BCE 损失和 Tversky 损失组合时,召回率和精确度之间的差异明

显缩小，更好的平衡了 *FN* 和 *FP* 的误判。

表 1　不同损失函数混合组合结果

Table 1　Results of different loss functions of mixed combination

损失函数	recall	precision
BCE	0.522 1	0.653 1
BCE+Dice Loss	0.883 3	0.936 7
BCE+Tversky Loss	0.935 9	0.913 6

因此，本文实验使用的损失函数是基于上述交叉熵损失和 Tversky 损失的混合组合，可表示为

$$Loss = 0.5L_{BCE} + 0.5L_{\text{Tversky}} \tag{13}$$

3　实验结果与分析

3.1　数据集

本文实验所用的口腔移植骨 CBCT 图像数据，来源于上海交通大学医学院附属第九人民医院（伦理批号：SH9H-2022-TK53-1），由 9 例种植牙患者在术后经同一台 CBCT 扫描仪扫描得到，图像矩阵为 651×651×651 体素。对于每位种植牙患者术后 CT 图像的口腔移植骨区域，均由经验丰富的口腔临床医生手工标注，再将标注好的标签和原始 CBCT 图像转换为更适用于计算机处理的 NIFTI 格式存储。随后，将三维 NIFTI 格式的 CBCT 图像及标注图像分别沿着横断面转换成 651 张 2D 图片后，手动去除无植骨区域的扫描切片，得到 430 张 2D 图片及相应的移植骨区域图片，再经过重采样得到 256 像素×256 像素的 2D 图片。最后，采用随机旋转、翻转、添加高斯噪声等在线数据增强的方法来扩充训练数据的数量，最终得到 2 000张图片。实验按照 8∶2 划分数据集，其中训练数据集为 1 600 张图片，测试数据集为 400 张图片。

3.2　实验配置与环境搭建

实验采用基于 Python3.7 的 PyTorch 深度学习框架；网络训练与测试平台为：Intel（R） Xeon（R） Platinum 8350C 处理器、NVIDIA RTX A5000 显卡（24 GB显存），在 Windows 10 操作系统上运行。

在网络训练中，把预处理好的 CBCT 切片和相对应的口腔移植骨区域图片送入本文提出的网络中，使用随机梯度下降算法（SGD）进行网络参数的迭代优化。每一次送入网络的批量大小设置为 4，权重衰减 weight decay 为 10^{-5}，动量参数 momentum 设置为 0.9，共训练 400 轮。为防止训练过程中权重参数陷入局部最优点，将学习率设置为随训练的轮次动态衰减。初始学习率设置为 0.01，学习率的衰减公式为：$lr = base\_lr \times (1.0 - n/N)^{\beta}$，其中 n 为当前迭代次数，N 为训练迭代的总次数，衰减指数 $\beta = 0.9$。

3.3　模型对比实验

为了评价 ResATUNet 模型的性能，本文使用平均交并比（Mean Intersection over Union，MIoU）、Dice 系数（Dice Similarity Coefficient ， DSC）和平均表面距离（Average Surface Distance ，ASD）3 种评估指标。在相同的实验条件下，对 UNet 及其最新的变体模型在上述口腔移植骨分割数据集进行了对比实验，实验结果见表 2。

表 2　不同模型在口腔移植骨数据集分割精度

Table 2　Segmentation accuracy of different models in the dental implant bone dataset

模型	MIoU ↑	DSC ↑	ASD ↓
UNet（2015）	0.809 2	0.876 4	0.644 6
UNet++（2018）	0.831 7	0.890 7	0.745 5
ResUNet（2017）	0.814 8	0.887 1	0.574 4
Axial-DeepLab（2020）	0.837 2	0.905 8	0.563 7
MedT（2021）	0.826 8	0.903 1	0.629 5
UNext（2022）[26]	0.849 2	0.912 8	0.489 0
ResATUNet	**0.871 3**	**0.935 1**	**0.473 1**

注：加粗字体为最优结果

由表 2 结果可以看出，本文提出的改进模型 ResATUNet 在 MIoU、DCS 和 ASD 三种评估指标方面都有较突出的性能表现。ResATUNet 在测试集上的 MIoU、平均 Dice 系数、平均表面距离分别为：87.13%、93.51%和0.473 1，相对于 UNext 网络，性能提升了 2.21%、2.23%和 1.59%。究其原因，UNext 在编码部分为了获得更大感受野，经过多轮卷积和下采样，以至于损失很多有用的特征信息；MedT 在轴向自注意力模块做出了调整，网络也设计了两个分支，但其在全局分支依然采用 ViT 中图像分块，很有可能把口腔移植骨区域分到不同的图像块中，在建立全局信息时难免会损失不少空间细节信息。

3.4　消融实验

通过一组消融实验，验证了加入轴向 Transformer、细节处理模块 DIM 和多尺度特征融合模块（Mutti-SFM）的影响。验证结果见表 3。

表 3　控制变量验证模型性能

Table 3　Control variables to verify model performance

模型	MIoU ↑	DSC ↑	ASD ↓
ResUNet	0.814 8	0.887 1	0.574 4
ResUNet+ Mutti-SFM	0.821 4	0.893 8	0.586 2
ResUNet+Axial-Transformer	0.853 2	0.916 9	0.598 1
ResUNet+Axial-Transformer+DIM	0.865 1	0.927 4	0.480 1
ResAT-UNet	**0.871 3**	**0.935 1**	**0.473 1**

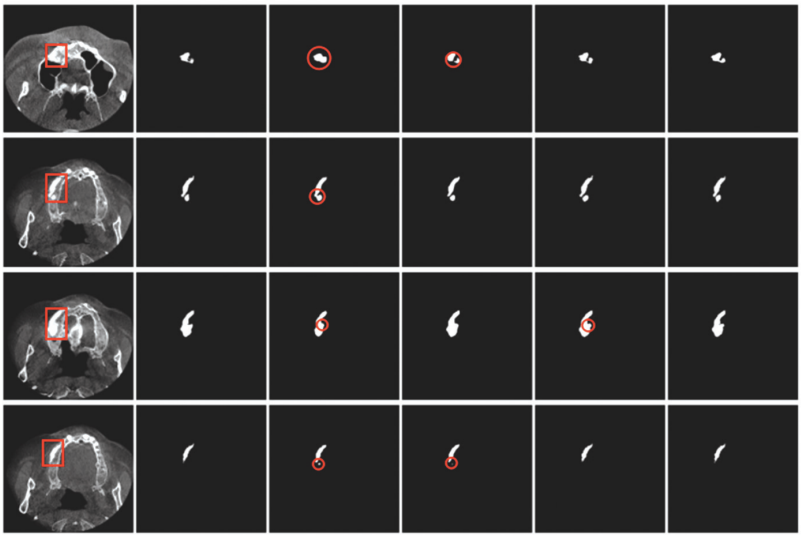
注：加粗字体为最优结果。

本文提出的模型是在 ResUNet 基础上进行修改的,加入多尺度特征融合模块,不仅在相邻特征间进行跳跃连接,而且将所有下采样的特征经过多倍上采样进行多尺度的融合。由表 3 可知,相较于 ResUNet,当使用 Axial-Transformer 替换 ResNet 中的 3×3 卷积来捕获图像的全局信息,模型性能得到明显的改善。其中,MIoU 为 85.32%,提高了3.84%;DSC 为 91.69%,提高了 2.98%。加入细节识别模块 DIM 后,平均表面距离 *ASD* 降为 0.480 1,下降了 11.8%,表明本文提出的细节识别模块,对口腔移植骨 CBCT 图像的空间几何细节处理得到显著增强。

总之,本文所提模型在关注全局信息的同时,也更加注重图像边缘等空间几何特征,从而提高了分割精度。

3.5 模型可视化结果

图 4 给出了不同模型在测试集上的分割效果可视化图,其中在原图中用红色方框凸显口腔移植骨分割区域,在预测图中用红色圆圈标出分割结果与真实标签有差别的地方。由图中可见,UNet 和 UNet++都存在欠分割和过分割问题,尤其是在不连贯的移植骨区域,分割缺陷更加明显。加入 Tansformer 之后的 Axial-DeepLab,虽然欠分割和过分割问题得到明显改善,但仍有边缘细节处理欠佳,把部分背景区域预测为移植骨区域。相较之下,本文提出的 ResATUNet 无论在过分割还是欠分割问题上处理得都很好,而且在边缘等低级空间特征的细节处理方面更出色,性能得到显著提升,更接近于专业医生所标注的真实分割区域标签,具备了口腔医学的临床意义。



(a) 原始图像 (b) 真实标签 (c) UNet (d) UNet++ (e) Axial-DeepLab (f) 本文算法

图 4 不同模型分割结果可视化

Fig. 4 Visualization of segmentation results of different models

4 结束语

本文提出了一种基于改进的 ResUNet 网络的口腔移植骨 CBCT 图像的自动分割算法。对于稀缺的数据集数量,实验采用了随机旋转、翻转、加入高斯噪声等在线数据增强方法进行数据的扩充。在网络的编码阶段,使用 Axial-Transformer 替换 ResNet 各阶段中的卷积运算,获取图像特征的全局信息。在浅层网络中设计了一个新颖的细节识别模块,其中包括了带有敏感系数的通道敏感注意力和空间注意力,前者用来获取所有通道特征之间的依赖关系,后者用来关注这些通道特征图中感兴趣的区域。在网络的解码阶段,设计了一个多尺度特征融合模块,消融实验结果证明,Mutti-SFM 比单独使用跳跃连接性能更好。在模型训练中,采用二元交叉熵损失和 Tversky 损失混合的损失函数,相较于交叉熵损失和 Dice 损失的组合,更灵活地平衡假阳性和假阴性之间的关系,减少误判区域,提高了分割精度。最终实验表明,本文提出的口腔移植骨区域自动分割算法的性能优于 Axial-DeepLab、Unext 等先进算法,有能力成为口腔临床医生在植骨区域分割任务上的辅助诊断工具,对于进一步研究和探索具有重要的指导意义。

尽管本文提出的 ResAT-UNet 网络已经表现出非常优异的口腔移植骨区域的分割性能,但由于数据集数量有限,针对小目标区域的泛化能力未能进一步验证,计划通过扩充数据集的规模来改善精细分割的能力,后续研究将会在 3D 数据上进行口腔

移植骨区域分割做相应的尝试和探索。

参考文献

[1] 黄嘉筑，林雪峰. 老年下颌牙列缺失患者种植覆盖义齿治疗的临床评估[J]. 华西口腔医学杂志，2019，37(4)：428-432.

[2] 赖红昌，史俊宇. 上颌窦提升术[J]. 口腔疾病防治，2017，25(1)：8-12.

[3] HUANG J，HU J，LUO R，et al. Linear measurements of sinus floor elevation based on voxel-based superimposition of cone beam computed tomography images[J]. Clinical Implant Dentistry and Related Research，2019，21(5)：1048-1053.

[4] STARCH-JENSEN T，JENSEN J D. Maxillary sinus floor augmentation：a review of selected treatment modalities[J]. Journal of Oral and Maxillofacial Research，2017，8(3)：e3.

[5] GERRESSEN M，RIEDIGER D，HILGERS R D，et al. The volume behavior of autogenous iliac bone grafts after sinus floor elevation：a clinical pilot study[J]. Journal of Oral Implantology，2015，41(3)：276-283.

[6] MAZZOCCO F，LOPS D，GOBBATO L，et al. Three-dimensional volume change of grafted bone in the maxillary sinus[J]. International Journal of Oral & Maxillofacial Implants，2014，29(1)：178-184.

[7] OTSU N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems，Man，And Cybernetics，1979，9(1)：62-66.

[8] YEN J C，CHANG F J，CHANG S. A new criterion for automatic multilevel thresholding[J]. IEEE Transactions on Image Processing，1995，4(3)：370-378.

[9] TREMEAU A，BOREL N. A region growing and merging algorithm to color segmentation[J]. Pattern Recognition，1997，30(7)：1191-1203.

[10] WANG C. Research of image segmentation algorithm based on wavelet transform[C]//Proceedings of the 2015 IEEE International Conference on Computer and Communications (ICCC). IEEE，2015：156-160.

[11] LITJENS G，KOOI T，BEJNORDI B E，et al. A survey on deep learning in medical image analysis[J]. Medical Image Analysis，2017，42：60-88.

[12] 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，Switzerland：Springer，2015：234-241.

[13] ÇIÇEK Ö，ABDULKADIR A，LIENKAMP 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，Switzerland：Springer，2016：424-432.

[14] XIAO X，LIAN S，LUO Z，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). IEEE，2018：327-331.

[15] ZHOU Z，RAHMAN SIDDIQUEE M M，TAJBAKHSH N，et al. Unet++：A nested u-net architecture for medical image segmentation[M]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham，Switzerland：Springer，2018：3-11.

[16] HUANG H，LIN L，TONG R，et al. Unet 3+：A full-scale connected unet for medical image segmentation[C]//Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics，Speech and Signal Processing (ICASSP). IEEE，2020：1055-1059.

[17] CHEN L C，PAPANDREOU G，KOKKINOS I，et al. DeepLab：Semantic image segmentation with deep convolutional nets，atrous convolution，and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intellgence，2018，40(4)：834-848.

[18] 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). New York：IEEE，2016：565-571.

[19] DOSOVITSKIY A，BEYER L，Kolesnikov A，et al. An image is worth 16×16 words：Transformers for image recognition at scale[C]//Proceedings of the International Conference on Learning Representations. 2020：122-142.

[20] RAMACHANDRAN P，PARMAR N，VASWANI A，et al. Stand-alone self-attention in vision models[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019：68-80.

[21] WANG H，ZHU Y，GREEN B，et al. Axial-deeplab：Stand-alone axial-attention for panoptic segmentation[C]//Proceedings of the European Conference on Computer Vision. Cham，Switzerland：Springer，2020：108-126.

[22] VALANARASU J M J，OZA P，HACIHALILOGLU I，et al. Medical transformer：Gated axial-attention for medical image segmentation[C]//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham，Switzerland：Springer，2021：36-46.

[23] HWOO S，PARK J，LEE J Y，et al. Cbam：Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018：3-19.

[24] HU J，SHEN L，SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018：7132-7141.

[25] SALEHI S S M，ERDOGMUS D，GHOLIPOUR A. Tversky loss function for image segmentation using 3D fully convolutional deep networks[C]//International Workshop on Machine Learning in Medical Imaging. Cham，Switzerland：Springer，2017：379-387.

[26] VALANARASU J M J，PATEL V M. UNeXt：MLP-based rapid medical image segmentation network[C]//Proceedings of the International Conference on Medical Image Computing and Cumputer Assisted Intervention. Cham Switzerland：Springer Nature，2022：23-33.