

基于改进 UCTransNet 的海洋微藻图像分割模型

窦宇¹, 陈宏远¹, 谭华超¹, 袁贵鸿¹, 江彦博¹, 刘丹^{1,2}

(1.大连海洋大学信息工程学院, 辽宁 大连 116023;

2.大连海洋大学设施渔业教育部重点实验室, 辽宁 大连 116023)

✉ 601096472@qq.com; 18473001460@163.com; tanhuachao_work@163.com; 858774793@qq.com;
1426101935@qq.com; liudan@dlou.edu.cn



摘要:海洋微藻是海洋生态系统的基石,对其进行分割识别可以监测海洋水质并防治藻华。UCTransNet 是使用 Transformer 模块替代 UNet 中跳跃连接模块的分割模型,但 UCTransNet 过于重视图像的通道信息而忽略了图像的空间信息。针对此情况,提出一种将空间与通道融合的注意力机制,并将其加入 UCTransNet 中,得到 CSAM-UCTransNet。该模型加强了编码器与译码器之间的联系。实验表明,CSAM-UCTransNet 对海洋微藻样本的分割精度相较于 UCTransNet 提升了 4.88%。与 UNet、Attention-UNet、UNet++ 等分割算法相比,该模型分割精度更高,对细节的处理效果更好。

关键词:海洋微藻;图像分割;UNet 网络;UCTransNet 网络;注意力机制

中图分类号:TP391 **文献标志码:**A

A Segmentation Model for Marine Microalgae Images Based on Improved UCTransNet

DOU Yu¹, CHEN Hongyuan¹, TAN Huanchao¹, YUAN Guihong¹, JIANG Yanbo¹, LIU Dan^{1,2}

(1.School of Information Engineering, Dalian Ocean University, Dalian 116023, China;

2. Facilities Key Laboratory of Ministry of Fisheries and Education, Dalian Ocean University, Dalian 116023, China)

✉ 601096472@qq.com; 18473001460@163.com; tanhuachao_work@163.com; 858774793@qq.com;
1426101935@qq.com; liudan@dlou.edu.cn

Abstract: Marine microalgae are the cornerstone of marine ecosystems, and their segmentation and identification can monitor marine water quality and prevent algal blooms. UCTransNet is a segmentation model that replaces the skip connection module in UNet with the Transformer module. However, UCTransNet places so much emphasis on the channel information of images that it ignores the spatial information of images. Aiming at this problem, this paper proposes a Channel-Space Attention Module (CSAM), and it is added to UCTransNet to obtain CSAM-UCTransNet. This model strengthens the connection between the encoder and decoder. The experiment shows that CSAM-UCTransNet improves the segmentation accuracy of marine microalgae samples by 4.88% compared to UCTransNet. Compared with other segmentation algorithms, such as UNet, Attention-UNet and UNet++, the proposed model has higher segmentation accuracy and better handling of details.

Key words: marine microalgae; image segmentation; UNet network; UCTransNet network; attention mechanism

0 引言 (Introduction)

在赤潮发生的早期,快速鉴定赤潮藻类对防治藻华有着十分重要的意义。图像分割算法是将图像在像素级别进行分类

的算法。早期的数字图像处理分割主要采用阈值分割、区域分割、聚类等方法^[1]。近年来,深度学习技术开始应用到图像分割领域。SHELHAMER 等^[2]提出了 FCN(Fully Convolutional Networks)网

络模型,实现了图片端到端的分割。RONNEBERGER 等^[3]提出了 UNet 网络模型,U 形分割模型开始兴起。ZHOU 等^[4]提出了 UNet++,使用密集型跳跃连接加强编码译码器之间的联系。OKTAY 等^[5]提出了 Attention-UNet,将注意力机制引入 UNet 模型之中。

近年来,多头注意力机制兴起,CHEN 等^[6]提出了 TransUNet 网络模型,使用 Transformer 作为译码器。VALANARASU 等^[7]提出了 Medical Transformer,在 Transformer 的基础上加入了可学习的门控机制。JI 等^[8]将多尺度的语义特征整合到一个框架中,提出了 MCTrans 模型。GAO 等^[9]把 Transformer 模块同时加入 UNet 的编码与译码模块,提出了 UTNet 网络。JI 等^[10]使用渐进归一化自注意力网络 PNS-Net 来获取更大的感受野。ZHANG 等^[11]将多个网络融合,提出了 MBT-Net 网络模型。WANG 等^[12]提出了使用 Transformer 代替跳跃连接的模型——UCTransNet,同时在译码阶段加入了通道注意力模块^[13]。但是,UCTransNet 过于重视特征图通道之间的联系而忽略了其空间的特征。本文在 UCTransNet 的基础上引入空间注意力机制^[14],得到 Channel-Space Attention Module (CSAM)模块,进一步提高了分割效果。

1 算法改进 (Algorithm improvement)

1.1 UCTransNet 网络模型

UCTransNet 网络模型整体的编解码器依然采用 UNet 模型架构。取消了原始的跳跃连接,将各级编码完成的特征图 E_i 输入带有 Transformer 的通道交通融合模块(CCT)中,然后输出到译码器中。在译码阶段采用通道交叉注意力模块 Channel wise Cross-Attention(CCA)将 CCT 的输出和上采样的输出进行融合。

原始 UCTransNet 模型架构如图 1 所示,在 CCT 模块中各级特征图首先经过层归一化(LN),使输入数据的分布更规范、差异性更小。在归一化结束后,将各级数据 Concat 作为多头交叉注意力机制(MCA)的 V 和 K 输入 MCA 中,同时将层归一化之后的各级 E_i 作为 MCA 的 Q 输入。

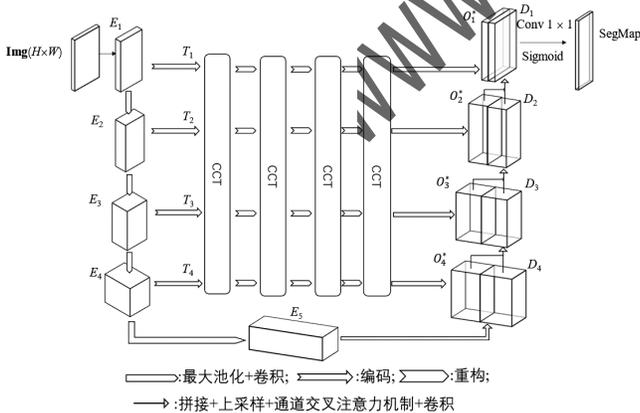


图 1 原始 UCTransNet 模型架构

Fig. 1 UCTransNet model architecture

CCT 模块架构如图 2 所示,与原始多头注意力机制不同的是,MCA 是沿着通道轴进行注意力操作,可以加强通道联系,节省计算量,MCA 架构如图 3 所示。经过 MCA 后,各级特征图进行归一化和全连接层 MLP(Multi-Layer Perceptron),然后输出到译码器上。通过 CCT 模块的各级数据加强了通道

上的语义联系,为后续译码奠定了基础。在译码阶段,CCT 输出的 O_i 与译码器上采样的特征图 D_i 一起输入通道交叉注意力模块(CCA),经过一层全局池化和全连接层与 O_i 相乘,形成融合编码和译码数据的特征图 O_i^* ,然后经上采样输出。

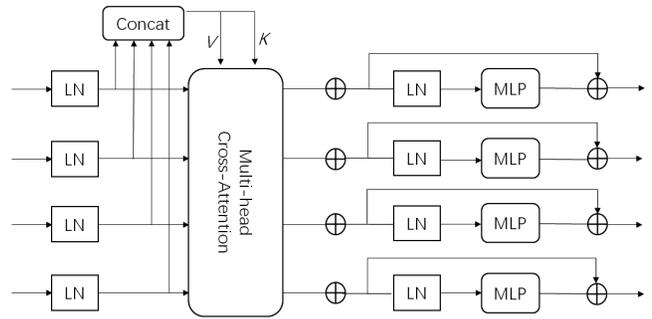


图 2 CCT 模块架构

Fig. 2 CCT module architecture

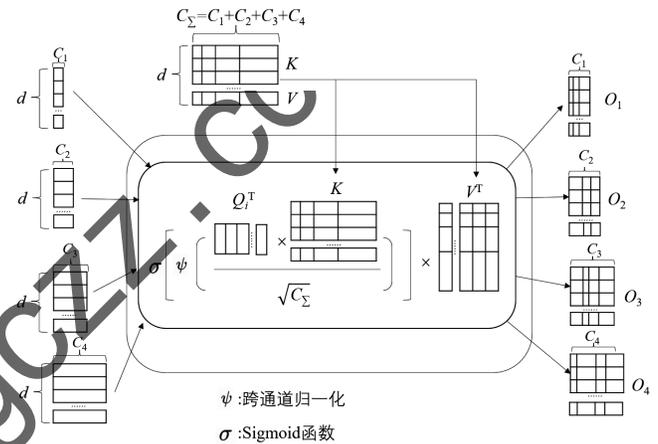


图 3 MCA 架构

Fig. 3 MCA architecture

CCA 模型架构如图 4 所示,CCA 加强了 D_i 与 O_i 在通道上的语义联系,并且提高了较为重要的通道的权重。译码器将经过 4 次 CCA 模块与上采样的特征图,输入 1×1 的卷积层与 S 型函数(Sigmoid)中,最终输出与原始图片尺寸一致的分割结果。

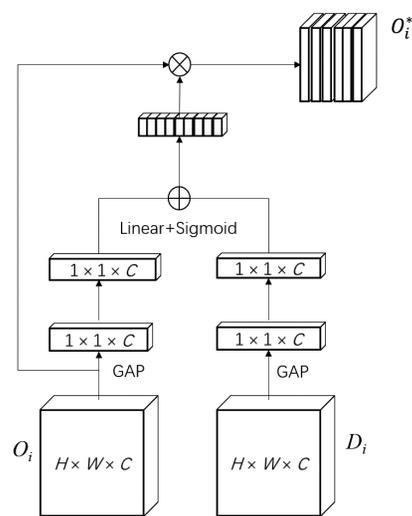


图 4 CCA 模型架构

Fig. 4 CCA model architecture

1.2 CSAM 网络架构

由于显微照片下的藻类细胞过小,为了进一步突出特征图的空间与通道特征,同时消除编码与译码器之间的语义鸿沟。本研究在 UCTransNet 的基础上进一步改进了 CCA 模块,提出了通道空间融合注意力机制(CSAM),CSAM 模型架构如图 5 所示。

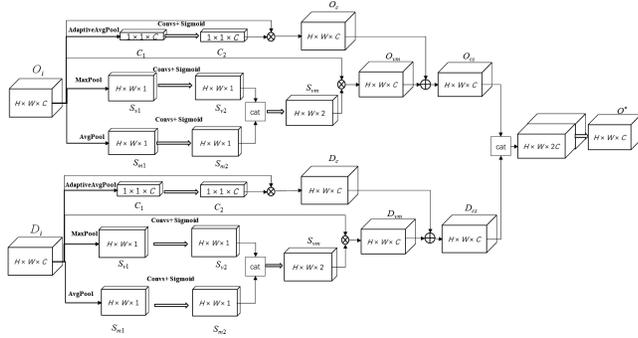


图 5 CSAM 模型架构

Fig. 5 CSAM model architecture

不同于 CCA 模块比较重视通道之间的特征融合,CSAM 在保留 CCA 通道能力的同时,加强了特征图空间上的权重,同时在空间和通道多维度上融合编码与译码的数据,使模型的分割结果更精细、准确。CSAM-UCTransNet 网络结构如图 6 所示。

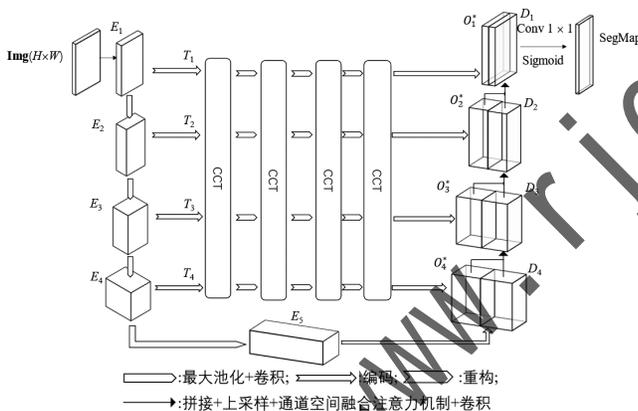


图 6 CSAM-UCTransNet 网络结构

Fig. 6 CSAM-UCTransNet network structure

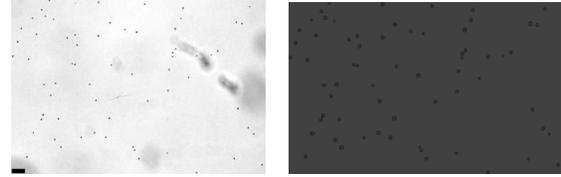
CSAM 的输入为 CCT 输出的特征图 O_i 以及译码器上采样得到的特征图 D_i 。CSAM 网络的主干是由三个池化层交叉连接形成的。以 D_i 为例,设其输入大小为 H, W, C ,分别代表高度、宽度、通道数。首先 D_i 经过通道层方向上的自适应平均池化压缩为 $1 \times 1 \times C$ 的特征图 C_1 ,将 C_1 经过 1×1 的卷积与 Sigmoid 函数得到 C_2 。将输入 D_i 与 C_2 相乘得到通道加权后的特征图 D_c 。同时,将输入 D_i 在空间上先进行全局平均池化得到 $H \times W \times 1$ 的平均空间特征图 S_{v1} , S_{v1} 经过 1×1 的卷积与 Sigmoid 函数得到 S_{v2} 。同时,为了更加全面地突出数据的空间特性,再将 D_i 进行最大池化得到 $H \times W \times 1$ 的最大空间特征图 S_{m1} 。 S_{m1} 经过 1×1 卷积与 Sigmoid 函数得到 S_{m2} 。将 S_{v2} 与 S_{m2} 经 cat 操作,融合为一个 $H \times W \times 2$ 的特征图 S_{sm} 。将输入 D_i 与 S_{sm} 相乘得到空间加权后的特征图 D_{sm} 。将 D_{sm}

与 D_c 相加得到融合了空间与通道多重特征的 D_{cs} 。将 D_{cs} 与 CCT 输出的经过相同操作的 O_{cs} 特征图结合,最后经过 1×1 卷积与 Sigmoid 函数得到最终输出的结果 O^* 。

2 实验(Experiment)

2.1 数据集与预处理

本文选用 LeicaDM4B 低倍显微镜,收集小球藻培养液作为实验样本。共计拍摄全尺寸小球藻显微图片 48 张,同时使用 labeme 软件对图片进行标记,藻类全尺寸图片及标签如图 7 所示,其中图 7(a)是采集的小球藻原图,图 7(b)是小球藻标签。



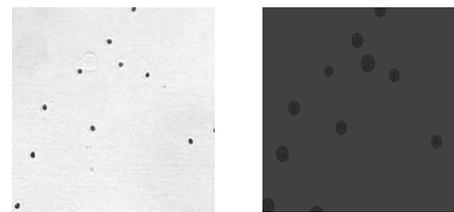
(a)小球藻原图

(b)小球藻标签

图 7 藻类全尺寸图片及标签

Fig. 7 Full size images and labels of algae

图 7 中有 21 张图片的尺寸为 1920×1200 ,27 张图片的尺寸为 2448×1920 。图像尺寸过大,直接进行训练会导致设备内存溢出。但是,若将图片的尺寸直接缩放为 256×256 ,则图片中的大量信息会损失。所以,将图像缩放并切割为 792 张图片尺寸为 480×480 的图片,切割完毕之后将 label 中像素值小于 800 的图片删除,减少图片中藻类过少带来的误差,剩余 401 张图片。每张 480×480 的图片含有 10~100 个小球藻,切分之后的图片及标签如图 8 所示。



(a)原图

(b)标签

图 8 切分之后的图片及标签

Fig. 8 Images and labels after segmentation

本文对数据集做了以下预处理。

- (1)将数据集以 3 : 1 : 1 的比例随机划分为训练集、验证集和测试集。
- (2)在训练之前,对图像进行数据增强,即对图片进行旋转、翻转、裁剪等操作。

2.2 评价指标

本文使用图像分割领域常用的评价指标 Dice 相似系数 (Dice Similarity Coefficient, DSC) 和交并比 (Intersection over Union, IoU)。

Dice 相似系数是用于度量两个集合相似度的函数,可以用于计算预测图像和标签图像之间的相似度,其取值范围为 $[0, 1]$,DSC 的值越接近于 1,表示分割效果越好,计算公式如下:

$$DSC(T, P) = \frac{2|T \cap P|}{|T| + |P|} \quad (1)$$

其中: $|T|$ 表示真实的分割区域, $|P|$ 表示模型预测的分割区域。

IoU 表示预测的分割区域和真实分割区域的交集比上两者的并集, 同样 IoU 的值越接近于 1, 表示分割效果越好, 计算公式如下:

$$IoU = \frac{|T \cap P|}{|T \cup P|} \quad (2)$$

2.3 实验设置

使用编程语言 Python 和 Pytorch 框架进行基本的程序运行和模型搭建。使用 Todesk 远程连接 NVIDIA GeForce RTX 3060, 其内存为 12 GB。使用 Adam 作为优化器进行迭代。由于显卡内存有限, 因此训练时设置 *batch-size* 为 2, 并且每次训练结束后将模型在验证集上进行验证, 记录下验证后的 Dice 相似系数。若这一次的 Dice 相似系数高于以前历次验证结果, 则保存当前模型。迭代 200 次, 保存 200 次中的最佳模型。

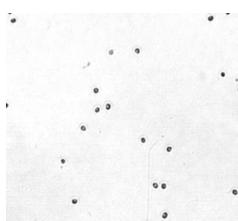
2.4 实验结果分析

本文使用 UNet、UNet++、Attention-UNet、UCTransNet、CSAM-UCTransNet 模型在同一服务器上设置相同的学习率和 *batch-size* 进行分割训练, UNet 是医学图像分割中最常用的网络模型之一, UNet++ 在 UNet 的基础上使用了密集跳跃连接; Attention-UNet 在 UNet 的译码模块中加入了注意力门 (Attention Gate) 机制。UCTransNet 使用多头注意力模块代替 UNet 跳跃连接。各个模型的分割评价指标结果如表 1 所示; 各个模型的分割效果对比图如图 9 所示。

表 1 各个模型的分割评价指标结果

Tab.1 Results of segmentation evaluation indicators for each model

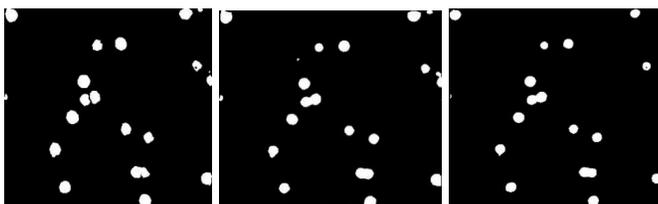
模型	DSC/%	IoU/%
UNet	56.96	40.94
UNet++	61.28	45.15
Attention-UNet	61.50	44.79
UCTransNet	68.81	53.13
CSAM-UCTransNet	73.69	59.16



(a) 全尺寸小球藻原图



(b) 全尺寸小球藻标签



(c) UCTransNet 分割效果 (d) UNet++ 分割效果 (e) UNet 分割效果



(f) Attention-UNet 分割效果 (g) CSAM-UCTransNet 分割效果

图 9 各个模型的分割效果对比图

Fig. 9 Comparison of segmentation results of various models

目前, CSAM-UCTransNet 网络的 Dice 相似系数最高, 超过了 70%。这是因为 CSAM-UCTransNet 在保留了 UCTransNet 完整的 CCT 模块的基础上, 在译码阶段增加了 CSAM 机制对数据的空间和通道进行多尺度融合, 弥补了编码译码阶段的语义鸿沟, 同时提高了模型对图片细节的感应能力。在 5 组对比模型中, UNet 的效果最差, Dice 相似系数仅为 56.96%, 这是因为 UNet 仅用最基础的跳跃连接实现编码译码之间的联系, 这使得模型对于数据细节把控较差。UNet++ 是在 UNet 的基础上使用密集的跳跃连接, 具有更强的特征提取能力, 能对不同尺度的特征进行融合提取。虽然 UNet++ 融合了各级感受野的特征, 但是由于各级感受野只跟相邻层级进行融合, 导致在模型译码阶段容易忽略信息的细节及边缘。Attention-UNet 在 UNet 的译码模块加入注意力门机制, 其计算效率比 UNet++ 更高, 并且能更好地把握全局信息。但是, Attention-UNet 是直接对图片中的细节信息进行加强, 导致它极易受噪声干扰, 将杂质一起分割出来。UCTransNet 使用当前较为流行的 Transformer 模块代替 UNet 的跳跃连接, 并且在译码阶段加入通道注意力机制, 但是 UCTransNet 过于重视图像的通道信息而忽视了图像的空间信息, 相较于 CSAM-UCTransNet 网络, UCTransNet 对于小目标样本分割效果相对差一点。

图 9 为各个模型对验证集上图片的分割效果。对比图为全尺寸显微照片切割后尺寸为 480×480 的图片。图 9(c) 呈现了 UCTransNet 的分割效果, UCTransNet 对小目标样本敏感度不足, 对小球藻的边缘分割效果相对不足。图 9(d) 呈现了 UNet++ 的分割效果, 其对小球藻的边缘分割效果较好, 但是相邻细胞易出现粘连情况, 影响整体准确性。图 9(e) 呈现了 UNet 的分割效果, 与 UNet++ 的分割效果一样, UNet 模型对单个细胞的分割区域过大, 导致相邻细胞互相粘连。图 9(f) 呈现了 Attention-UNet 的分割效果, 虽然 Attention-UNet 对细胞边缘分割更为细致, 但是出现较多的杂质。图 9(g) 呈现了 CSAM-UCTransNet 的分割效果, 其对小球藻实现了较为准确的分割, Dice 相似系数达到了 73.69%, 同时 IoU 达到了 59.16%, 结果最好。

3 结论 (Conclusion)

本文在 UCTransNet 的基础上, 改进译码器的注意力机制, 提出了一种结合了通道空间融合注意力机制的网络模型 CSAM-UCTransNet。该模型在译码阶段不仅考虑了数据的通

道信息,还将 CCT 模块输出的数据与译码器的空间信息和通道信息进行特征融合,进一步加强了模型在编码和译码之间的联系,并且强化了模型对样本在空间上的特征提取能力。相较于其他 U 形图像分割模型,CSAM-UTransNet 的分割精度更高,得到的效果图更好。后续将进一步对模型进行改进,同时引入 LoRA^[15]等参数高效性微调方法,进一步提高网络模型的精度和收敛速度。

参考文献 (References)

- [1] GUO Y M, LIU Y, GEORGIU T, et al. A review of semantic segmentation using deep neural networks[J]. International journal of multimedia information retrieval, 2018, 7: 87-93.
- [2] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(4): 640-651.
- [3] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]// NAVAB N, HORNEGGER J, WELLS W M, et al. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Cham: Springer, 2015: 234-241.
- [4] ZHOU Z W, SIDDIQUEE M M R, TAJBAKSHI N, et al. UNet++: a nested U-net architecture for medical image segmentation[C]// STOYANOV D, TAYLOR Z, CARNEIRO G, et al. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018 and 8th International Workshop, ML-CDS 2018. Cham: Springer, 2018, 11045: 3-11.
- [5] OKTAY O, SCHLEMPER J, LEFLOUCC L, et al. Attention U-net: learning where to look for the pancreas [DB/OL]. (2018-05-20) [2023-05-08]. <https://doi.org/10.48550/arXiv.1804.03999>.
- [6] CHEN J N, LU Y Y, YU Q H, et al. TransUNet: transformers make strong encoders for medical image segmentation[DB/OL]. (2021-02-08) [2023-05-08]. <https://doi.org/10.48550/arXiv.2102.04306>.
- [7] VALANARASU J M J, OZA P, HACIHALILOGLU I, et al. Medical transformer: gated axial-attention for medical image segmentation[DB/OL]. (2021-07-06) [2023-05-08]. <https://doi.org/10.48550/arXiv.2102.10662>.
- [8] JI Y F, ZHANG R M, WANG H J, et al. Multi-compound transformer for accurate biomedical image segmentation [C]// BRUIJNE M D, CATTIN P C, COTIN S, et al. Medical Image Computing and Computer Assisted Intervention—MICCAI 2021. Cham: Springer, 2021: 326-336.
- [9] GAO Y H, ZHOU M, METAXAS D N. UTransNet: a hybrid transformer architecture for medical image segmentation [C]// BRUIJNE M D, CATTIN P C, COTIN S, et al. Medical Image Computing and Computer Assisted Intervention—MICCAI 2021. Cham: Springer, 2021: 61-71.
- [10] JI G P, CHOU Y C, FAN D P, et al. Progressively normalized self-attention network for video polyp segmentation [C]// BRUIJNE M D, CATTIN P C, COTIN S, et al. Medical Image Computing and Computer Assisted Intervention—MICCAI 2021. Cham: Springer, 2021: 142-152.
- [11] ZHANG Y L, HIGASHITA R, FU H Z, et al. A multi-branch hybrid transformer network for corneal endothelial cell segmentation [C]// BRUIJNE M D, CATTIN P C, COTIN S, et al. Medical Image Computing and Computer Assisted Intervention—MICCAI 2021. Cham: Springer, 2021: 99-108.
- [12] WANG H N, CAO P, WANG J Q, et al. UTransNet: rethinking the skip connections in U-net from a channel-wise perspective with transformer [DB/OL]. (2022-01-25) [2023-05-08]. <https://doi.org/10.48550/arXiv.2109.04335>.
- [13] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [DB/OL]. (2019-05-16) [2023-05-08]. <https://doi.org/10.48550/arXiv.1709.01507>.
- [14] MNIH V, HEES N, GRAVES A, et al. Recurrent models of visual attention [J]. Machine learning, 2014 (24): 1-12.
- [15] HU E J, SHEN Y L, WALLIS P, et al. LoRA: low-rank adaptation of large language models [J]. Computation and language, 2021(16): 15-26.

作者简介:

窦宇 (1995-), 男, 硕士生。研究领域: 海洋微藻图像分割算法。

陈宏远 (1998-), 男, 硕士生。研究领域: 深度学习, 计算机视觉。

谭华超 (1997-), 男, 硕士生。研究领域: 深度学习, 计算机视觉。

袁贵鸿 (2001-), 男, 硕士生。研究领域: 深度学习, 计算机视觉。

江彦博 (2000-), 男, 硕士生。研究领域: 深度学习。

刘丹 (1978-), 女, 博士, 副教授。研究领域: 涉海智能图像处理理论与应用研究。本文通信作者。