

# 基于 Transformer 特征通道融合的舌像分割

薛玮珠<sup>1</sup>, 张博<sup>2</sup>, 姚瑶<sup>2</sup>, 熊玉洁<sup>2</sup>, 夏春明<sup>2</sup>

- 浙江机电职业技术学院 智能制造学院, 浙江 杭州 310053;
- 上海工程技术大学 电子电气工程学院, 上海 201620

收稿日期: 2023-06-26

基金项目: 国家自然科学基金(62006150), 上海市科技创新行动计划技术标准项目(21DZ2203100)

第一作者: 薛玮珠, 女, 副教授, 主要从事智能设计方面的研究。E-mail: xwz720118@163.com

**摘要:** 针对舌像分割过程中, 舌体边缘不连续, 复杂背景干扰舌像等问题, 提出一种基于 Transformer 特征通道融合的舌像分割方法。首先通过多级特征融合模块对舌像进行卷积, 得到具有位置和特征信息的特征图; 然后引入倒置特征金字塔网络模块匹配特征维度; 最后将 U-Net 网络的跳跃连接替换成 UCTransNet 的 CTrans 模块, 进行舌像的特征通道融合, 减少背景信息特征的干扰, 实现图像的准确分割。本文选取了 Dice 值、平均交并比(MIoU)作为评价标准, 通过在自采集的舌像数据集上进行训练评估和验证, Dice 值为 96.81%, MIoU 值为 93.89%。这表明本文方法在舌像数据集上具有较好的分割效果, 可以准确提取舌体特征; 本文方法可用于舌诊的标准化研究, 提高舌诊的准确性和可靠性, 且在其他医疗图像数据集上的泛化能力较强。

**关键词:** 深度学习; 图像分割; Transformer; 舌诊

中图分类号: TP319; R318

文献标志码: A

文章编号: 1671-8836(2024)06-0704-11

## Tongue Image Segmentation Based on Transformer Feature Channel Fusion

XUE Weizhu<sup>1</sup>, ZHANG Bo<sup>2</sup>, YAO Yao<sup>2</sup>, XIONG Yujie<sup>2</sup>, XIA Chunming<sup>2</sup>

- School of Intelligent Manufacturing, Zhejiang Institute of Mechanical and Electrical Engineering, Hangzhou 310053, Zhejiang, China;
- College of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

**Abstract:** To address challenges in tongue image segmentation, such as discontinuous tongue edges and interference from complex backgrounds, this paper proposes a traditional Chinese medicine tongue image segmentation method based on Transformer feature channel fusion. First, multi-level feature maps containing both positional and feature information are obtained through a multi-stage convolution module. Next, an inverted pyramid network is introduced to match the dimensions of the multi-level feature maps. Finally, the skip connections of the traditional U-Net network are replaced with the CTrans module of UCTransNet to capture contextual information in the image better and achieve accurate segmentation of medical images. Dice coefficient and mean intersection-over-union (MIoU) are selected as evaluation criteria. Training, evaluation, and validation on a self-collected dataset of tongue images yield a Dice value of 96.81% and an MIoU value of 93.89%, indicating strong segmentation performance. The proposed method has a good segmentation effect on the tongue image dataset, and can accurately extract the features of the tongue body. This method can be used for standardized research in tongue diagnosis, improving the accuracy and reliability of tongue diagnosis. Additionally, it demonstrates strong generalization capabilities on other medical image datasets.

**Key words:** deep learning; image segmentation; Transformer; tongue diagnosis

引用格式: 薛玮珠, 张博, 姚瑶, 等. 基于 Transformer 特征通道融合的舌像分割[J]. 武汉大学学报(理学版), 2024, 70(6): 704-714. DOI: 10.14188/j.1671-8836.2023.0134.

XUE Weizhu, ZHANG Bo, YAO Yao, et al. Tongue Image Segmentation Based on Transformer Feature Channel Fusion [J]. J Wuhan Univ (Nat Sci Ed), 2024, 70(6): 704-714. DOI: 10.14188/j.1671-8836.2023.0134(Ch).

## 0 引言

图像分割是计算机视觉领域中的一项重要任务,常用于自动驾驶、工业检测、医疗诊断领域。在医疗诊断领域中主要用来辅助医生诊断患者身体情况。中医舌诊是中医辨证论治的主要依据之一,在中医诊断学中具有重要地位<sup>[1~3]</sup>。医生通过观察舌体的形态、颜色、质地等特征,判断患者体内的病理变化和脏腑功能状态。传统的中医舌诊受到操作者与操作环境等多种因素影响,具有主观性和不准确性<sup>[4~7]</sup>。随着科技的不断发展和中医的现代化,在医学图像处理领域中,舌像分割广泛应用于临床诊断、口腔疾病检测,为舌诊分析提供了可靠的辅助<sup>[8~10]</sup>。

传统的舌像分割方法主要基于图像处理技术。舌像通常具有明显的颜色差异,可以通过设置一个固定的颜色阈值将舌体和背景分割开,阈值分割算法相对简单,实现起来比较容易,不需要复杂的计算和训练过程。姜朔等<sup>[11]</sup>在处理舌像时使用了Otsu阈值法进行二值化处理,结合先验知识进行聚类,以及形态学修正方法,使舌体自适应地选取结构元素的大小,修正分割范围,较好地实现了舌像的分割。然而,这种方法对光照和噪声敏感,难以处理复杂舌像情况。

在舌像分割的过程中,舌体具有较强的边缘信息,可以通过滤波来检测舌体的边缘,并据此将舌体和背景进行分割。傅之成等<sup>[12]</sup>提出一种结合径向边缘检测与Snake模型的分割方法,首先使用径向边缘检测得到初始轮廓,其次通过色彩对消除嘴唇影响,再通过形态滤波,高斯滤波和设定图像权重,最后获得输出结果,但滤波的选择针对不同的舌像存在主观性,需要多次尝试以获取较好的边缘检测效果。

舌体具有相对连续的区域,且舌体和背景之间存在较大的颜色或亮度差异,可以通过区域生长算法来实现舌像的分割。首先选择一个舌体种子点,然后根据一定的生长准则,将与种子点相邻且颜色或亮度相似的像素逐步合并到舌体区域,直到舌体完全生长出来。黄展鹏等<sup>[13]</sup>首先通过均值偏移对图像进行预处理,然后利用最大类间方差法对彩色舌像进行自动分类,最后基于分类后的子块特征进行区域生长,实现舌像中舌体区域的自动分割。区域生长算法基于像素的相似性进行生长,对于目标不连续的情况,如舌像中的裂纹、齿痕<sup>[14~17]</sup>等,区域生长算法无法有效处理。

近年来,深度学习在医学影像领域取得了显著的进展,其中基于深度学习的分割方法引人注目。U-Net<sup>[18]</sup>网络是一种常用的分割方法,它是由卷积神经网络(Convolutional Neural Networks, CNNs)<sup>[19~23]</sup>构成,具有对称的编码器和解码器结构,可以在分割任务中有效地捕捉到局部和全局的特征信息,以跳跃连接的方式拼接多级特征,生成分割结果,但U-Net的跳跃连接模块存在一定的局限性,并不是每个跳跃连接都是有效的,甚至有些跳跃连接会对分割性能产生负面影响。Wang等<sup>[24]</sup>提出一种从Transformer的通道级视角重新思考U-Net中的跳跃连接的方法UCTransNet,但依然存在高层特征和底层细节特征无法有效结合的问题。

传统的舌像分割方法简单易实现,利用舌体的边缘信息,舌体和背景之间的颜色或亮度差异实现分割,缺点是滤波的选择存在主观性,分割结果受实际图像质量的影响较大。通过融合浅层和深层特征图的多级特征融合模块和CTrans模块中的CCA(Channel-wise Cross-Attention)模块可以降低背景对分割结果的干扰。CTrans模块中的CCT(Channel Cross fusion with Transformer)模块在融合局部特征方面比传统的跳跃连接更好,更适合边缘连续,位于图像中心的舌像。

本文针对现有方法分割舌体边缘效果差,背景复杂等问题,提出一种改进的舌像分割方法,通过引入多级特征融合(Multi-Level Feature Fusion, ML)和倒置特征金字塔(Reverse Feature Pyramid Network, RFPN)模块(倒置特征金字塔网络模块是对特征金字塔(Feature Pyramid Network, FPN)<sup>[25]</sup>的改进),用UCTransNet中CTrans模块替代U-Net的跳跃连接,将更加丰富的特征适当地组合到深度特征图中,以提高舌像分割的准确性和鲁棒性。

## 1 相关工作

### 1.1 特征金字塔网络

特征金字塔网络是一种用于图像分割和目标检测任务的深度学习网络结构,FPN的设计灵感来自于图像金字塔的概念,它的主要目标是在不同尺度的特征图上有效地提取图像特征<sup>[26~29]</sup>。这种金字塔结构使得网络能够同时处理不同尺度的物体,并在保留全局上下文信息的同时捕获局部细节,从而提升了物体检测和分割的准确性。它的多尺度特征表示使得网络能够对不同大小和尺度的物体进

行准确检测和分割,同时具有较强的鲁棒性和泛化能力。因为这些优点,FPN已成为深度学习领域中一种重要的网络结构,并广泛应用于许多计算机视觉任务中。

### 1.2 U-Net

U-Net是一种用于图像分割的卷积神经网络结构,它由编码器和解码器两部分组成。编码器负责将输入图像通过一系列的卷积和池化操作转换为具有高级语义信息的特征图。解码器则负责对编码器得到的特征图进行逐步的上采样,并通过反卷积操作还原为与原始输入图像相同尺寸的分割结果。这种编码器-解码器的对称结构使得网络能够逐层提取图像的低级到高级特征,并将这些特征映射到分割结果中<sup>[30-33]</sup>。因此,U-Net在许多图像分割应用中表现出色。然而传统U-Net使用简单的跳跃连接方式,直接将编码器和解码器的特征图进行拼接,这种方式会导致特征图之间的信息传递不充分,影响分割性能。

### 1.3 UCTransNet

为了解决由于编码器和解码器阶段特征图不匹配的问题,UCTransNet通过引入Transformer模型重新设计U-Net中的跳跃连接,并提出了一个CTrans模块,CTrans模块里包括两个子模块,分别是通道交叉融合(CCT)模块和通道交叉注意力(CCA)模块。CTrans通过通道维度上的注意力机制实现特征图之间的信息传递,替代原有的跳跃连接,经过两个子模块引导融合后的多尺度通道信息与解码器特征之间连接更有效。这种方式允许网络进行更加灵活和全面的建模,从而更好地捕捉图像中的上下文信息,实现医学图像的准确自动分割。

## 2 本文方法

本文沿用U-Net的网络架构,首先在U-Net前加入多级特征融合模块,将更加丰富鲜明的特征融合到深度特征图中,提高模型的特征表达能力,增强鲁棒性;其次,将经过多级融合处理的输出图像输入到对应的RFPN网络层中,有效地提取不同层级中的多尺度特征图;最后,用UCTransNet中CTrans模块替代U-Net的跳跃连接和解码器,获得分割图。CCT和CCA模块能够解决语义层次不一致的问题,有利于复杂医学图像的分割。舌像分割网络的整体框架如图1所示。

UCTransNet网络由编码器、跳跃连接和解码

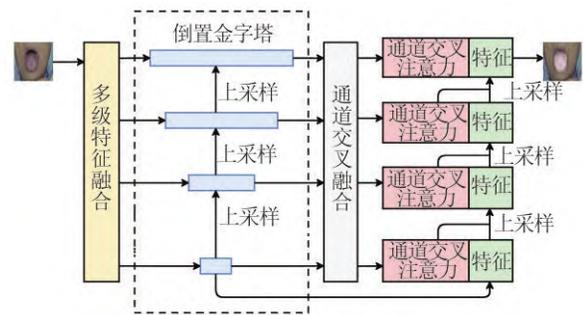


图1 方法整体结构

Fig. 1 Overall structure of the method

器组成,编码器端则是原始的U-Net网络的编码器,负责提取特征信息,跳跃连接为CCT模块,解码器为CCA模块,跳跃连接和解码器统称为CTrans模块,如图2(a)所示。我们主要针对编码器端进行了改进,增加了两个子模块,分别是多级特征融合模块以及倒置金字塔模块。多级特征融合模块可提取更丰富的位置和特征信息,倒置特征金字塔模块可提取不同层级中的特征图,匹配后续模块的输入维度,承上启下,如图2(b)所示。

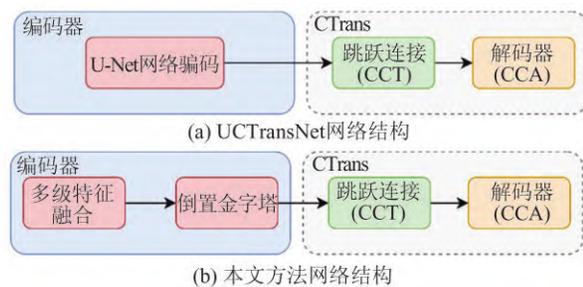


图2 UCTransNet与本文方法网络结构

Fig. 2 The network structure of UCTransNet and our method

### 2.1 多级特征融合

以往的注意力U-Net是通过简单的裁剪和复制操作完成的信息组合,将位置和特征信息集成到高维特征图中,这样简单的连接操作会导致有意义的信息丢失。然而,经过每一级处理的图像特征仍然存在着丰富的位置和特征信息,因此有必要对其进行改进。为了提取更多有效特征,需将多级融合作用于跳跃连接之前,即进行多层拼接和卷积运算。舌像分割任务主要面临的挑战和难点在于边界模糊,复杂背景干扰。浅层特征能够学习低级纹理和边缘信息,融合后的浅层特征包含不同尺度下的边缘特征,可以减少边界模糊的影响;深层特征通过学习抽象和高级的语义信息,对背景进行有效的抑制,不同尺度的浅层特征和深层特征在经过融合后,增强了模型表达能力,有助于模型更好地区分目标与背景,减少背景对分割结果的干扰。多级

特征融合模块如图3中间红色虚线左半所示。

首先,输入舌像经过一层卷积神经网络后,形成一个宽×高为224×224、通道数为64的特征图 $X_1$ ,对 $X_1$ 进行四次逐级卷积操作,前三次使其特征图宽高变为原来的一半,通道数变为原来的两倍。第四次卷积操作仅改变特征图的宽高,为上一层特征图的1/2,而通道数不变。其次,对前四个阶段的特征图,进行上采样和融合操作,得到新的特征图。最后,各层分别输出多级特征融合后的特征图, $Q_1, Q_2, Q_3$ 为中间特征图,起到连接前后特征图的作用。多级特征融合的目标是通过将来自不同层次的特征进行融合,充分利用它们的互补性,得到更丰富的位置信息和更鲜明的判别性特征表示。

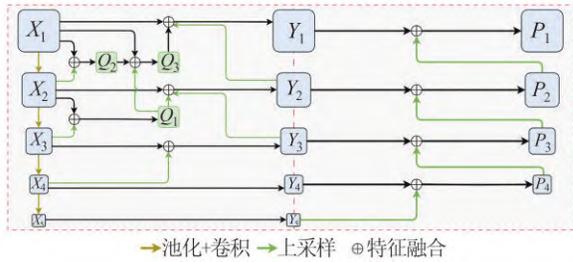


图3 多级特征融合(左半)和倒置特征金字塔模块(右半)  
Fig. 3 Multi-level feature fusion(left) and reverse feature pyramid network(right)

### 2.2 倒置特征金字塔网络(RFPN)

为了有效地提取不同级层中的多尺度特征图信息,匹配CTrans的CCT模块的输入维度,本文采用RFPN网络结构对特征图进行处理,这种结构在处理多尺度的特征图时,能够保留细节信息并提高模型性能。RFPN模块如图3中间红色虚线右半所示。RFPN网络的计算过程是从下到上的,输入图像的大小自下到上逐渐增大,左边的每一层特征图都与右边的对应层进行横向连接。横向连接模块将自底向上的倒置特征金字塔网络的路径与自上而下的多级融合网络的路径进行连接,有效增强高层特征。同级网络层的特征图大小相同,上下层的特征图通过特征融合的方法使上采样映射与对应映射进行合并,通过迭代计算,生成最终的分类分割图像。

RFPN网络的流程如下:首先,对输入的最后一层特征图 $Y_5$ 进行上采样操作,与 $Y_4$ 特征融合得到 $P_4$ ,并作为RFPN的最后一层输出;然后将融合后的特征图 $P_4$ ,与上一层的输入特征图 $Y_3$ 融合,得到 $P_3$ ,以此类推;最后得到各阶段融合后的特征图,共计4层。相比于FPN,RFPN能够更加灵活地匹配输入输出,在不损失特征信息的情况下,减少输出特征

图的数量,为后续CCT的输入提供可靠的特征信息。

### 2.3 通道交叉融合(CCT)模块

为了解决之前提到的跳跃连接问题,本文采用UCTransNet的通道交叉融合模块。舌体图像是连续区域,复杂程度比腺体、多器官图像高,常在图像的中心位置,通道交叉融合模块在融合局部特征方面比传统的跳跃连接更好,更适合边缘连续、位于图像中心的舌像。本文未对通道交叉融合模块网络结构作改变,其结构如图4所示。

该模块由三个部分组成:多尺度特征编码、多头交叉注意力和多层感知机(MLP)。首先将四个跳跃连接的输出 $E_i \in \mathbb{Q}^{\frac{HW}{r^i} \times C_i}, i=1,2,3,4$ ,标记并重建为二维的patch序列,patch的大小分别为 $P, \frac{P}{2}, \frac{P}{3}, \frac{P}{4}$ ,然后将序列化的patch结合起来,作为键(Key)和值(Value)输入多头交叉注意力模块中,最后,序列特征被送入带有残差结构的多层感知机。

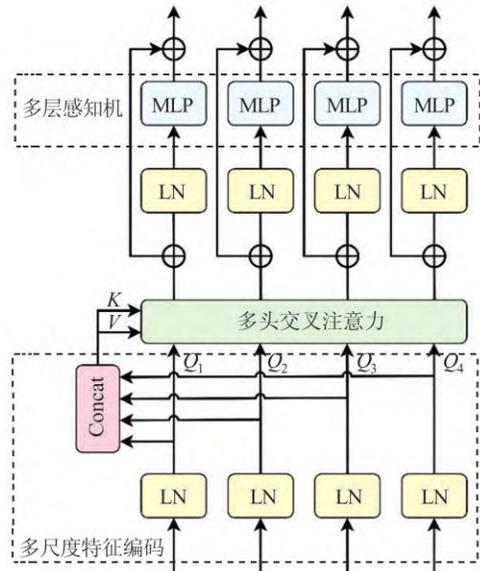


图4 CCT模块

Fig. 4 Channel-wise cross fusion transformer for encoder feature transformation

### 2.4 通道交叉注意力(CCA)模块

为了更好地融合通道Transformer与U-Net解码器之间的特征,本文在原始网络中添加了一个基于Transformer的通道交叉注意力模块,该模块能够合理利用Transformer的信息过滤特征,消除与解码器特征的歧义。采集舌像的地方包括病患家中,医院等,获取难度高,采集人员的拍摄方式不同,产生的舌像质量有区别,背景干扰问题是其中一个难

点,舌分割任务不仅需要分割出连续的舌体,而且要将舌体与背景分隔开,通道交叉注意力模块可以减少背景信息特征的干扰,使网络更好地理解多个输入之间的关系,辅助模型更好地捕捉多个输入之间的相关性,帮助算法更好地捕捉图像中边缘、轮廓等细节特征,提高分割精度。通道交叉注意力模块有三个组成部分,第一个部分是输入层,目的是接收输入数据,这些数据包含图像位置、空间等特征信息;第二部分是交叉注意力层,用来计算输入数据之间的相关性;第三部分是输出层,用来输出最终的分割结果。

基于 Transformer 的通道交叉注意力模块如图 5 所示。首先将第  $i$  层的 Transformer 输出  $O_i \in \mathbb{Q}^{C \times H \times W}$  与第  $i$  层的解码器特征图  $D_i \in \mathbb{Q}^{C \times H \times W}$  进行上采样和平均池化,得到  $M_i$ ,如(1)式:

$$M_i = L_1 \cdot G(O_i) \oplus L_2 \cdot G(D_i) \quad (1)$$

其中,  $L$  表示线性层,  $G$  表示平均池化。然后经过激活函数再与  $O_i$  叠加在一起,得到 CCA 的输出  $\hat{O}_i$ ,如(2)式所示:

$$\hat{O}_i = \sigma(M_i) \otimes O_i \quad (2)$$

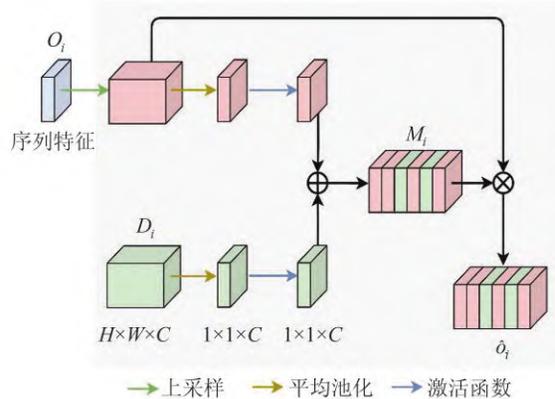


图 5 CCA 模块

Fig. 5 Channel-wise cross attention for feature fusion in decoder

## 2.5 损失函数

为了确保本文方法在训练时的准确性,采用了交叉熵损失函数(cross-entropy loss function)作为模型的损失函数度量模型预测结果与真实标签之间的差异,交叉熵损失函数如(3)式所示:

$$CE \text{ Loss} = -\frac{1}{N} \sum_n \sum_i y_{\text{true}}^{(n)}(i) \log(y_{\text{pred}}^{(n)}(i)) \quad (3)$$

其中,  $N$  表示批量中的样本数,  $y_{\text{true}}^{(n)}(i)$  表示第  $n$  个样本的实际标签中第  $i$  个类别的值,  $y_{\text{pred}}^{(n)}(i)$  表示第  $n$  个样本模型预测的第  $i$  个类别的概率。

## 3 实验和结果分析

### 3.1 数据集

在医生的协助下,用手机拍照的方式采集数据,经过挑选后,最终得到包含 100 张图片的舌像数据集(图片分辨率在  $984 \times 803$  到  $1436 \times 885$  之间),文件的后缀格式为 PNG。通过专业人员对数据集标注后,单个样本包含一张原图片和一张经过 Mask 后的图片,以 Mask 图片作为训练数据的标签。100 张图片以 8:2 的比例分为训练集和测试集,两个集合之间无重复数据,保证实验的可靠性。

为了验证该方法的泛化能力,实验选取了医学开源数据集腺体分割数据集(Gland Segmentation, GlaS)<sup>[34]</sup>和多器官核分割数据集(Multiple organ Nuclear Segmentation, MoNuSeg)<sup>[35]</sup>评估本文所提的方法。腺体分割数据集是用于训练深度学习模型的一种数据集,旨在识别和分割腺体图像中的腺体细胞结构。该数据集包含来自不同组织类型的腺体图像,每个图像都有标记的腺体结构,共计 165 张图片。MoNuSeg 数据集是一个多器官核分割数据集,由微观图像组成,该数据集包含 7 个器官的图像,涵盖了来自不同组织的多种细胞类型,包括细胞核、细胞膜、细胞质和细胞器等组织结构,共计 44 张图片。

### 3.2 实验环境

在 Windows 10 操作系统上进行实验,使用 Python 作为主要开发语言,PyTorch 作为深度学习框架,计算机的硬件配置包括 Intel Core i7-9900K CPU, NVIDIA GeForce RTX 3080 GPU(10 GB)和 32 GB 内存。开发环境采用 PyCharm Community Edition 2022.1.3, CUDA11.7。

### 3.3 评价指标

为了量化方法的性能并比较不同算法,实验采用了两个广泛应用的图像分割评估指标,分别是 Dice 值和 MIoU(Mean Intersection over Union)。

Dice(Sørensen-Dice)值用于衡量预测分割结果与真实标签之间的相似度,如(4)式所示:

$$\text{Dice} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \times 100\% \quad (4)$$

MIoU 值用于度量预测分割结果与真实标签之间的重叠程度,如(5)式所示:

$$\text{MIoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \times 100\% \quad (5)$$

其中, TP 表示真正例(True Positives),即预测结果与真实标签都为正例的像素数; FP 表示假正例

(False Positives),即预测结果为正例但真实标签为负例的像素数;FN表示假负例(False Negatives),即预测结果为负例但真实标签为正例的像素数。Dice值和MIoU值越接近100%表示预测结果与真实标签越相似。

### 3.4 参数设置

在优化模型性能的过程中,进行了超参数的调整,训练阶段参数设置如表1所示。

表1 模型参数设置一览表

Table 1 Overview of model parameter settings

参数名称	数值设置
迭代次数	120
学习率	0.0001
批量大小	4
优化器	Adam
图片调整后大小	224×224
Python哈希数	666

### 3.5 模型速度比较

模型的速度是评估其性能和实际应用的重要指标之一,实验测试了不同模型的速度,结果如图6所示。从箱型图6中橙色的中位线可以看出,U-Net的画面平均传输速度为20帧/s,MedT的平均传输速度为4帧/s,UCTransNet平均传输速度为12帧/s,本文方法平均传输速度为11帧/s。实验中,U-Net的参数数量为 $14.75 \times 10^6$ ,MedT的参数数量为 $98.3 \times 10^6$ ,UCTransNet的参数数量为 $66.43 \times 10^6$ ,本文方法的参数数量为 $85.56 \times 10^6$ 。本文方法的处理速度虽然比MedT慢,但参数数量比MedT少。医学图像常常是以离线的模式进行分析处理的,因此,本文处理速度是可接受的。

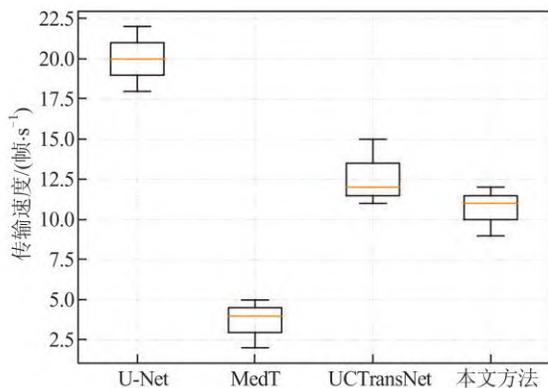


图6 本文方法与其他方法的测试速度对比

Fig. 6 Comparison of testing speed between the proposed method and other methods

### 3.6 消融实验及分析

消融实验中原始网络使用的是U-Net,添加的模块有MF(Multi-level Fusion)、RFPN、CCT以及CCA,共进行了4组实验,表2展示了这4组消融实验的数据结果。

1) 第一组实验采用U-Net作为基础网络。该实验数据包括了U-Net与CCT、CCA单独组合的分割结果,以及同时添加CCT、CCA的U-Net分割结果。在GlaS和MoMuSeg数据集上,同时添加CCT、CCA的结果比分别添加模块的效果好。在舌像数据集中,同时添加CCT、CCA的结果不如基础网络U-Net,这是因为在实际的应用场景中,舌像数据集在数据分布上不如公开的数据集均匀,注意力机制会对特征图集中中出现次数较多的特征产生偏见,导致结果降低。

2) 第二组实验采用了U-Net与RFPN结合的网络作为基础。该实验数据包括了在RFPN中与CCT、CCA分别组合的分割结果,以及同时添加CCT、CCA的RFPN分割结果。在GlaS和MoMuSeg数据集上同时添加CCT、CCA的RFPN表现良好。在舌像数据集中,单独添加CCT后的Dice值,MIoU值分别为96.37%,93.04%,高于以Baseline+RFPN为基础的,与基础网络U-Net的结果接近,说明RFPN+CCT的组合对分割效果是有提升的。

3) 第三组实验采用了U-Net与MF结合的网络作为基础。该实验数据包括了在MF中与CCT、CCA分别组合的分割结果,以及同时添加CCT、CCA的MF分割结果。Baseline+MF+CCT+CCA的Dice值、MIoU值无论在哪个数据集上都高于其他方法。

4) 第四组实验采用了U-Net与MF和RFPN结合的网络作为基础。该实验数据包括了在MF、RFPN中与CCT、CCA分别组合的分割结果,以及同时添加CCT、CCA的MF、RFPN分割结果。在舌像数据集上,Baseline+MF+RFPN+CCT+CCA的Dice值、MIoU值分别为96.81%和93.89%,高于其他方法;在GlaS和MoMuSeg数据集上,Baseline+MF+RFPN+CCT+CCA的Dice值、MIoU值虽然小于舌像数据集,但Baseline+MF+RFPN+CCT+CCA的Dice值、MIoU值高于其他方法。

### 3.7 与现有方法的对比

将本文方法与MedT<sup>[36]</sup>、U-Net<sup>[7]</sup>、UCTransNet<sup>[13]</sup>进行比较,结果见表3。为了比较公平,实验中使用了UCTransNet<sup>[13]</sup>所发布的代码和参数设置。从

表 2 本文方法在不同数据集上的消融实验表

Table 2 Ablation experiments of our proposed method on different datasets

方法		舌像数据集		GlaS		MoNuSeg	
		Dice	MIoU	Dice	MIoU	Dice	MIoU
第一组	Baseline(U-Net) <sup>[7]</sup>	96.47	93.26	85.45	74.78	76.45	62.86
	Baseline+CCT	95.38	91.44	89.09	80.78	79.31	65.97
	Baseline+CCA	94.86	90.84	87.09	78.10	76.84	63.85
	Baseline+CCT+CCA	96.00	92.42	89.84	82.24	79.87	65.63
第二组	Baseline+RFPN	96.10	92.57	87.76	80.92	79.04	65.63
	Baseline+RFPN+CCT	96.37	93.04	89.35	81.22	79.64	66.03
	Baseline+RFPN+CCA	95.80	92.02	89.78	82.03	78.78	64.83
	Baseline+RFPN+CCT+CCA	95.46	91.42	90.04	83.56	79.96	66.98
第三组	Baseline+MF	96.25	92.89	86.75	78.64	78.31	63.67
	Baseline+MF+CCT	94.98	90.72	88.23	81.04	78.63	65.47
	Baseline+MF+CCA	95.73	91.97	87.98	81.21	78.21	63.96
	Baseline+MF+CCT+CCA	96.63	93.57	89.92	83.02	79.87	65.92
第四组	Baseline+MF+RFPN	95.86	92.21	88.01	80.97	77.08	63.09
	Baseline+MF+RFPN+CCT	96.26	92.87	89.21	82.02	79.85	66.21
	Baseline+MF+RFPN+CCA	96.59	93.47	88.62	81.73	77.62	64.56
	Baseline+MF+RFPN+CCT+CCA	96.81	93.89	90.85	84.76	80.11	67.15

表 3 中可以看出本文方法在舌像数据集中的结果最好。

图 7 展示了本文方法与其他方法在舌像数据集上的可视化结果,红色方框区域表示该区域的分割结果真值标签有差异。第一行中,MedT 和 UC-TransNet 能完整地分割出舌体边缘,但在舌根区缺少了部分舌面信息。第二行中,对于舌尖区域,U-Net 和 UCTransNet 无法适当地处理尖锐的部分,存在舌面缺失和冗余。第三行中,MedT 和 U-Net 则在舌根区域出现了冗余,把真值中没有的部分给错误地补全了,而 UCTransNet 的分割结果存在明显的缺失。第四行中,U-Net 和 UCTransNet 无法平滑地分割舌体的右边缘。第一行到第四行中,本文

方法的分割结果不仅舌面保持完整统一,而且舌体的边缘清晰平滑。实验数据和可视化结果均表明,本文方法相比于其他方法在舌像分割上具有一定优势。

MedT<sup>[36]</sup>通过门控轴向注意力控制位置信息对局部编码的影响,如果一个相对位置编码被准确学习,相对于那些没有被准确学习的编码,门控机制会赋予它较高的权重。在舌像分割任务中,经过 MedT 推理产生的分割图像真值在舌体附近会产生离散点,而且无法很好地处理部分图像中心,且 MedT 在学习像素间的远程依存关系上较弱。

从表 3 中可看出,在 GlaS 上,本文方法 Dice 值比 U-Net 高 4.51 个百分点,比 UCTransNet 高 1.01

表 3 本文方法与其他分割方法的对比

Table 3 Comparison of our proposed method with the other segmentation methods

方法	舌像数据集		GlaS		MoNuSeg	
	Dice	MIoU	Dice	MIoU	Dice	MIoU
MedT <sup>[36]</sup>	91.86	85.17	87.14	77.33	77.50	64.10
U-Net <sup>[7]</sup>	96.47	93.26	86.34	76.81	73.97	59.42
UCTransNet <sup>[13]</sup>	96.00	92.42	89.84	82.24	79.87	66.68
本文方法	<b>96.81</b>	<b>93.89</b>	<b>90.85</b>	<b>84.76</b>	<b>80.11</b>	<b>67.15</b>

注:黑色粗体数字为最优值。

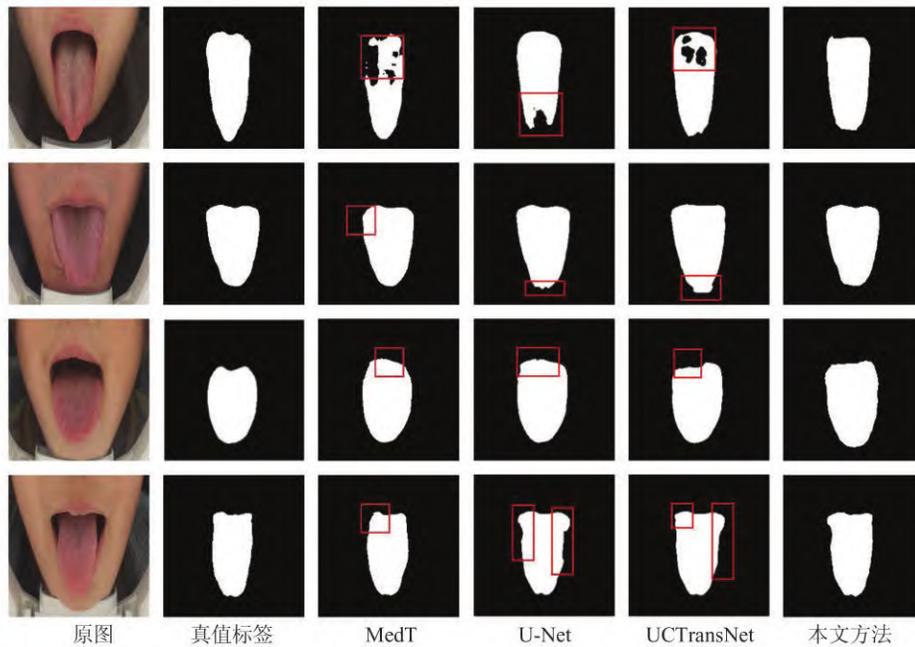


图 7 本文方法与其他方法在舌像数据集中的舌体分割图

Fig. 7 Tongue segmentation results of proposed method and other methods on tongue image dataset

个百分点;MIoU 值比 U-Net 高 7.95 个百分点,比 UCTransNet 高 2.52 个百分点。在 MoNuSeg 上,本文方法 Dice 值比 U-Net 高 6.14 个百分点,比 UCTransNet 高 0.24 个百分点;MIoU 值比 U-Net 高 7.73 个百分点,比 UCTransNet 高 0.47 个百分点。与 U-Net 相比,UCTransNet 在获取局部特征方面较强,能够很好地识别出细胞组织,本文对 UC-

TransNet 进行了改进,能更加精确地处理边缘信息。图 8 展示了本文方法与其他方法在 GlaS 和 MoNuSeg 两个数据集上的可视化结果。本文方法的可视化结果基本趋于真值标签。这些可视化结果更加直观地展示了分割效果,验证了本文提出改进方法具有较强的泛化能力。例如,对于细胞组织,U-Net 和 UCTransNet 无法妥当地处理细胞壁边缘,细

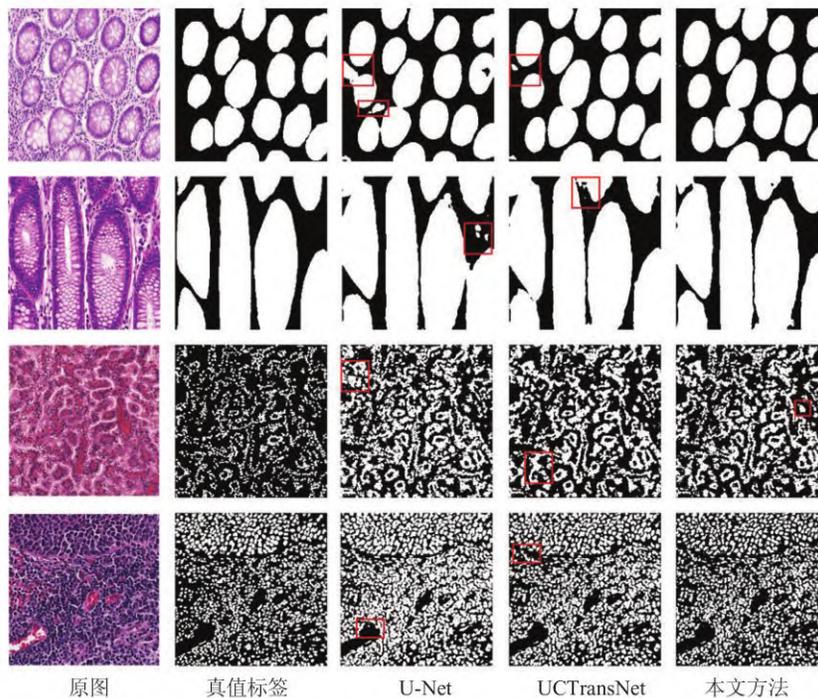


图 8 本文方法与其他方法在 GlaS(第一、二行)和 MoNuSeg(第三、四行)数据集上的比较结果

Fig. 8 Comparative results of our proposed method and other methods on the GlaS(Line 1, 2) and MoNuSeg(Line 3, 4) datasets

胞组织之间仍然存在重合部分,但本文方法能够完整清晰地识别出细胞的边缘。

#### 4 结 语

本文在 U-Net 框架中添加多级融合模块、倒置金字塔特征模块,并将 U-Net 中的跳跃连接替换成基于 Transformer 的通道交叉融合模块(CCT)和通道交叉注意力模块(CCA),实现精确且可靠的医学图像分割。本文的这种方法通过端到端的方式利用了通道式交叉融合模块和通道交叉注意力模块的优势(这种注意力机制可以自适应地调整不同通道的重要性),使得网络能够更加聚焦于关键信息,进而提高分割的准确性,显著地改善了当前舌像分割效果。随着数据集的不断积累和算法的不断优化,本文方法有望应用于更多的医学领域,协助医生更准确地诊断和治疗疾病。

#### 参考文献:

- [1] 孟宪友, 黄水清. 中医舌诊源流探析[J]. 辽宁中医杂志, 2016, **43**(5): 946-948. DOI: 10.13192/j.issn.1000-1719.2016.05.018.  
MENG X Y, HUANG S Q. Origin and development of tongue diagnosis of TCM[J]. *Liaoning Journal of Traditional Chinese Medicine*, 2016, **43**(5): 946-948. DOI: 10.13192/j.issn.1000-1719.2016.05.018 (Ch).
- [2] 王静, 常佩芬, 关静, 等. 中医舌诊客观化研究[J]. 中国中医基础医学杂志, 2022, **28**(9): 1531-1534. DOI: 10.19945/j.cnki.issn.1006-3250.2022.09.029.  
WANG J, CHANG P F, GUAN J, *et al.* Study on objectification research of tongue diagnosis in traditional Chinese medicine[J]. *Journal of Basic Chinese Medicine*, 2022, **28**(9): 1531-1534. DOI: 10.19945/j.cnki.issn.1006-3250.2022.09.029 (Ch).
- [3] 张冬, 庞稳泰, 王可仪, 等. 基于微观角度的中医舌诊客观化研究的现在与未来[J]. 世界科学技术—中医药现代化, 2022, **24**(11): 4574-4579. DOI: 10.11842/wst.20211103006.  
ZHANG D, PANG W T, WANG K Y, *et al.* The present and future of the research on the objectification of TCM tongue diagnosis based on the microscopic angle[J]. *World Science and Technology—Modernization of Traditional Chinese Medicine*, 2022, **24**(11): 4574-4579. DOI: 10.11842/wst.20211103006 (Ch).
- [4] 李丹溪, 关静, 李峰. 舌诊仪的发展及其在舌诊客观化研究中的应用现状[J]. 世界中医药, 2017, **12**(2): 456-460. DOI: 10.3969/j.issn.1673-7202.2017.02.054.  
LI D X, GUAN J, LI F. Development of tongue instrument and progress of its application in traditional Chinese medicine tongue characterization research[J]. *World Chinese Medicine*, 2017, **12**(2): 456-460. DOI: 10.3969/j.issn.1673-7202.2017.02.054 (Ch).
- [5] 肖群霞, 郑晓婷, 吴春城, 等. 舌诊在肝硬化中的应用研究[J]. 医学信息, 2020, **33**(19): 21-23. DOI: 10.3969/j.issn.1006-1959.2020.19.007.  
XIAO Q X, ZHENG X T, WU C C, *et al.* Research progress of tongue diagnosis in cirrhosis[J]. *Medical Information*, 2020, **33**(19): 21-23. DOI: 10.3969/j.issn.1006-1959.2020.19.007 (Ch).
- [6] 苏欢, 张志明, 雍文星, 等. 基于舌苔微生态中医舌诊客观化研究概况[J]. 河南中医, 2021, **41**(8): 1267-1270. DOI: 10.16367/j.issn.1003-5028.2021.08.0287.  
SU H, ZHANG Z M, YONG W X, *et al.* General situation of objectification of TCM tongue diagnosis based on microecology of tongue coating [J]. *Henan Traditional Chinese Medicine*, 2021, **41**(8): 1267-1270. DOI: 10.16367/j.issn.1003-5028.2021.08.0287 (Ch).
- [7] 孙璇, 王东军, 田春颖, 等. 中医舌诊客观化可视化分析[J]. 天津中医药, 2021, **38**(12): 1537-1543. DOI: 10.11656/j.issn.1672-1519.2021.12.10.  
SUN X, WANG D J, TIAN C Y, *et al.* Objective and visual analysis of tongue diagnosis in traditional Chinese medicine [J]. *Tianjin Journal of Traditional Chinese Medicine*, 2021, **38**(12): 1537-1543. DOI: 10.11656/j.issn.1672-1519.2021.12.10(Ch).
- [8] 齐真, 许家佗, 张志枫, 等. 基于数字图像处理技术的舌诊客观化临床应用进展[J]. 中华中医药杂志, 2015, **30**(8): 2849-2851.  
QI Z, XU J T, ZHANG Z F, *et al.* Progress in clinical application of tongue inspection objectivity based on digital image processing technique [J]. *China Journal of Traditional Chinese Medicine and Pharmacy*, 2015, **30**(8): 2849-2851(Ch).
- [9] 张丽倩, 李孟航, 高珊珊, 等. 面向计算机辅助舌诊关键问题的解决方案综述[J]. 计算机科学, 2021, **48**(7): 256-269. DOI: 10.11896/jsjxx.200800223.  
ZHANG L Q, LI M H, GAO S S, *et al.* Summary of computer-assisted tongue diagnosis solutions for key problems [J]. *Computer Science*, 2021, **48**(7): 256-269. DOI: 10.11896/jsjxx.200800223 (Ch).
- [10] 张林子, 周武, 张洪来. 自动舌诊技术的研究进展[J]. 中国中医基础医学杂志, 2023, **29**(5): 871-876. DOI: 10.19945/j.cnki.issn.1006-3250.2023.05.035.  
ZHANG L Z, ZHOU W, ZHANG H L. Research progress of automatic tongue diagnosis technology[J]. *Journal of Basic Chinese Medicine*, 2023, **29**(5): 871-876. DOI: 10.19945/

- j.cnki.issn.1006-3250.2023.05.035(Ch).
- [11] 姜朔, 胡洁, 夏春明, 等. 基于Otsu阈值法与形态学自适应修正的舌像分割方法[J]. 高技术通讯, 2017, **27**(2): 150-155. DOI: 10.3772/j.issn.1002-0470.2017.02.006.
- JIANG S, HU J, XIA C M, *et al.* A tongue image separation method based on Otsu threshold method and morphological adaptive correction[J]. *Chinese High Technology Letters*, 2017, **27**(2): 150-155. DOI: 10.3772/j.issn.1002-0470.2017.02.006 (Ch).
- [12] 傅之成, 李晓强, 李福凤. 基于径向边缘检测和Snake模型的舌像分割[J]. 中国图象图形学报, 2009, **14**(4): 688-693.
- FU Z C, LI X Q, LI F F. Tongue image segmentation based on snake model and radial edge detection[J]. *Journal of Image and Graphics*, 2009, **14**(4): 688-693 (Ch).
- [13] 黄展鹏, 黄益栓, 易法令, 等. 基于最大类间方差法和区域生长的舌体自动分割[J]. 时珍国医国药, 2017, **28**(12): 3062-3064. DOI: 10.3969/j.issn.1008-0805.2017.12.088.
- HUANG Z P, HUANG Y S, YI F L, *et al.* An automatic tongue segmentation algorithm based on Otsu and region growing[J]. *Lishizhen Medicine and Materia Medica Research*, 2017, **28**(12): 3062-3064. DOI: 10.3969/j.issn.1008-0805.2017.12.088 (Ch).
- [14] 杨燕平, 瞿岳云. 裂纹舌的研究概况[J]. 湖南中医杂志, 2008, **24**(6): 96-98. DOI: 10.16808/j.cnki.issn1003-7705.2008.06.064.
- YANG Y P, QU Y Y. General situation of research on cracked tongue[J]. *Hunan Journal of Traditional Chinese Medicine*, 2008, **24**(6): 96-98. DOI: 10.16808/j.cnki.issn1003-7705.2008.06.064 (Ch).
- [15] 杨朝辉. 计算机舌诊中裂纹舌图像的診断分类研究[D]. 哈尔滨: 哈尔滨工业大学, 2010.
- YANG Z H. Research on diagnostic classification for fissured tongue images in computerized tongue diagnosis[D]. Harbin: Harbin Institute of Technology, 2010 (Ch).
- [16] 李敏. 病理性齿痕舌与临床常见证候的相关性研究[J]. 新中医, 2011, **43**(11): 48-50. DOI: 10.13457/j.cnki.jncm.2011.11.024.
- LI M. Correlation study on the relationship between pathological tooth marked tongue and clinically common symptoms[J]. *Journal of New Chinese Medicine*, 2011, **43**(11): 48-50. DOI: 10.13457/j.cnki.jncm.2011.11.024 (Ch).
- [17] 潘赐明, 朱培超, SISHIR Sharma, 等. 齿痕舌源流考及临床意义探讨[J]. 陕西中医, 2021, **42**(9): 1267-1269. DOI: 10.3969/j.issn.1000-7369.2021.09.031.
- PAN C M, ZHU P C, SISHIR S, *et al.* Study on the origin and clinical significance of toothed tongue[J]. *Shaanxi Journal of Traditional Chinese Medicine*, 2021, **42**(9): 1267-1269. DOI: 10.3969/j.issn.1000-7369.2021.09.031 (Ch).
- [18] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241. DOI:10.1007/978-3-319-24574-4\_28.
- [19] 王丽冉, 汤一平, 陈朋, 等. 面向舌体分割的两阶段卷积神经网络设计[J]. 中国图象图形学报, 2018, **23**(10): 1571-1581.
- WANG L R, TANG Y P, CHEN P, *et al.* Two-phase convolutional neural network design for tongue segmentation[J]. *Journal of Image and Graphics*, 2018, **23**(10): 1571-1581 (Ch).
- [20] 颜建军, 徐姿, 郭睿, 等. 基于Mask R-CNN的舌图像分割研究[J]. 世界科学技术——中医药现代化, 2020, **22**(5):1532-1538.
- YAN J J, XU Z, GUO R, *et al.* Research on tongue image segmentation based on Mask R-CNN[J]. *Modernization of Traditional Chinese Medicine and Materia Medica—World Science and Technology*, 2020, **22**(5):1532-1538 (Ch).
- [21] YAN B, ZHANG S, YANG Z J, *et al.* Tongue segmentation and color classification using deep convolutional neural networks[J]. *Mathematics*, 2022, **10**(22): 4286. DOI: 10.3390/math10224286.
- [22] LIU W, CHEN J M, LIU B, *et al.* Tongue image segmentation and tongue color classification based on deep learning [J]. *Digital Chinese Medicine*, 2022, **5**(3):253-263. DOI:10.1016/j.dcm.2022.10.002.
- [23] MATHEW J K, SATHYALAKSHMI S. ExpACVO-Hybrid deep learning: Exponential anti corona virus optimization enabled hybrid deep learning for tongue image segmentation towards diabetes mellitus detection[J]. *Biomedical Signal Processing and Control*, 2023, **83**: 104635. DOI: 10.1016/j.bspc.2023.104635.
- [24] WANG H N, CAO P, WANG J Q, *et al.* UCTransNet: Rethinking the skip connections in U-Net from a channel-wise perspective with transformer[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, **36**(3): 2441-2449. DOI: 10.1609/aaai.v36i3.20144.
- [25] LIN T Y, DOLLÁR P, GIRSHICK R, *et al.* Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2017: 936-944. DOI: 10.1109/CVPR.2017.106.
- [26] DENG A P, LIU J H, CHEN Q Q, *et al.* Visual tracking with FPN based on transformer and response map enhancement[J]. *Applied Sciences*, 2022, **12**(13): 6551.

- DOI: 10.3390/app12136551.
- [27] 李莉, 乔璐, 张浩洋. 结合 FPN 改进 R-FCN 的肺结节检测算法[J]. 计算机应用与软件, 2022, **39**(4): 179-184. DOI: 10.3969/j.issn.1000-386x.2022.04.028.
- LI L, QIAO L, ZHANG H Y. Improved R-FCN lung nodule detection algorithm based on FPN [J]. *Computer Applications and Software*, 2022, **39**(4): 179-184. DOI: 10.3969/j.issn.1000-386x.2022.04.028 (Ch).
- [28] 刘鹏, 毕誉轩, 张天翼, 等. 注意力机制优化的全尺寸目标检测方法[J]. 电子测量与仪器学报, 2023, **37**(2): 193-203. DOI: 10.13382/j.jemi.B2205954.
- LIU P, BI Y X, ZHANG T Y, *et al.* Full-size object detection method optimized by attention mechanism [J]. *Journal of Electronic Measurement and Instrumentation*, 2023, **37**(2): 193-203. DOI: 10.13382/j.jemi.B2205954 (Ch).
- [29] HE J Z, WANG J, HAN Z Y, *et al.* Cancer detection for small-size and ambiguous tumors based on semantic FPN and transformer[J]. *PLoS One*, 2023, **18**(2): e0275194. DOI: 10.1371/journal.pone.0275194.
- [30] 上官天钧, 丁学明, 王霞红, 等. 基于注意力 Unet 的多尺度胎儿图像分割方法[J]. 控制工程, 2023, **30**(4): 722-729. DOI: 10.14107/j.cnki.kzgc.20220041.
- SHANGGUAN T J, DING X M, WANG X H, *et al.* Multi-scale fetal image segmentation method based on attention Unet[J]. *Control Engineering of China*, 2023, **30**(4): 722-729. DOI: 10.14107/j.cnki.kzgc.20220041 (Ch).
- [31] XU Y, HOU S K, WANG X Y, *et al.* A medical image segmentation method based on improved UNet 3+ network[J]. *Diagnostics*, 2023, **13**(3): 576. DOI: 10.3390/diagnostics13030576.
- [32] HUANG X P, CHEN J X, CHEN M Z, *et al.* TDD-UNet: Transformer with double decoder UNet for COVID-19 lesions segmentation[J]. *Computers in Biology and Medicine*, 2022, **151**(Pt A): 106306. DOI: 10.1016/j.compbiomed.2022.106306.
- [33] 程俊, 李红岩, 郎许峰, 等. 融合 Seg-UNet 的中医面像分割网络模型研究[J]. 世界科学技术——中医药现代化, 2022, **24**(10): 4073-4081. DOI: 10.11842/wst.20220515001.
- CHEN J, LI H Y, LANG X F, *et al.* Research on TCM image segmentation network model integrating Seg-UNet [J]. *Modernization of Traditional Chinese Medicine and Materia Medica—World Science and Technology*, 2022, **24**(10): 4073-4081. DOI: 10.11842/wst.20220515001 (Ch).
- [34] SIRINUKUNWATTANA K, PLUIM J P W, CHEN H, *et al.* Gland segmentation in colon histology images: The glas challenge contest [J]. *Medical Image Analysis*, 2017, **35**: 489-502. DOI: 10.1016/j.media.2016.08.008.
- [35] KUMAR N, VERMA R, SHARMA S, *et al.* A dataset and a technique for generalized nuclear segmentation for computational pathology[J]. *IEEE Transactions on Medical Imaging*, 2017, **36**(7): 1550-1560. DOI: 10.1109/TMI.2017.2677499.
- [36] VALANARASU J M J, OZA P, HACIHALILOGLU I, *et al.* Medical transformer: Gated axial-attention for medical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2021: 36-46. DOI: 10.1007/978-3-030-87193-2\_4.

□