

文章编号: 2095-2163(2024)02-0083-07

中图分类号: TP391

文献标志码: A

融合残差结构和注意力机制的下颌骨 CT 图像分割方法

方梓涛, 刘丹, 吴扬东, 何玲

(贵州大学 现代制造技术教育部重点实验室, 贵阳 550025)

摘要: 针对传统深度学习方法在下颌骨 CT 图像分割中存在的问题, 本文提出一种融合残差结构和注意力机制的改进 Unet 网络。将注意力机制融入 Unet 的解码器, 构建上采样注意力模块, 在不同维度上进行平均池化和最大池化, 使网络更加关注下颌骨区域的信息, 增强下颌骨分割能力; 将残差结构融入 Unet 网络的编码器, 解决深度网络训练时的网络退化和梯度消失问题; 采用迁移学习训练的方法, 避免因下颌骨图像数据不足导致的网络收敛慢的问题。对比实验表明, 改进 Unet 网络平均交并比达到 94.68%, 各评价指标均优于 FCN、DeeplabV1 和 SegNet 网络。

关键词: 下颌骨; Unet; 残差结构; 注意力机制; 迁移学习

Mandible CT image segmentation method fused with residual structure and attention mechanism

FANG Zitao, LIU Dan, WU Yangdong, HE Ling

(Key Laboratory of Advanced Manufacturing Technology, Ministry of Education, Guizhou University, Guiyang 550025, China)

Abstract: Aiming at the problems existing in traditional mandible CT image segmentation methods, this paper proposes an improved Unet network that combines residual structure and attention mechanism. By integrating the attention mechanism into Unet decoder, the designed model constructs an upsampling attention module and performs average pooling and maximum pooling in different dimensions, making the network pay more attention to the information of the mandible region and enhance the ability of mandible segmentation. The residual structure is introduced into the encoder to solve the problem of network degradation and gradient disappearance during deep network training. The transfer learning is used to avoid the problem of slow convergence caused by insufficient mandible image data. Comparative experiments show that the Mean Intersection over Union of the improved Unet network reaches 94.68%, and all evaluation indicators are better than FCN, DeeplabV1 and SegNet networks.

Key words: mandible; Unet; residual structure; attention mechanism; transfer learning

0 引言

下颌骨位于面部骨骼的下部, 是 30 多种不同病变的常见部位^[1]。下颌骨的不同解剖区域密度不同, 导致在下颌骨 CT 图像中下颌骨组织的灰度值与面部其他骨组织的灰度值非常相似, 难以确定一个合适的阈值实现有效分离。此外, 由于受到 CT 图像中噪声或金属伪影的影响, 检测图像中经常出现髁突和牙齿部分的弱边缘和假边缘, 且图像中还有牙齿和周围软组织, 均增加了下颌骨分割的难度, 导致传统自动分割方法分割费时且效果不佳^[2-5]。

近年来, 深度学习在医学图像分割中得到广泛应用^[6], 研究者也开始将其应用于下颌骨图像分割

中。Yan 等^[7]提出了基于端到端的可训练深度学习方法, 通过强制执行对称的卷积和反卷积进行下颌骨自动分割; Xue 等^[8]提出了用于头颈部 CT/MR 图像多器官分割的深度混合类神经网络, 集成的卷积神经网络进行不同的初始化, 同时执行分割, 可以获得更有效的图像特征; 卢苇等^[9]改进 Mask-R CNN 网络, 通过复用依次卷积来提高底层特征的利用率, 进行下颌骨 CT 图像分割。然而这些方法不能有效提取下颌骨区域信息, 无法精确分割下颌骨。

Unet 网络^[10]由于其基于全卷积网络的端对端结构, 能够提取不同尺度的特征并进行特征融合, 提高分割精度, 成为医学图像分割的主流网络框架。Qiu 等^[11]提出了一种采用 Unet 架构的方法, 将从 3

作者简介: 方梓涛(1998-), 男, 硕士研究生, 主要研究方向: 医学图像处理与三维重建; 吴扬东(1969-), 男, 博士, 副教授, 主要研究方向: 设计理论与方法; 何玲(1975-), 女, 博士, 副教授, 主要研究方向: 智能康养装备、智能机器人。

通讯作者: 刘丹(1975-), 女, 博士, 副教授, 主要研究方向: 智能制造技术、智能康养装备。Email: dliu@gzu.edu.cn

收稿日期: 2023-02-13

哈尔滨工业大学主办 ◆ 系统开发与应用

个正交平面获得的 2D 分割组合成 3D 分割。Park 等^[12]提出了一种基于 Unet 的下颌骨和上颌骨自动 3D 分割框架,不仅可以从 CT 中准确描绘下颌骨和上颌骨,还可以提高多中心数据的兼容性。然而,3D 深度学习网络结构复杂,参数量巨大,训练时间长,无法满足下颌骨快速分割的要求。而笔者在针对下颌骨的分割实验中发现,Unet 的网络深度对下颌 CT 图像分割不足,难以提取 CT 中复杂的下颌骨区域信息,导致分割效果不佳。

针对 Unet 网络出现的问题,研究者们提出了各种改善医学图像分割效果的 Unet 变体网络^[13-15]。其中,注意力机制的引入可以增强网络中间层信息的利用率,为目标特征分配更多权重以实现目标特征增强。然而,网络深度的增加导致高层语义信息退化,引起梯度消失等问题。此外,由于下颌骨医学图像缺乏公开数据集,自建数据集的数据量小,使用深度网络进行训练时,大量参数无法得到充分训练,难以获得较高的分割精度。迁移学习^[16]的出现,解决了深度学习中数据量过小而无法充分训练神经网络的问题。目前,迁移学习方法在医学图像领域已得到较多应用^[17]。蒋瑞等^[18]采用基于联合特征学习和多重迁移学习的方法,对肝脏病变进行分类,提高了分类的准确率。姚永康^[19]采用对抗式迁移学习,对皮肤病变图像进行了分割,获得了较好的分割效果。迁移学习方法在医学图像领域的应用,为下颌骨 CT 图像分割提供了新的思路。

本文基于 Unet 网络,对下颌骨 CT 图像分割进行研究,针对深度学习方法分割下颌骨时存在的问题,提出了一种融合残差结构和注意力机制的改进 Unet 网络,提高网络提取下颌骨特征的能力;并针对下颌骨领域自建数据集数据量小的问题,采用基于迁移学习的模型训练方法,使网络参数得到充分训练,实现下颌骨区域快速精确分割。

1 Unet 网络改进

1.1 改进网络模型整体结构

Unet 网络包括编码器、解码器以及预测模块 3 部分。编码器通过下采样卷积层对输入图像进行初步特征提取,通过最大池化降低图像尺寸;解码器通过上采样和卷积层对图像进行深层特征提取。在特征提取过程中,将下采样获得的特征和上采样获得的特征进行拼接 (Concatenate) 操作,沿通道进行拼接,实现特征融合。最后通过 1×1 的卷积调整图像的通道数,对图像中的像素点进行分类。

改进的 Unet 网络结构如图 1 所示。将注意力机制引入 Unet 的解码器,在上采样和卷积层之间加入卷积注意力模块 (Convolutional Block Attention Module, CBAM), 构建上采样注意力模块 (Up-CBAM)。将残差结构融入 Unet 的编码器,利用卷积模块 (Conv Block) 和恒等模块 (Identity Block) 替换原编码器中的部分卷积层。CBAM 模块通过调整各通道维度和空间维度的权重,使网络更加关注医学图像中下颌骨区域的特征信息,有效提取下颌骨区域特征,降低噪声影响,强化网络对下颌骨的分割能力。当融入 CBAM 模块使网络层数增加时,残差结构能够有效解决网络模型退化以及梯度消失问题。图像在输入网络前统一将尺寸缩放为 512×512 。

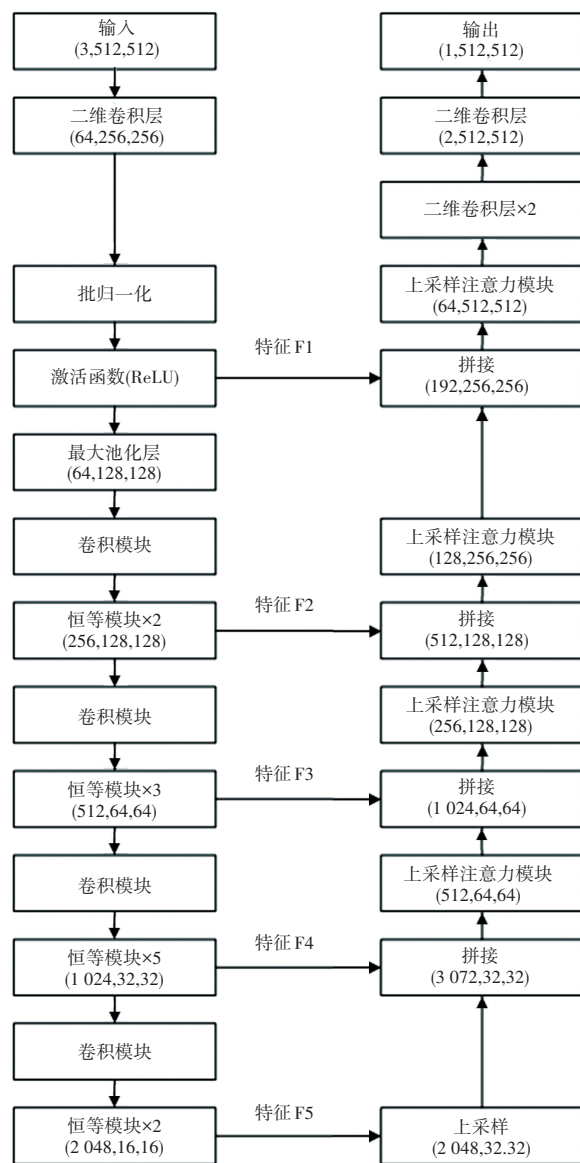


图 1 改进的 Unet 网络结构

Fig. 1 Improve Unet structure

1.2 上采样注意力模块设计

在 Unet 网络的解码器中引入注意力机制, 构建上采样注意力模块(UP-CBAM), 结构如图 2 所示。引入注意力机制可以调整网络的权重, 在训练 Unet 模型的过程中, 对医学图像中下颌骨的特征分配更多权重, 增强网络对下颌骨区域特征的提取能力。对于输入的特征, 先通过 CBAM 模块计算注意力, 然后进行两次卷积, 调整通道数为 $C1$, 最后进行一次上采样, 将高和宽扩大两倍, 获得输出特征 $F_{out} = C1 \times 2H \times 2W$ 。

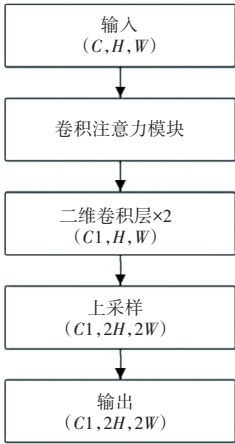


图 2 上采样注意力模块

Fig. 2 Upsampling attention module

卷积注意力模块 CBAM 是一种混合注意力机制, 其中包括通道注意力模块(Channel Attention Module, CAM)和空间注意力模块(Spatial Attention Module, SAM), 结构如图 3 所示。对于输入特征 $F_{in} = C \times H \times W$ (其中 C, H, W 分别为特征图的通道数、特征的宽和高), 首先由通道注意力模块 CAM 计算得到通道注意力权重 $M_c = C \times 1 \times 1$, 与 F_{in} 相乘获得调整后的特征 F_c ; 再由空间注意力模块 SAM 计算得到空间注意力权重 $M_s = 1 \times H \times W$, 与 F_c 相乘获得输出特征 F_{out} 。

1.2.1 通道注意力模块

通道注意力模块结构如图 4 所示。对于输入的特征 F , 首先通过最大池化层(MaxPool)和平均池化层(AvgPool)提取信息, 分别得到两个特征 F_{max}^c 和 F_{avg}^c , 然后将这两个特征输入到多层感知器 MLP (Multilayer Perceptron) 中计算得出结果并相加, 最后经过 Sigmoid 激活函数的映射, 得到通道注意力 M_c 。

将多层感知器中隐含层的大小设置为 $C/r \times 1 \times 1$ (r 为压缩率), 则通道注意力 M_c 的计算公式为 $M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c)))$ (1)

式中: σ 表示 Sigmoid 函数, $W_0 = C/r \times C$, $W_1 = C \times C/r$ 。

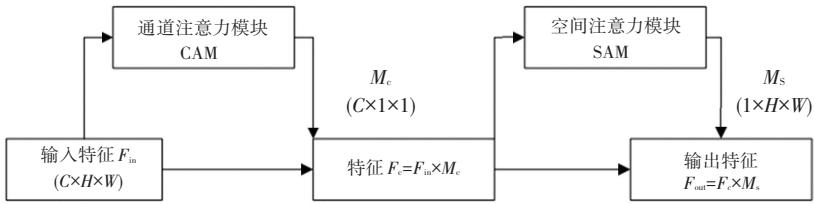


图 3 CBAM 模块

Fig. 3 Convolutional block attention module

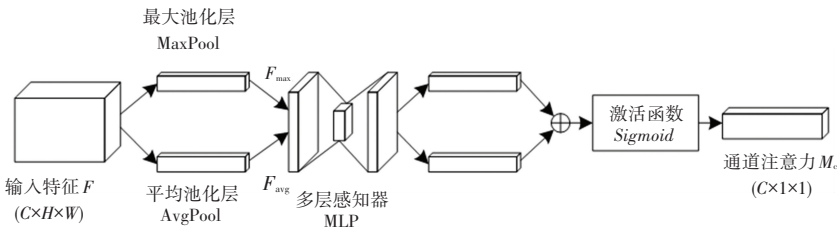


图 4 通道注意力模块

Fig. 4 Channel attention module

1.2.2 空间注意力模块

空间注意力模块结构如图 5 所示。模块输入为经过通道注意力模块的输出特征 $F_c = C \times H \times W$, 分别经过平均池化层和最大池化层得到 $F_{avg}^s = 1 \times H \times W$ 和 $F_{max}^s = 1 \times H \times W$, 然后按照通道拼接, 得到维度

为 $2 \times H \times W$ 的特征; 再经过一次通道数为 1, 卷积核尺寸为 7×7 的卷积运算, 最后经过 Sigmoid 激活函数的映射得到空间注意力 M_s , 其计算公式如下:

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (2)$$

式中: σ 表示 Sigmoid 函数, $f^{7 \times 7}$ 表示卷积核大小为 7×7 的卷积运算。

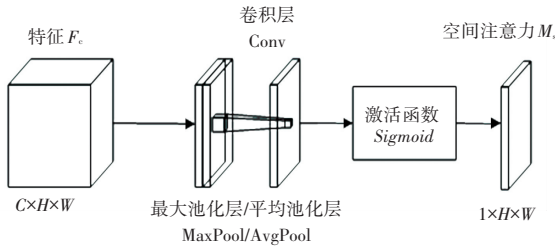


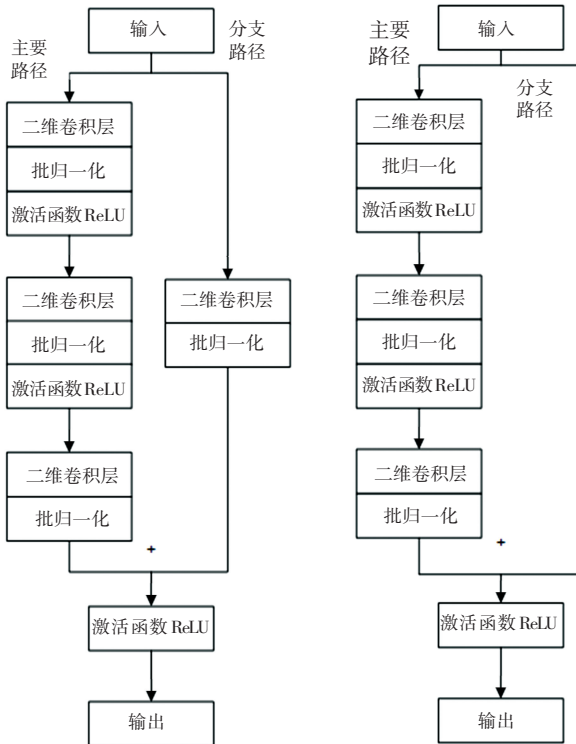
图5 空间注意力模块

Fig. 5 Spatial attention module

1.3 残差结构设计

编码器中的残差结构包含主要路径和分支路径两条路径,通过跳跃连接将两条路径的计算结果相加,再经过激活函数输出,从而解决深层网络退化和梯度消失问题。在恒等模块中,主要路径需依次经过两次二维卷积层、批归一化层、激活函数 ReLU 和一次二维卷积层、批归一化层,而分支路径则直接输出,将两条路径的输出结果相加,再通过 ReLU 激活函数,得到最终输出结果。

卷积模块和恒等模块结构如图6所示。由图中可见,卷积模块与恒等模块的不同在于:在分支路径不再直接输出而是经过一个CB层计算后再输出。



(a) 卷积模块结构 (b) 恒等模块结构

图6 残差结构

Fig. 6 Residual structure

1.4 损失函数

本文将下颌骨图像分割归为二分类问题,下颌骨类别为阳性,背景为阴性。二分类交叉熵损失函数(Cross Entropy Loss, CE)是深度学习二分类问题中常用的损失函数,形式如下:

$$Ce(p, y) = \begin{cases} -\log(p), & \text{if } y = 1 \\ -\log(1-p), & \text{otherwise} \end{cases} \quad (3)$$

其中, $p \in [0, 1]$, 令 $pt = \begin{cases} p, & \text{if } y = 1 \\ 1-p, & \text{otherwise} \end{cases}$,

则:

$$Ce(p, y) = Ce(pt) = -\log(pt) \quad (4)$$

聚焦损失函数(Focal Loss, FL)由二分类交叉熵损失函数演化而来,通过一个动态缩放因子,可以动态降低训练过程中易区分样本的权重,从而能够更加关注下颌骨区域,降低背景区域的影响,形式如下:

$$FL(pt) = -(1-pt)^\gamma \log(pt) \quad (5)$$

其中, γ 为聚焦参数(Focusing Parameter)。

Lin 等^[20]研究了不同 γ 值对损失函数分类性能的影响,得出 $\gamma = 2$ 时性能最佳。因此,本文取 $\gamma = 2$, 即:

$$FL(pt) = -(1-pt)^2 \log(pt) \quad (6)$$

2 实验与分析

2.1 数据集

本文数据来源于贵阳市某口腔医院,选取了1064张下颌骨CT数据,格式为Dicom。首先将该格式转化为JPG格式的图片,大小为 640×640 ,方便后续图片的标注与处理。在专业医生指导下,利用开源标注工具Labelme对图片中下颌骨区域进行标注,获得对应的标签。最后,将获得的数据集按照8:2的比例分割成训练集和验证集,最终训练集大小为851张,验证集大小为213张。

2.2 评价指标

在分割模型常用的评价指标中,选择准确率ACC(Accuracy)、精准率P(Precision)、召回率R(Recall)和平均交并比MIoU(Mean Intersection over Union)进行评估。

$$MIoU = (IoU_1 + IoU_0) / 2 = [TP / (TP + FP + FN) + TN / (TN + FP + FN)] / 2 \quad (7)$$

式中, IoU_1 为下颌骨类别的交并比, $IoU_1 = TP / (TP + FP + FN)$, IoU_0 为背景类别的交并比, $IoU_0 = TN / (TN + FP + FN)$ 。其中, TP 、 TN 、 FP 、 FN 分别为真阳性、真阴性、假阳性、假阴性。则:

$$ACC = (TP + TN) / (TP + TN + FP + FN) \quad (8)$$

$$P = TP / (TP + FP) \quad (9)$$

$$R = TP / (TP + FN) \quad (10)$$

2.3 训练方法

由于下颌骨 CT 图像数量少, 难以满足深度神经网络训练需求, 本文采用迁移学习方法训练改进 Unet 网络: 在 ImageNet 数据集上训练网络的编码

器, 获得预训练权重; 针对解码器没有得到预训练的问题, 训练分为两个阶段: 第一阶段冻结网络的编码器, 保持编码器的网络参数不变, 设置初始学习率, 仅训练解码器网络层参数; 第二阶段解冻编码器, 设置解冻学习率, 训练全部网络层参数。迁移学习过程如图 7 所示。

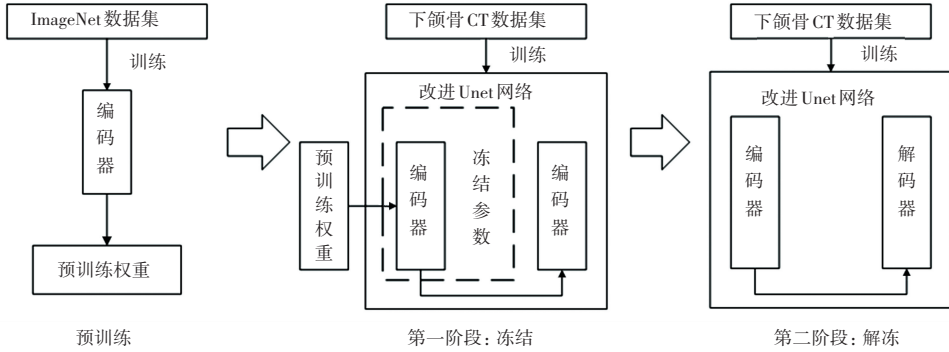


图 7 迁移学习方法

Fig. 7 Transfer learning method

2.4 参数设置

本文实验在 Windows 操作系统下完成, GPU 为 NVIDIA T1200 Laptop, 内存为 64 G。程序实现软件为 Python3.6, Cuda 版本为 11.2.162。同时安装了 Pytorch1.2 和其他第三方库。

总训练周期设置为 100, 冻结训练的周期设置为 25; 训练采用 Adam 优化器优化参数, 冻结阶段的学习率设置为 0.000 1, 解冻阶段的学习率设置为 0.000 01; 批量大小 (batchsize) 设置为 4。

2.5 结果分析

2.5.1 网络实验对比分析

基于下颌骨 CT 图像数据集, 将本文方法与 FCN、DeeplabV1 和 SegNet 3 种网络进行对比测试, 得到评价指标结果见表 1。

表 1 不同网络评价指标表现

Table 1 Evaluation results of different networks

网络	准确率 (ACC)	精准率 (P)	召回率 (R)	平均交并比 (MIoU)
FCN	0.975 5	0.940 5	0.677 4	0.659 3
DeeplabV1	0.984 7	0.975 2	0.797 8	0.784 1
SegNet	0.992 2	0.975 5	0.908 6	0.891 0
Ours	0.995 9	0.976 8	0.975 1	0.944 8

由表 1 可以看出, 改进的 Unet 网络表现优于其他方法, 其中召回率达到 97.51%, 相较 FCN、DeeplabV1 和 SegNet 分别提高 28.85%、16.07% 和 5.38%; 平均交并比达到 94.48%, 相较 FCN、

DeeplabV1 和 SegNet 分别提高 29.77%、17.73% 和 6.65%。结果表明, 融入残差结构和注意力机制并采用迁移学习方法, 能够有效提升 Unet 网络提取 CT 图像中下颌骨信息的能力, 提高了下颌骨分割精度。

各网络在下颌骨 CT 图像数据集上的分割效果见表 2。

表 2 不同网络下颌骨分割效果对比

Table 2 Comparison of mandible segmentation effects of different networks

标签	FCN	DeeplabV1	SegNet	文献[6]	本文方法

在下颌骨区域较大(见表 2 第一、三行)的 CT 图像中, 其他网络不能完整地分割出下颌骨区域, 出现了大面积缺失; 在下颌骨区域较小(见表 2 第二、

四行)的CT图像中,其他网络的分割结果在下颌骨区域周围出现了较多噪声;在图像整体清晰度较差(见表2第五行)的CT图像中,其他网络对下颌骨区域分割效果较差,部分网络完全没有分割出下颌骨区域。

而对于以上不同位置及成像质量的CT图像,改进的Unet网络均能较为准确地分割出下颌骨区域,最接近标签分割结果,本文方法的分割效果均明显优于其他方法。由此可见,融入注意力机制后,网络更加关注下颌骨区域的信息,学习了更多下颌骨

特征,减少了金属伪影等噪声的影响;而残差结构有效解决了因网络层数增加导致的网络退化及梯度消失问题;采用迁移学习,避免了下颌骨CT图像数据不足导致的网络训练收敛慢甚至无法收敛的问题。

2.5.2 消融实验

通过消融实验对改进Unet网络进行验证,分别讨论残差结构Resnet、注意力机制CBAM和采用迁移学习TL(Transfer Learning)对下颌骨分割精度的影响。在构建的下颌骨CT图像数据集上的消融实验结果见表3。

表3 消融实验

Table 3 Ablation experiments

模型				评价指标			
Unet	Resnet	CBAM	TL	准确率(ACC)	精准率(P)	召回率(R)	平均交并比(MIoU)
√	-	-	-	0.987 6	0.951 7	0.861 4	0.833 0
√	√	-	-	0.993 7	0.961 1	0.946 9	0.915 1
√	-	√	-	0.994 7	0.963 2	0.973 1	0.932 0
√	-	-	√	0.995 1	0.965 7	0.974 3	0.937 4
√	√	√	-	0.995 2	0.957 5	0.974 8	0.936 0
√	√	-	√	0.995 1	0.961 1	0.969 1	0.934 4
√	-	√	√	0.995 5	0.967 4	0.967 0	0.937 0
√	√	√	√	0.995 9	0.976 8	0.975 1	0.944 8

从表3可以看出,Unet网络分别融入残差结构和注意力机制及采用迁移学习后,相较原始Unet网络,召回率分别提高8.85%、11.17%和11.29%;平均交并比分别提高8.21%、9.90%和10.44%表明残差结构、注意力机制和迁移学习均能增强网络性能。同时融入残差结构和注意力机制,并采用迁移学习后,精准率达到97.68%,召回率达到97.51%,平均交并比达到94.48%,相较原始Unet网络分别提升2.51%、11.37%和11.18%。实验结果表明,残差结构、注意力机制和迁移学习能够相互促进,在其共同作用下,有效提升了Unet网络对下颌骨区域的分割能力。

3 结束语

本文针对人体下颌骨CT图像分割开展研究,基于Unet网络进行改进,融入残差结构和注意力机制。注意力机制使网络更加关注下颌骨区域信息,能够学习到更多下颌骨特征,降低了噪声对分割的影响;而残差结构解决了深度网络训练时的网络退化和梯度消失问题;采用基于迁移学习的模型训练方法,解决下颌骨CT图像数据量不足的问题,避免了训练模型收敛慢、无法收敛,提升了网络分割下颌

骨的性能。在下颌骨CT图像数据集上与其他方法进行了对比,本文所提方法的分割效果优于其他方法,显著提高了下颌骨CT图像分割的准确率,有助于提升医生在临床诊断中的效率。

参考文献

- [1] DUNFEE B L, SAKAI O, PISTEY R, et al. Radiologic and pathologic characteristics of benign and malignant lesions of the mandible[J]. Radiographics, 2006, 26(6): 1751-1768.
- [2] QIU B, VAN DER WEL H, KRAEIMA J, et al. Automatic segmentation of mandible from conventional methods to deep learning—a review[J]. Journal of Personalized Medicine, 2021, 11(7): 629.
- [3] TOROSDAGLI N, LIBERTON D K, VERMA P, et al. Robust and fully automated segmentation of mandible from CT scans [C]//Proceedings of 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, 2017: 1209-1212.
- [4] SPAMPINATO C, PINO C, GIORDANO D, et al. Automatic 3D segmentation of mandible for assessment of facial asymmetry [C]//Proceedings of 2012 IEEE International Symposium on Medical Measurements and Applications. IEEE, 2012: 1-4.
- [5] WANG L, GAO Y, SHI F, et al. Automated segmentation of dental CBCT image with prior-guided sequential random forests [J]. Medical Physics, 2016, 43(1): 336-346.
- [6] SHEN D, WU G, SUK H I. Deep learning in medical image analysis[J]. Annual Review of Biomedical Engineering, 2017, 19: 221-248.

- [7] YAN M, GUO J, TIAN W, et al. Symmetric convolutional neural network for mandible segmentation [J]. Knowledge - Based Systems, 2018, 159: 63-71.
- [8] XUE J, WANG Y, KONG D, et al. Deep hybrid neural-like P systems for multiorgan segmentation in head and neck CT/MR images [J]. Expert Systems with Applications, 2021, 168: 114446.
- [9] 卢苇, 刘丹, 邵敏, 等. 改进 Mask R-CNN 网络在医学图像识别与分割中的应用[J]. 计算机工程与应用, 2021, 57(24): 234-241.
- [10] OLAF R, PHILIPP F, THOMAS B. U-Net: Convolutional networks for biomedical image segmentation [J]. arXiv preprint arXiv:1505.04597, 2015.
- [11] QIU B, GUO J, KRAEIMA J, et al. Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network [J]. Physics in Medicine & Biology, 2019, 64(17): 175020.
- [12] PARK S, KIM H, SHIM E, et al. Deep learning-based automatic segmentation of mandible and maxilla in multi-center CT images [J]. Applied Sciences, 2022, 12(3): 1358.
- [13] 黄晓鸣, 何富运, 唐晓虎, 等. U-Net 及其变体在医学图像分割中的应用研究综述[J]. 中国生物医学工程学报, 2022, 41(5): 567-576.
- [14] JIANG J, SHARIF E, UM H, et al. Local block - wise self attention for normal organ segmentation [J]. arXiv preprint arXiv: 1909.05054, 2019.
- [15] SUN S, LIU Y, BAI N, et al. Attention anatomy: A unified framework for whole - body organs at risk segmentation using multiple partially annotated datasets [C] // Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020: 1-5.
- [16] YANG X, LI F, LIU H. TTL-IQA: Transitive transfer learning based no - reference image quality assessment [J]. IEEE Transactions on Multimedia, 2020, 23: 4326-4340.
- [17] 高爽, 徐巧枝. 迁移学习方法在医学图像领域的应用综述 [J]. 计算机工程与应用, 2021, 57(24): 39-50.
- [18] 蒋瑞, 刘哲, 宋余庆, 等. 基于联合特征学习和多重迁移学习的肝脏病变分类 [J]. 江苏大学学报(自然科学版), 2021, 42(5): 554-561, 568.
- [19] 姚永康. 基于对抗式迁移学习的皮肤病变图像分割方法研究 [D]. 西安: 西京学院, 2021.
- [20] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] // Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017: 2999 - 3007.