中图法分类号:TP 391　　**文献标识码**: A　　**文章编号**: 1006-8961(2024)03-0713-12

论文引用格式:Cao W J, Duan X H, Xu Z W and Sheng S. 2024. Application of U-Net channel transformation network in gland image segmentation. Journal of Image and Graphics, 29(03):0713-0724(曹伟杰, 段先华, 许振伟, 盛帅. 2024. U-Net通道变换网络在腺体图像分割中的应用. 中国图象图形学报, 29(03):0713-0724)[DOI:10.11834/jig.230233]

U-Net通道变换网络在腺体图像分割中的应用

曹伟杰,段先华*,许振伟,盛帅

江苏科技大学计算机学院,镇江　212100

摘　要:**目的**　腺体医学图像分割是将医学图像中的腺体区域与周围组织分离出来的过程,对分割精度有极高要求。传统模型在对腺体医学图像分割时,因腺体形态多样性和小目标众多的特点,容易出现分割不精细或误分割等问题,对此根据腺体医学图像的特点对U-Net型通道变换网络分割模型进行改进,实现对腺体图像更高精度分割。**方法**　首先在U-Net型通道变换网络的编码器前端加入ASPP_SE(spatial pyramid pooling_squeeze-and-excitation networks)模块与ConvBatchNorm模块的组合,在增强编码器提取小目标特征信息能力的同时,防止模型训练出现过拟合现象。其次在编码器与跳跃连接中嵌入简化后的密集连接,增强编码器相邻模块特征信息融合。最后在通道融合变换器(channel cross fusion with Transformer,CCT)中加入细化器,将自注意力图投射到更高维度,提高自注意机制能力,增强编码器全局模块特征信息融合。简化后的密集连接与CCT结合使用,模型可以达到更好效果。**结果**　改进算法在公开腺体数据集MoNuSeg(multi-organ nuclei segmentation challenge)和Glas(gland segmentation)上进行实验。以Dice系数和IoU(intersection over union)系数为主要指标,在MoNuSeg的结果为80.55%和67.32%,在Glas数据集的结果为92.23%和86.39%,比原U-Net型通道变换网络分别提升了0.88%、1.06%和1.53%、2.43%。**结论**　本文提出的改进算法在腺体医学分割上优于其他现有分割算法,能满足临床医学腺体图像分割要求。

关键词:医学图像分割;U-Net型通道变换网络(UCTransNet);密集连接;注意力机制;细化器

Application of U-Net channel transformation network in gland image segmentation

Cao Weijie, Duan Xianhua*, Xu Zhenwei, Sheng Shuai

School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang　212100, China

Abstract: **Objective** Adenocarcinoma is a malignant tumor originating from the glandular epithelium and poses immense harm to human health. With the rapid development of computer vision technology, medical imaging has become an important means for expert preoperative diagnosis. In the diagnosis of adenocarcinoma, doctors judge the severity of the cancer and grade it by analyzing the size, shape, and other external features of the glandular structure. Accordingly, achieving high-precision segmentation of glandular images has become an urgent requirement in clinical medicine. Glandular medical image segmentation refers to the process of separating the glandular region from the surrounding tissue in medical images, requiring high segmentation accuracy. Traditional models for segmenting glandular medical images can suffer from

收稿日期:2023-04-20;**修回日期**:2023-10-31;**预印本日期**:2023-11-07

*通信作者:段先华　dxh@just.edu.cn

基金项目:国家自然科学基金项目(62276118)

Supported by:National Natural Science Foundation of China(62276118)

such problems as imprecise segmentation and mis-segmentation owing to the diverse shapes of glands and presence of numerous small targets. To address this issue, this study proposes an improved glandular medical image segmentation algorithm based on UCTransNet. UCTransNet addresses solves the semantic gap between different resolution modules of the encoder and between the encoder and decoder, thereby achieving high precision image segmentation. **Method** First, a combination of the fusion of ASPP_SE and ConvBatchNorm modules is added to the front end of the encoder. The ASPP_SE module combines the ASPP module and channel attention mechanism. The ASPP module consists of three different dilation rates of atrous convolution, a 1 × 1 convolution, and an ASPP pooling. Atrous convolution injects holes into standard convolution to expand the receptive field, obtain dense data features, and maintain the same output feature map size. The ASPP module uses multi-scale atrous convolution to obtain a large receptive field, and fuses the obtained features with the global features obtained from the ASPP pooling to obtain denser semantic information than the original features. The channel attention mechanism enables the model to focus considerably on important channel regions in the image, dynamically select information in the image, and give substantial weight to channels containing important information. In the CCT (channel cross fusion with Transformer), modules with higher weight of important information will achieve better fusion. The ConvBatchNorm module enhances the ability of the encoder to extract the features of small targets, while preventing overfitting during model training. Second, a simplified dense connection is embedded between the encoder and the skip connections, and the CCT in the model performs global feature fusion of the features extracted by the encoder from a channel perspective. Although the global attention ability of the CCT is strong, its problem is a weak local attention ability, and the ambiguity between adjacent encoder modules has not been solved. To solve this problem, a dense connection is added to enhance the local information fusion ability. The dense connection passes the upper encoder module through convolution pooling to obtain the lower encoder module and performs upsampling on the lower encoder to make its resolution consistent with the upper encoder module. The two encoder modules are concatenated on the channel, and the resolution does not change after concatenation. After concatenation, the upper encoder module obtains the feature information supplement of the lower encoder module. Consequently, the semantic fusion between adjacent modules is enhanced, the semantic gap between adjacent encoder modules is reduced, and the feature information fusion between adjacent encoder modules is improved. A refiner is added to the CCT, which projects the self-attention map to a higher dimension, and uses the head convolution to enhance the spatial context and local patterns of the attention map. This method effectively combines the advantages of self-attention and convolution to further improve the self-attention mechanism. Lastly, a linear projection is used to restore the attention map to the initial resolution, thereby enhancing the global feature information fusion of the encoder. A fusion ASPP_SE and ConvBatchNorm modules are added to the front end of the UCTransNet encoder to enhance its ability to extract small target features and prevent overfitting. Second, a simplified dense connection is embedded between the encoder and skip connection to enhance the fusion of adjacent module features. Lastly, a refinement module is added to the CCT to project the self-attention map to a markedly high dimension, thereby enhancing the global feature fusion ability of the encoder. The combination of the simplified dense connection and CCT refinement module improves the performance of the model. **Result** The improved algorithm was tested on the publicly available gland data sets MoNuSeg and Glas. The Dice and intersection over union (IoU) coefficients were the main evaluation metrics used. The Dice coefficient is a similarity measure used to represent the similarity between two samples. By contrast, the IoU coefficient is a standard used to measure the accuracy of the result's positional information. Both metrics are commonly used in medical image segmentation. The test results on the MoNuSeg data set were 80.55% and 67.32%, while those on the Glas data set were 92.23% and 86.39%. These results represent improvements of 0.88% and 1.06%, and 1.53% and 2.43%, respectively, compared those of the original UCTransNet. The improved model was compared to existing popular segmentation networks and was found to generally outperform them. **Conclusion** The proposed improved model is superior to existing segmentation algorithms in medical gland segmentation and can meet the requirements of clinical medical gland image segmentation. The CCT module in the original model was further optimized to fuse global and local feature information, thereby achieving better results.

Key words: medical image segmentation; U-Net from a channel-wise perspective with Transformer (UCTransNet); dense connection; self-attention mechanism; refinement module

0 引 言

腺癌是一种来源于腺上皮的恶性肿瘤，对人体健康有极大危害。近年来人们生活作息不规律，导致腺癌成为严重损害人体健康的恶性肿瘤之一（张家宝和肖志勇，2022）。计算机视觉技术快速发展使医学图像成为专家术前诊断的重要手段（赵樱莉等，2023）。在腺癌诊断中，医生通过分析腺体结构大小、形状和其他一些外在表现，来判断腺癌的严重程度并对其进行分级。如何准确高效分割腺体细胞成为临床医学的迫切要求。传统的人工分割方法（任楚岚 等，2022）需要很多先验知识，耗费大量人力和时间，同时存在分割主观性强、分割缺失等问题，限制了分割准确性和效率，降低了临床诊断的可靠性（蒋希 等，2022）。因此开发一种快速、高精度的图像分割机制，对临床腺体病症的判断具有重要意义。

常用医学图像分割方法分为传统的机器视觉方法和基于深度学习的方法。传统方法包括基于阈值的图像分割、基于区域生长的图像分割和基于聚类的图像分割等。这些方法利用图像自身的灰度、色彩和纹理等特征进行分割。孟月波等人（2019）提出一种具有边缘保持的多尺度马尔可夫随机场模型，将分割问题转化为优化问题，有效保持了图像分割结果的边缘，获得了更好的分割结果。但是传统分割方法在面对医学图像分割的复杂情况下，难以满足其高精度要求。随着人工智能视觉技术的迅速发展，深度学习方法在医学图像分割领域得到广泛应用，通过构建多层深度神经网络模型，训练标注好的数据集，实现高精度分割。目前为止，流行的分割框架有 FCN（fully convolutional network）（Shelhamer 等，2015）、DeepLab（deep convolutional nets）（Chen 等，2014）和 U-Net（Ronneberger 等，2015）等，其中 U-Net 由于其结构比较简单、所需数据集少、不易出现过拟合现象等备受瞩目。很多学者开始对 U-Net 进行相关性研究，提出许多基于 U-Net 的变体结构，例如 U-Net++（Zhou 等，2018）、V-Net（volumetric convcets）（Milletarì 等，2016）、3D U-Net（Çiçek 等，2016）和 Attention U-Net（Oktay 等，2018）等。这些研究通过改进 U-Net 的结构或针对特定场景加入适用的模块，进一步提升模型性能。但由于 U-Net 中的卷积神经网络（convolutional neural network，CNN）存在局部感知性（徐光宪 等，2023），无法准确把握图像全局信息，所构建模型存在一定局限性。

Transformer 的出现给医学图像分割带来新突破，将全局注意力的 Transformer 和关注局部性的 CNN 结合使用，可以有效提高医学图像分割的精度和效率（Dosovitskiy 等，2021），同时兼顾全局特征和局部细节，是一种具有潜力的研究方向（许聪 等，2023）。许多学者根据这个理论提出两者结合使用的模型，如 TransUNet（Transformers make strong encoders）（Chen 等，2021）、Swin-UNet（hybrid vision Transformer with U-Net）（Cao 等，2022）、UCTransNet（U-Net from a channel-wise perspective with Transformer）（Wang 等，2022）、SMESwin UNet（semi-supervised with a Swin Transformer-based U-Net）（Wang 等，2022）。Chen 等人（2021）利用多层次的 Transformer 来提取图像全局的信息，提高了精度，但是训练时间过长。Cao 等人（2022）将 U-Net 中的 CNN 模块替换为 Swin Transformer 模块，增强了 U-Net 特征提取能力。Wang 等人（2022）以 U-Net 作为网络的主要结构，提出一种 Transformer 构成的通道融合变换器（channel cross fusion with Transformer，CCT）来替换 U-Net 中的跳跃连接，从通道的角度将编码器多尺度上下文特征信息进行交叉关注融合，产生了更精确的分割性能。

腺体原始图像及其医学图像的分割存在差异，具体如下：1）腺体图像包含大量小目标，易造成分割不足或误分割；2）不同图像中的细胞形态、大小变化差异大（赵宝奇 等，2021），要求模型具有较强的泛化能力；3）腺体图像中分割信息范围广，要求模型能够有效捕获全局信息（霍璐 等，2021）。

为此，本文采用U-Net与Transformer的最新结合模型，即U-Net型通道变换网络U-Net from a channel-wise perspective with Transformer（UCTransNet）作为参考模型，对腺体医学图像的特点进行针对性改进。本文主要从3个方面进行改进：1）改进编码器结构，在编码器开端加入 ASPP_SE（atrous spatial pyramid pooling_squeeze and excitation）模块和ConvBatchNorm模块，解决模型提取小目标特征信息不足的问题，同时避免过拟合的问题；2）在编码器和跳跃连接中加入简化后的密集连接，以增强编码器相邻模块特征信息融合；3）在通道融合变换器中引入细化器（Zhou 等，2021），将 Transformer 的注意力特征图投射到更高维度空间，增

强自注意机制能力，解决模型融合速度缓慢问题。密集连接与CCT模块相互结合可以使模型达到更好的效果，改进的算法在腺体医学图像分割上优于现有其他分割算法，可以满足临床腺体图像分割的要求。

1 U-Net型通道变换网络

U-Net型通道变换网络的结构如图1所示，模型的基本结构采用U-Net结构，分为编码器、解码器和跳跃连接模块。编码器捕获低、高级语义特征信息，解码器将捕获到的特征信息上采样到像素级别的分割图，跳跃连接补充编码器在卷积、池化操作中丢失的空间信息到解码器中。

U-Net型通道变换网络模型采用通道融合变换器来替换跳跃连接，将不同分辨率的编码器模块提取的特征信息分割为不同批次大小的图像块，将其重塑为二维批次序列$T_i(i = 1, 2, 3, 4)$输入到通道融合变换器，利用Transformer的自相关性融合不同层次的编码器模块特征。从通道的角度将编码器多尺度上下文特性信息与交叉注意力融合在一起，捕捉局部跨通道交互，通过协作学习，有效融合尺度语义间隙的多尺度通道特征。U-Net型通道变换网络有效解决了编码器不同分辨率模块之间以及编码器和解码器之间的语义缺口，实现了更高精度的图像分割。

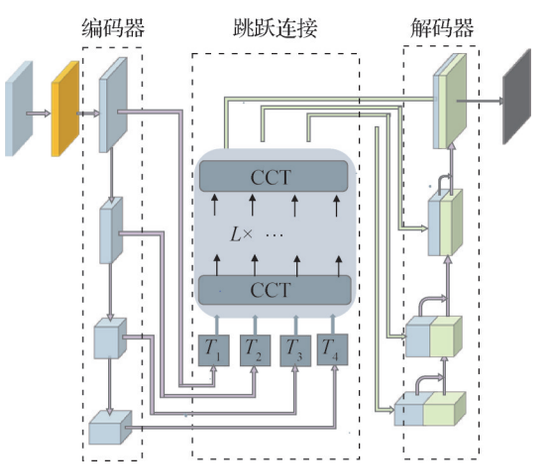


图1 U-Net型通道变换网络

Fig. 1 UCTransNet

2 改进的U-Net型通道变换网络

2.1 改进编码器结构

腺体医学图像中存在许多小目标信息，对腺体医学图像高精度分割具有重要作用。U-Net型通道变换网络的编码器部分仅采用简单的卷积、池化操作，无法充分提取这些小目标特征信息。为了解决这个问题，本文在编码器的开端加入ASPP_SE模块。ASPP_SE模块融合了ASPP模块和通道注意力机制，ASPP模块利用多尺度空洞卷积，通过不同尺度的膨胀率来获得更大的感受野、更密集的数据特征，可促进编码器小目标特征信息的提取。通道注意力可对图像中的通道信息进行动态选择，在之后CCT模块融合起到重要作用。在ASPP_SE模块充分提取图像特征之后加入ConvBatchNorm模块，防止出现过拟合现象。

本文具体的改进方法如图2所示，输入图像会先经过ASPP_SE与ConvBatchNorm模块进行不同膨胀率信息的提取，提取出原编码器模块无法提取的小目标信息，再通过原编码器进一步提取特征。

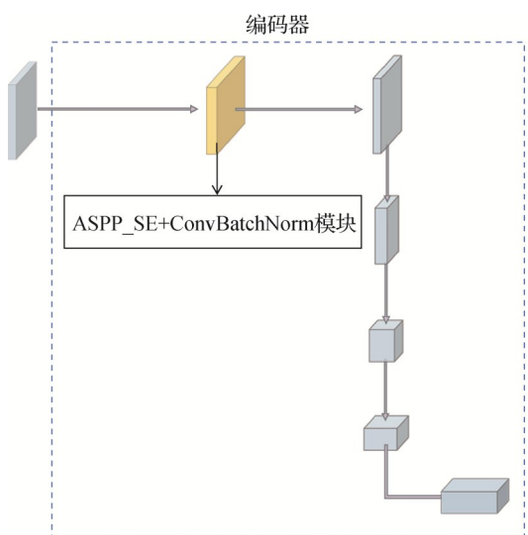


图2 改进编码器结构图

Fig. 2 Improved encoder structure diagram

ASPP模块由3个不同膨胀率的空洞卷积、1个1×1卷积、1个ASPP Pooling组成，模块结构如图3所示。空洞卷积是在标准卷积的基础上注入空洞来扩大感受野，获得更密集的数据特征，同时不改变图像输出特征图的尺寸。ASPP Pooling是一个自适应均值池化，可以提取各个通道的特征，进而提取全局的特征。ASPP利用不同膨胀率的空洞卷积，从不同的尺度提取输入特征，再将所获取的特征与ASPP Pooling得到的全局特征进行特征融合，得到比原有

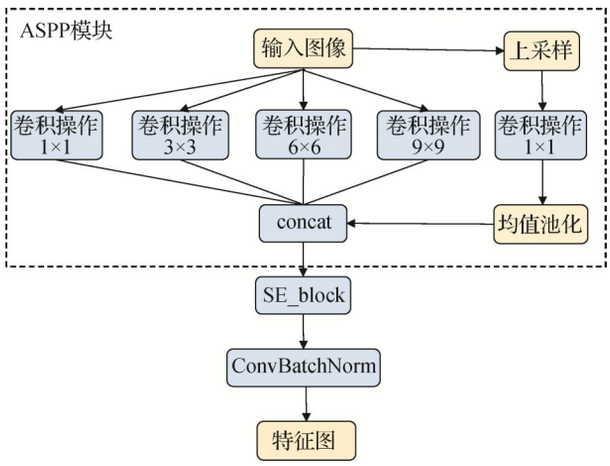


图3　ASPP_SE+ConvBatchNorm模块结构图

Fig. 3　ASPP_SE+ConvBatchNorm module structure diagram

特征更密集的语义信息。本文修改空洞卷积的膨胀率进行实验对比，得出在膨胀率为3，6，9的时候，模型的精度是最高的。

在ASPP后端加入SE通道注意力机制，可以让模型更加关注图像中的重要通道区域，对图像中的信息进行动态选择，在CCT模块进行通道融合中信息权重大的模块会得到更好的融合。SE模块由一个全局平均层和两个全连接层组成。全局平均池化层用来进行空间压缩，对输入为$W \times C \times H$的特征图实现全局平均池化，得到$1 \times 1 \times C$的特征图，然后通过两个全连接层，将这些特征映射到一个可训练的参数上，用于学习通道注意力。最后将得到的通道权重与特征图对应通道进行二维矩阵相乘，以加强重要信息的权重，得到更好的模型结果，SE模块结构如图4所示。

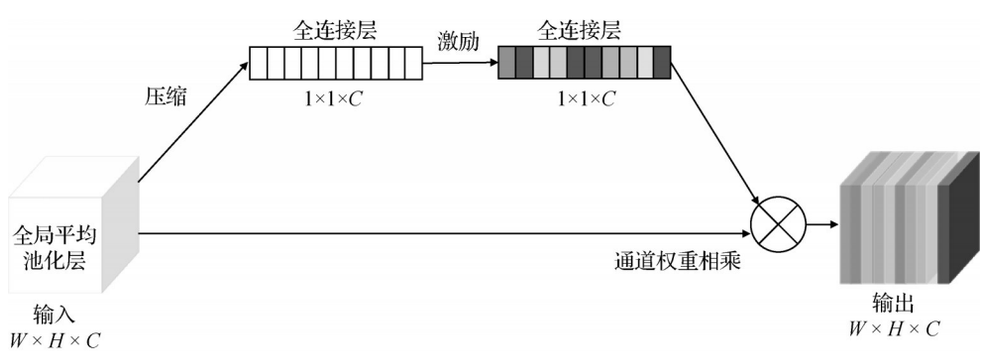


图4　SE模块结构图

Fig. 4　SE module structure diagram

由于腺体图像数据集的稀少，为防止模型训练过程中出现过拟合现象，在提取完特征之后加入ConvBatchNorm模块，该模块由一个3×3卷积层，一个BatchNorm层和ReLU（rectified linear unit）激活函数组成，其中，卷积层进行特征的提取，BatchNorm层可以将数据归一化，减少不同batch之间数据抖动的情况发生，防止过拟合，加快网络收敛的速度。ReLU激活函数可以有效地防止梯度消失。

2.2　编码器和跳跃连接间加入密集连接

U-Net型通道变换网络模型利用通道融合变换器来替换跳跃连接，通道融合变换器可以从通道的角度对编码器提取的特征进行全局特征信息的融合。虽然通道融合变换器全局关注能力很强，但存在局部关注能力过弱的问题，相邻编码器模块之间的歧义性并未得到良好的解决。为解决此问题，本文在模型编码器和跳跃连接的连接处，嵌入密集连接，以U-Net++中跳跃连接的密集结构为参考。但实验时发现加入密集连接后，精度反而下降，可能是过拟合现象导致。通过进行不同层数和不同加入位置的实验对比，发现采用1层密集连接时，密集连接块在CCT模块之前，模型的精度是最高的。密集连接的结构图如图5所示。

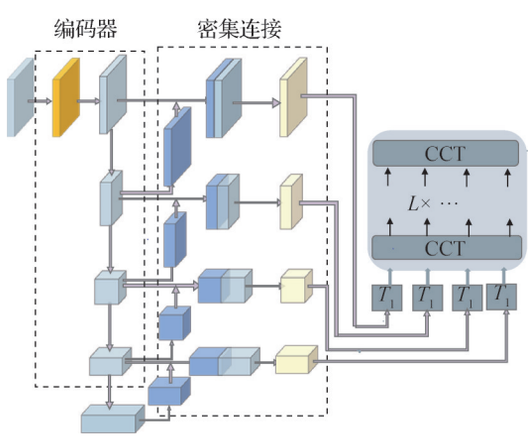


图5　密集连接结构图

Fig. 5　Structure diagram of dense connections

上层编码器模块经过卷积池化得到下层编码器模块，对下层编码器进行上采样操作，使其与上层编码器模块分辨率一致，将这两层编码器进行通道上的拼接，拼接后分辨率不发生改变，但是通道数为两者通道数之和，对拼接后的模块进行ConvBatchNorm模块处理，修正通道数。经过拼接后，上层编码器模块得到了下层编码器模块的特征信息补充，增强了相邻模块之间的语义融合，减少了编码器相邻模块之间的语义差距，再将融合后的模块拆分为小图像块，作为线性序列输入到通道融合变换器。

2.3 自注意力优化

在通道融合变换器融合编码器各个分辨率序列块信息与全局特征信息时，注意力机制起到了关键作用，自注意力机制的强弱代表其融合特征信息的能力。通过增强自注意力能力，可进一步提高模型的分割精度，CCT模块结构如图6所示。图中$\boldsymbol{Q}_i$ (i = 1, 2, 3, 4)为由U-Net通道变换网络编码器部分中不同分辨率大小的序列块线性转化得到，$\boldsymbol{K}$、$\boldsymbol{V}$为编码器不同分辨率大小序列块拼接得到的全局信息，d为序列块的长度，C_Σ为不同分辨率序列块的维度之和，$\psi()$为实例归一化，$\sigma()$为softmax函数。$\boldsymbol{O}_i$ (i = 1, 2, 3, 4)为$\boldsymbol{Q}_i$、$\boldsymbol{K}$、$\boldsymbol{V}$经过自注意力计算得到的不同分辨率序列块与全局特征信息融合的结果。

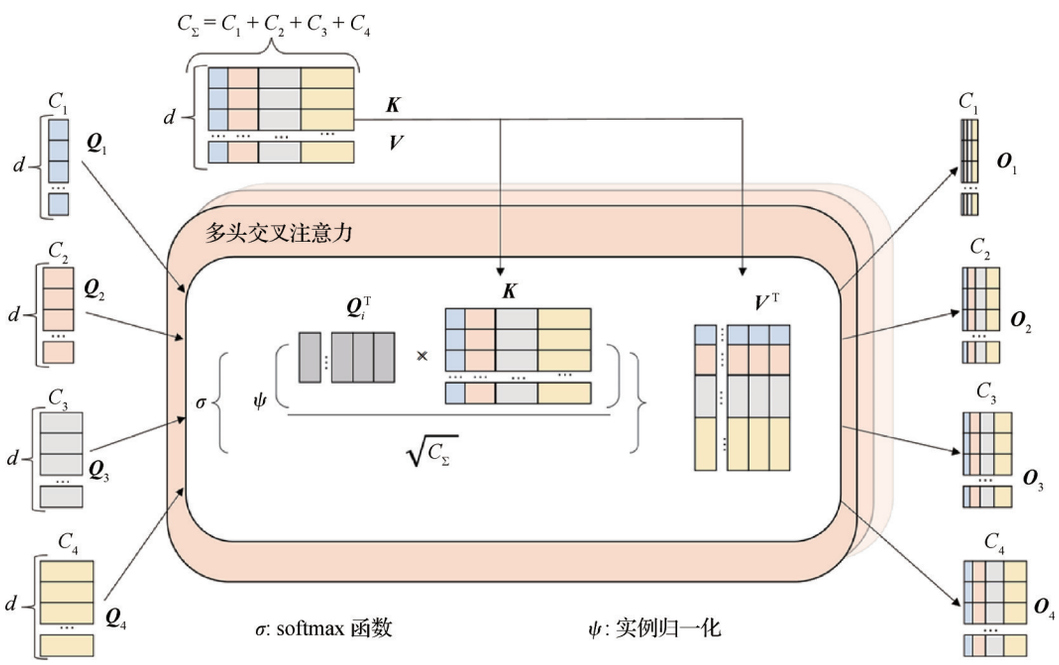


图6 CCT模块结构图

Fig. 6 CCT module structure diagram

自注意力机制计算为

$$\boldsymbol{O}_i = \sigma\left(\psi\left(\frac{\boldsymbol{Q}_{i(i=1,2,3,4)}^{\mathrm{T}}\boldsymbol{K}}{\sqrt{C_\Sigma}}\right)\right)\boldsymbol{V}^{\mathrm{T}} \tag{1}$$

式中，$\boldsymbol{Q}_i \in \mathbf{R}^{C_i \times d}$，$\boldsymbol{K} \in \mathbf{R}^{C_\Sigma \times d}$，$\boldsymbol{V} \in \mathbf{R}^{C_\Sigma \times d}$，通过自注意力机制，不同的序列块$\boldsymbol{Q}_i$关注不同的语义模式，从拼接后的全局信息$\boldsymbol{K}$中自发融合足够多的通道信息，减少编码器不同分辨率模块之间的特征信息差距，有利于模型进行复杂图像的分割。

目前，自注意力增强的大多数研究都是通过设计更复杂的结构或者训练方法，这种方式存在问题，随着训练的加深，会导致自注意力内的不同模块之间越来越相似，大大降低自注意力的性能。为了解决这个问题，进一步提升模型精度，本文在通道融合变换器的自注意力机制中加入细化器DLA（distributed local attention），用于细化注意力的关注图，将多头注意力投射到更高维度空间用于促进多样性的扩展，再利用头部卷积来增强注意力图的空间上下文和局部模式，有效地结合自我注意力和卷积的优点，最后再用一个线性投影，将注意力图恢复到初始的分辨率。细化器由一个线性扩展层、一个线性回溯层和3×3的卷积层组成，其具体结构如图7所示。

具体而言，细化器主要对CCT的自注意机制中

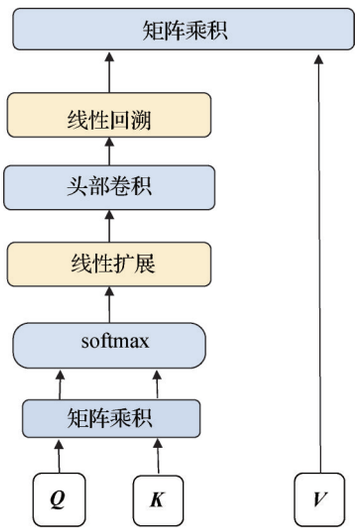


图 7　细化器模块结构图

Fig. 7　Structure diagram of the refiner module

注意力图 $\boldsymbol{A}$ 进行进一步优化，其步骤如图 8 所示。注意力图 $\boldsymbol{A}$ 为 $\boldsymbol{Q}_i$ 与 $\boldsymbol{K}$ 进行相关性计算得到的结果，计算为

$$\boldsymbol{A} = \sigma\left[\psi \frac{\boldsymbol{Q}_i^{\mathrm{T}} \boldsymbol{K}}{\sqrt{C_{\Sigma}}}\right] \tag{2}$$

式中，$\boldsymbol{A} \in \mathbf{R}^{C_i \times C_{\Sigma}}$。在得到注意力图 $\boldsymbol{A}$ 后，使用一个线性投影 $\boldsymbol{W}_d \in \mathbf{R}^{C' \times C_i}$ 将注意力图 $\boldsymbol{A} = \left[\boldsymbol{A}^1, \boldsymbol{A}^2, \cdots, \boldsymbol{A}^{C_i}\right]$ 投影到一个新的注意力图 $\tilde{\boldsymbol{A}} = [\tilde{\boldsymbol{A}}^1, \tilde{\boldsymbol{A}}^2, \cdots, \tilde{\boldsymbol{A}}^{C'}]$，其中 $C' > C_i$，通过线性扩展可以使得注意力图 $\boldsymbol{A}$ 的长度增加，所存储的信息量增多，具体计算为

$$\tilde{\boldsymbol{A}}^{c'} = \sum_{i=1}^{C} \boldsymbol{W}_d(c', i) \times \boldsymbol{A}^i, c' = 1, \cdots, C' \tag{3}$$

经过线性扩展后，本文使用新的注意力图来聚合特征，注意力扩展隐含的增加了注意力机制的头部数量，摆脱了特征划分。重新使用可学习的标量对 $\boldsymbol{Q}_i$ 进行权重划分，有效的提高注意力的多样性。之后，本文利用卷积对扩展后的注意力矩阵进行卷积操作，对特定位置进行重新加权，再对局部聚集的特征进行求和，该方法在全局上下文聚合（自我关注）和局部上下文建模（卷积）之间建立很好的协同关系其计算为

$$\tilde{\boldsymbol{A}}_{i,j}^{C'} = \sum_{a,b=1}^{k} \boldsymbol{W}_{a,b} \times \tilde{\boldsymbol{A}}_{i-\left\lfloor \frac{k}{2} \right\rfloor + a, j-\left\lfloor \frac{k}{2} \right\rfloor + b}^{C'} \tag{4}$$

式中，$\tilde{\boldsymbol{A}}_{i,j}^{C'}$ 为经过头部卷积处理后的数值，卷积核的大小根据不同数据集进行测验调节，本文采用卷积核 $k = 3$ 的卷积进行头部卷积计算。最后利用一个线性投影将注意力图还原回原来的注意力图 $\boldsymbol{A}$，细化器不改变输入模块的分辨率大小，经过细化器处理后的输出的特征矩阵可直接输入到之后的模块中。

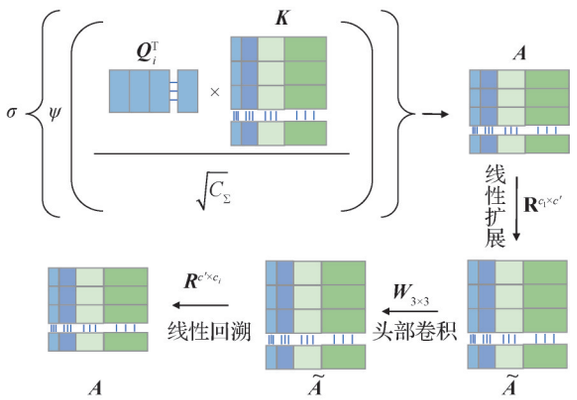


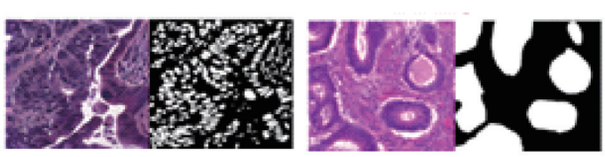
图 8　细化器运算步骤图

Fig. 8　Refiner operation steps diagram

3　实验分析

3.1　实验数据集

本实验采用 MoNuSeg（multi-organ nuclei segmentation challenge）数据集和 Glas（gland segmentation）数据集，如图 9 所示。二者均为公共数据集，MoNuSeg 数据集是通过仔细注释几名患有不同器官肿瘤并在多家医院被诊断的患者的组织图像获得的，该数据集将能够训练细胞分割模型实现高精度的细胞核分割技术。Glas 数据集由来自 T165 期或 T16 期结直肠腺癌的 3 个 H&E（hematoxylin and eosin stains）染色组织学切片的 42 幅图像组成。在染色分布和组织结构方面都表现出很高的受试者间变异性。



(a) MoNuSeg数据集　　(b) Glas数据集

图 9　医学数据集

Fig. 9　Medical datasets((a)MoNuSeg;(b)Glas)

3.2　实验平台

网络优化实验基于 python3. 7 环境，采用 pytorch 作为深度学习的框架，利用 CUDA11. 0 和 CUDNN，加快 GPU 运算。设置 Adam 作为模型的优化器，初始学习率设定为 0. 000 1，每个批次训练 4 幅图像，

输入图像的分辨率为244×244像素，序列块的分辨率为16×16像素。在硬件方面，本文基于Unbutu系统的计算机上进行实验，该计算机配备NVIDIA GeForce RTX 2080，12 GB显存，足够运行本文所有的网络优化实验。

3.3 模型评估方法

本文采用医学图像分割算法中的常用性能指标平均：索伦森—骰子系数（Dice）和交并比（intersection over union，IoU）系数对改进后的模型进行评估。实验结果采用3次5折交叉验证法，将原始数据分成5组，将每个子数据集分别做一次验证集，结果加和平均得到最后的结果。

Dice是一种集合相似度度量函数，用来表示两个样本的相似度，是医学图像分割中使用频率最高的指标。*Dice*系数越高，表示二者之间的相似性越高。

IoU是一种用来衡量结果位置信息准确度的一个标准。IoU的值越大，推测的物体区域就越准确。

3.4 实验结果与分析

3.4.1 消融实验

为准确论证所阐述改进点的有效性，本文以参考模型在两个数据集上的实验结果为基准，逐一进行对比参照实验，除改进点不同其余实验环境均相同，在MoNuSeg数据集上的实验结果如表1所示。

表1 在MoNuSeg数据集上的消融实验结果

Table 1 Ablation experiments on MoNuSeg dataset

组别	ASPP_SE	密集连接	细化器	Dice	IoU
1	–	–	–	0.796 7	0.665 1
2	√	–	–	0.802 2	0.671 9
3	–	√	–	0.801 4	0.670 5
4	–	–	√	0.800 7	0.669 4
6	√	√	–	0.804 3	0.673 8
7	√	√	√	**0.805 5**	**0.675 7**

注：加粗字体表示各列最优结果，“√”表示加入了该改进点，“–”表示未加入。

本文采用同样的方式对Glas数据集进行改进点测试，实验结果如表2所示。

根据表1与表2所展示的数据可以得到，本文提出的改进方法在提高腺体图像分割精度上具有一定有效性，模型加入创新点后，在两个数据集上的分割精度显著提升，不同改进点之间相互结合，可产生更好的实验效果。在观察结果时，可看到加入密集连接后，精度提升幅度最明显，说明跳跃连接中将局部融合与全局融合搭配使用具有一定参考性。

表2 在Glas数据集上的消融实验结果

Table 2 Ablation experiments on Glas dataset

组别	ASPP_SE	密集连接	细化器	Dice	IoU
1	–	–	–	0.907 0	0.839 6
2	√	–	–	0.918 0	0.854 2
3	–	√	–	0.919 2	0.854 8
4	–	–	√	0.915 2	0.850 7
6	√	√	–	0.921 5	0.860 4
7	√	√	√	**0.922 3**	**0.863 9**

注：加粗字体表示各列最优结果，“√”表示加入了该改进点，“–”表示未加入。

3.4.2 密集连接层数对比

本实验第2个创新点以U-Net++网络结构中的多层密集连接为参考，密集连接层融合相邻模块特征信息，CCT模块负责融合全局模块特征信息。但简单地将两个结构拼接在一起，并未起到很好的作用，甚至精确度发生了下降。研究发现，可能产生了过拟合现象。本文在加入其他模块的情况下，进行多次实验对比，发现在减少密集连接层数后，精确度得到了大幅提升，选用1层密集连接的时候实验效果最好，本文以Glas数据集为例，将不同层数的实验结果进行展示，如图10所示。

根据实验所展示的结果可以看到，当层数为1的时候，模型的准确度是最高的，任意增加密集连接

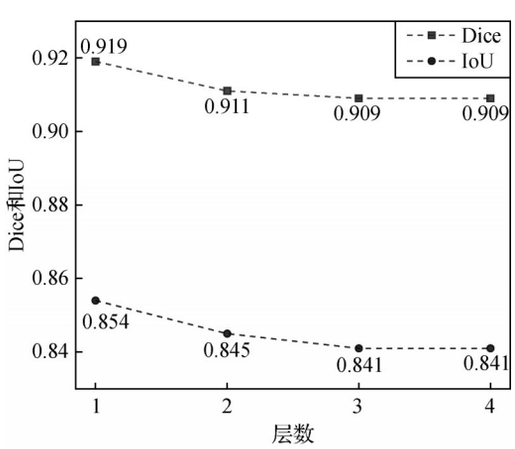


图10 不同层数对比图

Fig. 10 Comparison chart of the number of different layers

的层数后，模型的精度大幅度降低，产生了过拟合现象，层数2、3、4的结果接近。为了进一步验证密集连接的效果，本文采用包含不同层数密集连接的分割网络对腺体图像进行分割预测，得出分割结果图，并将其与原网络的分割结果图进行对比。

不同层数的分割结果如图11所示。从图11可以看出，层数为1层时，模型对凹陷区域和缝隙的关注更加明显。例如红色方块标注的地方为一块凹陷区域，当层数为2、3和4层的时候，此处凹陷被填充，原网络此处的效果没有加入层数为1的密集连接的分割效果好；蓝色方块标记的缝隙，原网络此处缝隙被忽略，添加1层密集连接后，模型对腺体医学图像的局部特征把握能力增强，此处的缝隙分割效果变好，随着层数增加，模型细节关注能力减少，分割的缝隙变得细小，模型分割效果也变得越来越差。

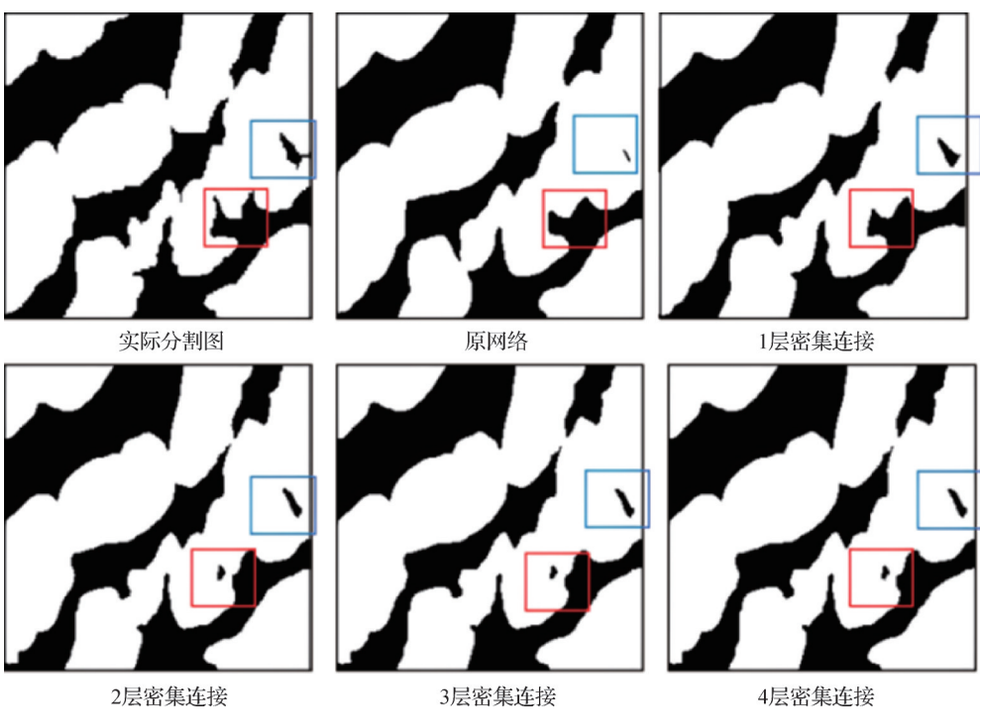


图11　不同层数预测图

Fig. 11　Prediction charts for different layers

3.4.3　密集连接层添加位置对比

加入密集连接层后，本文对密集连接与CCT之间组合进行研究，设置了3种结构进行训练，具体展示如图12所示，并对其实验结果进行分析。

第1种搭配结构是将密集连接层加在CCT模块的前面，先局部信息融合再全局信息融合；第2种是将密集连接层添加在CCT模块的后面，先全局信息融合再进行局部信息融合；第3种是将密集连接层与CCT模块并联，局部融合与全局融合同时进行，将所得到的结果通过卷积进行融合。

3.4.4　其他检测算法对比

为了验证本文中的改进方法对腺体图像的分割性能，本文将改进后的算法与U-Net及其他的一些变体，如U-Net++、TransUNet、Swin-UNet、UCTransNet、

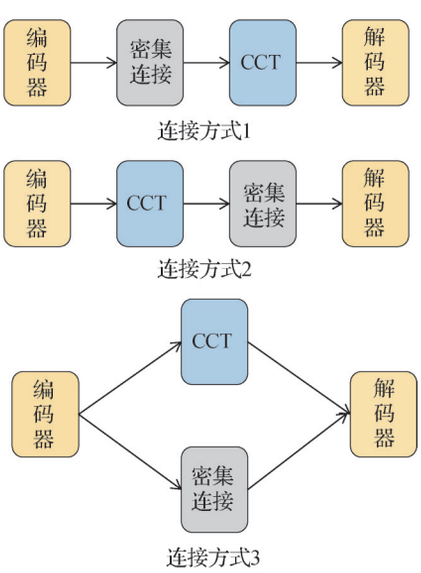


图12　密集连接与CCT结构图

Fig. 12　Dense connection with CCT structure diagram

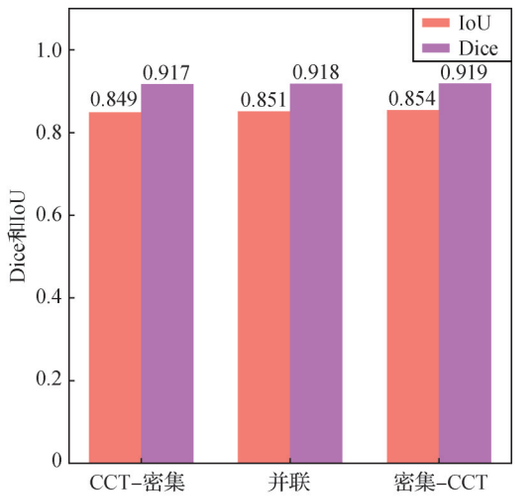


图 13　密集连接与 CCT 不同结构对比

Fig. 13　Comparison of dense connections and CCT structures

DoubleUNet(DCA)(Ates 等,2023)和 HistoSeg(quick attention with multi-loss function for multi-structure segmentation)(Wazir 和 Fraz,2022)模型进行对比,以 Dice 系数和 IoU 系数作为对比指标,采用 early stop 操作,训练 50 次,观察验证集的数值是否存在上升趋势,若 50 次后数值没有上升,模型训练终止。为了防止过拟合,采用 3 次 5 折交叉实验,将训练数据分成 5 个子集,依次将每个子集作为验证集,其余子集作为训练集,可以保证模型在不同的数据子集上进行训练和验证,对 3 次的结果求取平均值作为最后的效果。

实验结果如表 3 所示,通过表 3 可以看到,本文所使用的模型在腺体医学分割的精确度上优于其他 UNet 和 Transformer 结合的分割算法,对比最新的工作 DoubleUNet(DCA)和 HistoSeg,在 Glas 数据集上优于 HistoSeg,在 MoNuSeg 上略低于该网络,后续本文将对其进行研究,对本文网络进一步优化。各个网络分割效果如图 14 所示。

表 3　不同算法的对比

Table 3　Comparison of different algorithms

模型	参数量/M	Flops/G	Glas 数据集		MoNuSeg 数据集	
			Dice/%	IoU/%	Dice/%	IoU/%
U-Net	14.80	50.30	85.45	74.78	76.45	62.86
U-Net++	74.50	94.60	87.56	79.13	77.01	63.04
TransUNet	105.01	56.70	88.40	80.40	78.53	64.83
Swin-UNet	82.30	67.30	89.58	82.06	77.69	63.77
UCTransNet	65.60	63.20	90.70	83.96	79.67	66.51
DoubleUNet (DCA)	30.68	57.82	89.90	81.68	79.50	65.97
HistoSeg	50.81	54.90	90.80	84.05	**81.20**	**68.45**
本文	70.50	61.40	**92.23**	**86.39**	80.55	67.55

注:加粗字体表示各列最优结果。

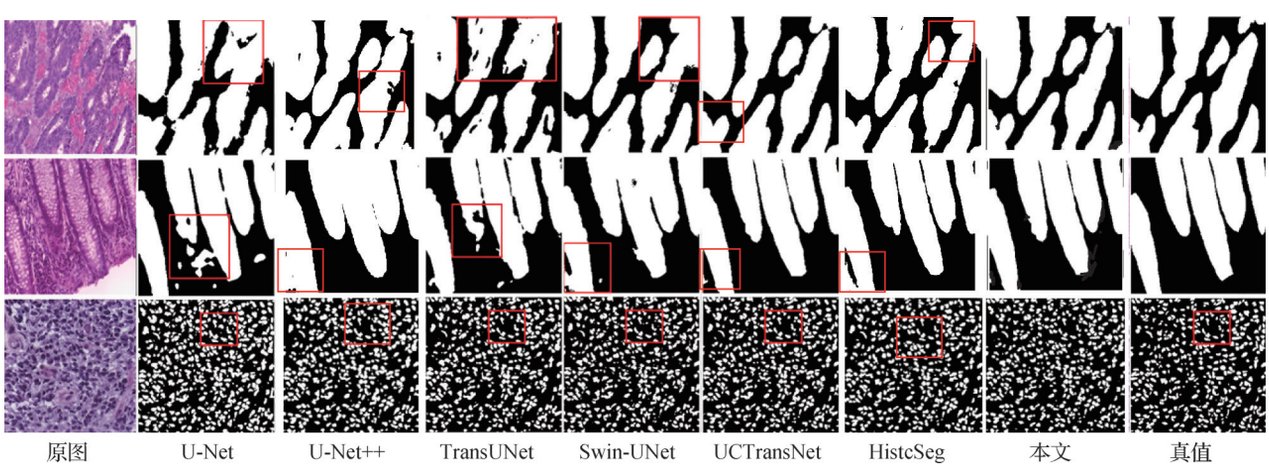


图 14　分割结果对比图

Fig. 14　Comparison of the segmentation results

4　结　论

腺体医学图像分割的难点在于分割目标多,分割目标小,存在分割不准确、漏分割等问题。针对这些问题,本文对 U-Net 型通道变换网络进行了一系列改进,首先在编码器开端加入融合 ASPP_SE 模块与 ConvBatchNorm 模块的组合,利用多尺度空洞卷

积来获取更多小目标信息，提升模型提取小目标信息的能力，同时避免模型训练出现过拟合现象；其次在编码器与跳跃连接中引入简化后的密集连接，补充局部信息的融合；最后在通道融合变换器（CCT）中加入细化器，增强CCT的自注意力能力，增强模型对编码器全局信息的融合，密集连接与增强后的CCT模块结合使用，可增强模型整体语义融合能力。本文提出的改进方法可以更有效地注意到腺体医学图像中的微小信息，同时更加注重图像全局信息，在分割精度上有了很大的提高。但是本实验只针对腺体进行了研究，对于其他的医学图像未进行太多的训练与实验。未来将会对更加复杂和多样的医学图像进行研究，使模型可以更好地应用于更多的医学分割任务。

参考文献（References）

Ates G C, Mohan P and Celik E. 2023. Dual cross-attention for medical image segmentation. Engineering Applications of Artificial Intelligence, 126: #107139 [DOI: 10.1016/j.engappai.2023.107139]

Cao H, Wang Y Y, Chen J, Jiang D S, Zhang X P, Tian Q and Wang M N. 2022. Swin-UNet: UNet-like pure Transformer for medical image segmentation//Proceedings of 2022 European Conference on Computer Vision. Tel Aviv, Israel: Springer [DOI: 10.1007/978-3-031-25066-8_9]

Chen J N, Lu Y Y, Yu Q H, Luo X D, Adeli E, Wang Y, Lu L, Yuille A L and Zhou Y Y. 2021. TransUNet: Transformers make strong encoders for medical image segmentation [EB/OL]. [2022-12-05]. https://arxiv.org/pdf/2102.04306.pdf

Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2014. Semantic image segmentation with deep convolutional nets and fully connected CRFs//Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA: [s.n.]

Çiçek O, Abdulkadir A, Lienkamp S S, Brox T and Ronneberger O. 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation//Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens, Greece: Springer [DOI: 10.1007/978-3-319-46723-8_49]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N. 2021. An image is worth 16 × 16 words: Transformers for image recognition at scale//Proceedings of the 9th International Conference on Learning Representations. [s. l.]: OpenReview.net

Huo L, Hu X X, Xiao Q, Gu Y J, Chu X and Jiang L. 2021. Automatic segmentation of breast and fibroglandular tissues in DCE-MR images based on nnU-Net. Chinese Journal of Magnetic Resonance, 38(3): 367-380 (霍璐, 胡晓欣, 肖勤, 顾雅佳, 褚旭, 姜娈. 2021. 基于nnU-Net的乳腺DCE-MR图像中乳房和腺体自动分割. 波谱学杂志, 38(3): 367-380) [DOI: 10.11938/cjmr20212883]

Jiang X, Yuan Y X, Wang Y P, Xiao Z X, Zhu M L, Chen Z H, Liu T M and Shen D G. 2022. A 20-year retrospect and prospect of medical imaging artificial intelligence in China. Journal of Image and Graphics, 27(3): 655-671 (蒋希, 袁奕萱, 王雅萍, 肖振祥, 朱美芦, 陈泽华, 刘天明, 沈定刚. 2022. 中国医学影像人工智能20年回顾和展望. 中国图象图形学报, 27(3): 655-671) [DOI: 10.11834/jig.211162]

Liu Y, Wu R R, Tang L and Song N N. 2022. U-Net-based mediastinal lymph node segmentation method in bronchial ultrasound elastic images. Journal of Image and Graphics, 27(10): 3082-3091 (刘羽, 吴蓉蓉, 唐璐, 宋宁宁. 2022. U-Net支气管超声弹性图像纵膈淋巴结分割. 中国图象图形学报, 27(10): 3082-3091) [DOI: 10.11834/jig.210225]

Meng Y B, Liu G H, Xu S J and Feng F. 2019. Image segmentation method using multi-resolution Markov random field model with edge-preserving. Journal of Xi'an Jiaotong University, 53(3): 56-65 (孟月波, 刘光辉, 徐胜军, 冯峰. 2019. 一种具有边缘保持的多尺度马尔可夫随机场模型图像分割方法. 西安交通大学学报, 53(3): 56-65) [DOI: 10.7652/xjtuxb201903009]

Milletarì F, Navab N and Ahmadi S A. 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation//Proceedings of the 4th International Conference on 3D Vision (3DV). Stanford, USA: IEEE: #79 [DOI: 10.1109/3DV.2016.79]

Oktay O, Schlemper J, Fdgoc L L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla N Y, Kainz B, Glocker B and Rueckert D. 2018. Attention U-Net: learning where to look for the pancreas [EB/OL]. [2023-04-20]. https://arxiv.org/pdf/1804.03999v3.pdf

Ren C L, Wang N and Zhang Y.2022.Review of medical image segmentation methods. Network Security Technology and Application, 254(2):49-50. (任楚岚,王宁,张阳.2022.医学图像分割方法综述.网络安全技术与应用,254(2):49-50)

Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the 18th Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer [DOI: 10.1007/978-3-319-24574-4_28]

Shelhamer E, Long J and Darrell T. 2014. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: #7298964 [DOI: 10.1109/CVPR.2015.7298965]

Wang H N, Cao P, Wang J Q and Zaïane O R. 2022. UCTransNet: rethinking the skip connections in U-Net from a channel-wise perspective with Transformer//Proceedings of the 36th AAAI Confer-

ence on Artificial Intelligence, AAAI 2022, 34th Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, the 12th Symposium on Educational Advances in Artificial Intelligence. [s.l.]: AAAI Press: #20144 [DOI: 10.1609/aaai.v36i3.20144]

Wazir S and Fraz M M. 2022. HistoSeg: quick attention with multi-loss function for multi-structure segmentation in digital histology images//Proceedings of the 12th International Conference on Pattern Recognition Systems (ICPRS). Saint-Etienne, France: IEEE: #9854067 [DOI: 10.1109/ICPRS54038.2022.9854067]

Xu C, Hao H Y, Wang Y, Ma Y H, Yan Q F, Chen B, Ma S D, Wang X G and Zhao Y T. 2023. Vessel segmentation of OCTA images based on latent vector alignment and Swin Transformer. Journal of Image and Graphics, 28(9): 2927-2939 (许聪, 郝华颖, 王阳, 马煜辉, 阎岐峰, 陈浜, 马韶东, 王效贵, 赵一天. 2023. 融合隐向量对齐和Swin Transformer的OCTA血管分割. 中国图象图形学报, 28(9): 2927-2939) [DOI: 10.11834/jig.220482]

Xu G X, Feng C and Ma F. 2023. Review of medical image segmentation based on UNet. Journal of Frontiers of Computer Science and Technology, 17(8): 1776-1792 (徐光宪, 冯春, 马飞. 2023. 基于UNet的医学图像分割综述. 计算机科学与探索, 17(8): 1776-1792) [DOI: 10.3778/j.issn.1673-9418.2301044]

Zhang J B and Xiao Z Y. 2023. Gland and colonoscopy segmentation method combining self-attention and convolutional neural network. Laser and Optoelectronics Progress, 60(2): #0217002 (张家宝, 肖志勇. 结合自注意力与卷积神经网络的腺体及息肉分割方法. 2023. 激光与光电子学进展, 60(2): #0217002 [DOI: 10.3788/LOP212696]

Zhao B Q, Yu F, Sun J M, Li X M, Yuan L and Xiao L. 2021. Glandular cell segmentation method combined with dense connective blocks and self-attention mechanism. Journal of Computer-Aided Design and Computer Graphics, 33(7): 991-999 (赵宝奇, 尉飞, 孙军梅, 李秀梅, 袁珑, 肖蕾. 2021. 结合密集连接块和自注意力机制的腺体细胞分割方法. 计算机辅助设计与图形学学报, 33(7): 991-999) [DOI: 10.3724/SP.J.1089.2021.18625]

Zhao Y L, Ding W L, You Q H, Zhu F L, Zhu X J, Zheng K and Liu D D. 2023. Classification of whole slide images of breast histopathology based on spatial correlation characteristics. Journal of Image and Graphics, 28(4): 1134-1145 (赵樱莉, 丁维龙, 游庆华, 朱峰龙, 朱筱婕, 郑魁, 刘丹丹. 2023. 融合空间相关性特征的乳腺组织病理全切片分类. 中国图象图形学报, 28(4): 1134-1145) [DOI: 10.11834/jig.211133]

Zhou D Q, Shi Y J, Kang B Y, Yu W H, Jiang Z H, Li Y, Jin X J, Hou Q B and Feng J S. 2021. Refiner: refining self-attention for vision Transformers [EB/OL]. [2023-02-15]. https://arxiv.org/pdf/2106.03714.pdf

Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2018. UNet++: a nested U-Net architecture for medical image segmentation//4th International Workshop, DLMIA 2018, and 8th International Workshop on Deep Learning in Medical Image Analysis -and-Multimodal Learning for Clinical Decision Support-. Granada, Spain: Springer [DOI: 10.1007/978-3-030-00889-5_1]

作者简介

曹伟杰,男,硕士研究生,主要研究方向为医学图像分割。E-mail:cwj15536308910@163.com

段先华,通信作者,男,教授,研究生导师,主要研究方向为图像分割。E-mail:dxh@just.edu.cn

许振伟,男,硕士研究生,主要研究方向为目标检测与图像分割。E-mail:a84249040z@qq.com

盛帅,男,硕士研究生,主要研究方向为目标检测与图像分割。E-mail:2855656766@qq.com