

基于利尿肾图的新型肾实质自动分割算法

朱郭慧, 孙占全

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2024年1月4日; 录用日期: 2024年3月21日; 发布日期: 2024年3月28日

摘要

尽管深度学习图像分割技术在医学图像处理中取得了良好的效果,但由于利尿肾动态显像具有噪声明显、对比度低、图像质量差和边界不清晰的特点,大多数现有的分割方法在获取感受野和提取图像特征信息方面仍然面临很大的挑战。为了解决上述问题,本文提出了一种新的肾实质自动分割算法,该算法采用的是Swin-Transformer编码器与解码器,并且结合了特征融合模块、连续扩张卷积模块与深度注意力模块。特征增强模块中的空间金字塔池模块,可以弥补空间表征并产生多尺度表征。连续扩张卷积能通过获取多尺度上下文聚合来扩大感受野。注意力模块通过顺序捕获多尺度编码器特征之间的关系来解决编码器和解码器特征之间的语义差距。该算法应用于私人的利尿肾造影数据集。仿真实验结果表明,与其他深度学习分割方法比较,该方法可以显著提高肾脏分割性能。

关键词

医学图像分割, 语义分割, 注意力机制, 空间金字塔, 感受野, 膨胀卷积

A Novel Automatic Renal Parenchyma Segmentation Algorithm Based on the Diuretic Renal Image

Guohui Zhu, Zhanquan Sun

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Jan. 4th, 2024; accepted: Mar. 21st, 2024; published: Mar. 28th, 2024

Abstract

Although deep learning image segmentation techniques have achieved good results in medical

image processing, most of the existing segmentation methods still face great challenges in acquiring the sensory field and extracting the image feature information due to the fact that diuretic renal dynamic imaging is characterised by significant noise, low contrast, poor image quality, and unclear boundaries. In order to solve the above problems, this paper proposes a new automatic renal parenchyma segmentation algorithm, which employs a Swin-Transformer encoder and decoder, and combines a feature Enhancement module, a successive dilation convolution module, and a deep attention module. The spatial pyramid pooling module in the feature Enhancement module compensates for spatial representations and produces multi-scale representations. Sequential Expansion Convolution can expand the sensory field by capturing multiscale context aggregation. The Attention module addresses the semantic gap between encoder and decoder features by sequentially capturing relationships between multiscale encoder features. The algorithm is applied to a private diuretic nephrography dataset. The results of simulation experiments show that the method can significantly improve the performance of renal segmentation compared to other deep learning segmentation methods.

Keywords

Medical Image Segmentation, Semantic Segmentation, Attention Mechanisms, Spatial Pyramids, Sensory Fields, Inflated Convolution

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

肾脏疾病及其并发症是人类的医学难题, 严重威胁着人类的健康。小儿肾积水又是肾脏疾病中的常见病。而利尿肾图又是评价小儿肾脏功能的重要工具[1]。利尿肾图在评价分肾功能的同时, 还可以协助临床鉴别机械性尿路梗阻和非机械性尿路扩张。目前已有多种影像学检查方法已用于儿童肾积水的评估, 包括超声、静脉肾盂造影、CT 尿路成像等[2]。而对于肾积水评估的关键在于核医学科医生对于肾实质的人工标注。这是一个耗时又费力, 且不可重复的过程, 所以人工标注的准确性在很大程度上也会取决于医生的经验和理解。

在过去的几十年里, 许多学者对深度学习图像分割算法进行了研究[3]。深度学习由于其强大的非线性特征提取能力, 在图像分类、目标检测、语义分割等方面取得了巨大成功。其中, 卷积神经网络(CNN)就是最经典的算法, 改进的 CNN 可以在医学图像数据集训练样本不足的情况下, 完成精确的自动分割任务[4] [5]。例如, 以全卷积网络(Full Convolutional Network, FCN) [6]或 Unet [7]为代表的一系列模型被提出并广泛应用于医学, 利尿肾图也不例外。Ronneberger 等[7]借鉴了 FCN 的结构, 提出了对称编解码结构的 Unet, 通过网络间的跳过连接, 在生物医学图像分割中表现出优越性。

尽管 CNN 具有出色的表征能力, 但是卷积运算中, 由于 U 形结构感受野过小, 导致在建模过程中, 捕捉远距离关系方面受限。基于 CNN 的方法通常性能较弱, 特别是对于处理纹理和形状方面差异较大的目标。为了克服这种局限性, 人们提出了基于 CNN 的自注意力机制[8]。另一方面, Transformer 在计算机视觉领域也得到了广泛研究, 它完全不使用卷积算子, 而仅仅依赖于注意力机制[9]。Vision Transformer (ViT) [10]通过直接将具有全局自关注的 Transformer 应用于全尺寸图像, 在 ImageNet 数据集上实现了最先进的分类。受 ViT 的启发, TransUNet [11]进一步结合了 Transformer 和 U-Net 的优势, 用于医学图像分割。然而, TransUNet 的性能仍然受到医学图像数量的限制。而且, 传统的 Transformer 还有一个典型

的问题, 就是它的计算复杂度。为了降低复杂度, Swin-Transformer [12]使用本地窗口的自注意力来代替全局自注意力并划分移位的窗口, 在各窗口之间的前一层建立连接。

为了解决这些问题, 我们将所提出的算法应用于一个实际性的临床问题: 在上海交通大学附属新华医院提供的利尿肾图数据集中进行肾脏自动分割。我们的工作贡献可以概括如下:

1) 提出了一种基于 Swin-Transformer 的编码器-解码器结构, 并且优化了 U 型结构, 在编码器每个阶段插入几个连续的不同膨胀速率的膨胀卷积, 以获得足够的多尺度上下文聚合, 该网络比单一网络能提取更多的全局特征。此外, 利用特征增强模块通过扩展接受野来细化提取的深层特征, 实现了对利尿肾动态显像中肾实质的分割。

2) 在跳跃连接中加入交叉注意力模块, 它通过捕获多尺度编码器之间的通道和空间依赖关系来解决编码器和解码器之间的语义差距。

3) 设计了一个深度注意模块来融合提取的高级特征, 该模块融合编码器的最后一个特征提取层和来自特征融合模块的特征, 强制网络关注感兴趣区域。

2. 相关工作

2.1. 传统机器学习方法

传统方法在深度学习发展之前, 许多传统的基于机器学习的方法已经应用于医学图像处理。Halkar 等人[13]提出了一种单阈值方法, 可以根据利尿肾造影自动绘制肾脏轮廓。但是肾造影的对比度相对较低, 很难选择合适的阈值来有效区分前景和背景。Tomaru 等人[14]提出了一种改进的双阈值半自动分割方法, 但它仍需要手动操作指定两个肾脏的中心位置。Inoue 等人[15]提出了另一种半自动分割方法, 该方法将预设的矩形框手动移动到不同的 ROI, 利用每个部位不同的放射性峰值时间来减少非肾区域的干扰。这降低了对操作员的要求, 但仍占用了大部分时间。Tian 等人[16]提出了一种新的半自动分割方法, 减少了人工参与。然而, 当肾功能严重受损时, 很难定位肾脏区域。这些半自动方法或多或少需要人类的参与, 并且在一定程度上是主观的。同时, 这些方法完全依赖于图像, 并且算法的鲁棒性较差[17]。

2.2. 深度学习方法

Xie 等人提出了 SERUnet [18], 这是一种用于肾脏和肿瘤分割的级联 SE-ResNeXTUnet。首先, 进行粗分割, 在 CT 图像中找到肾脏的关键切片。然后将得到的 patch 作为输入, 对分割进行细化。Wu 等[19]使用级联的全卷积 DenseNet 对超声图像进行肾脏自动分割。通过预处理可以避免噪声和冗余信息的干扰。Chen 等人[20]提出了 DeepLabv3+, 通过增加一个简单而有效的解码器模块来扩展 DeepLabV3, 以提高分割性能。Hsiao 等人[21]提出了一种用于肾脏分割的编码器-解码器架构。实现了超参数优化过程, 它允许计算总肾体积来估计 ADPKD 的肾功能。Azad 等人[22]对 DeepLabv3+模型进行建模, 并通过使用 Swin-Transformer 作为架构的基本组件来建立 TransDeepLab。该模型在各种医学图像分割任务中展示了有效建立远程依赖关系的潜力。

受先前工作的启发, 提出了一种基于 Swin-Transformer 和注意力机制结合的编码器-解码器框架。它继承了 Swin-Transformer 的优点, 有效提取全局和局部信息, 对肾实质进行有效的分割。并且所提模型对架构进行了优化, 在跳跃连接中引入了深度注意力模块, 以有效进行信息交互。该方法通过对原始输入图像中的肾实质 ROI 区域进行分割, 我们的模型成功地生成了肾 ROI 区域, 实现了有效的分割。

3. 网络框架

3.1. 网络主体结构

基于利尿剂肾动态成像数据集, 我们提出了一种基于 Swin-Transformer 的肾图自动分割网络。我们

提出的网络结构是基于 U 型编码器 - 解码器网络构建的, 其总体框架如图 1 所示。与常规 U 型结构网络不同, 我们的网络结合改进了编码器、解码器和跳转连接, 优化了 U 型结构。给定医学图像 $I \in R^{3 \times H \times W}$ 的输入, 其中, $H \times W$ 表示图像的空间分辨率。图像先被输入 Swin-Transformer 模块, 用于捕捉全局信息并构建分层特征表示, 而后我们在每个阶段插入几个连续的不同扩展速率的扩展卷积, 以获得足够的多尺度上下文聚合。在最后一层下采样中引入特征增强模块和深度注意模块, 可以弥补空间表征并产生多尺度表征, 并且通过深度监督的方式来获得具有判别性的高级特征。最后, 通过 Swin-Transformer 解码器可以获得 $H \times W \times C$ 的分割图。

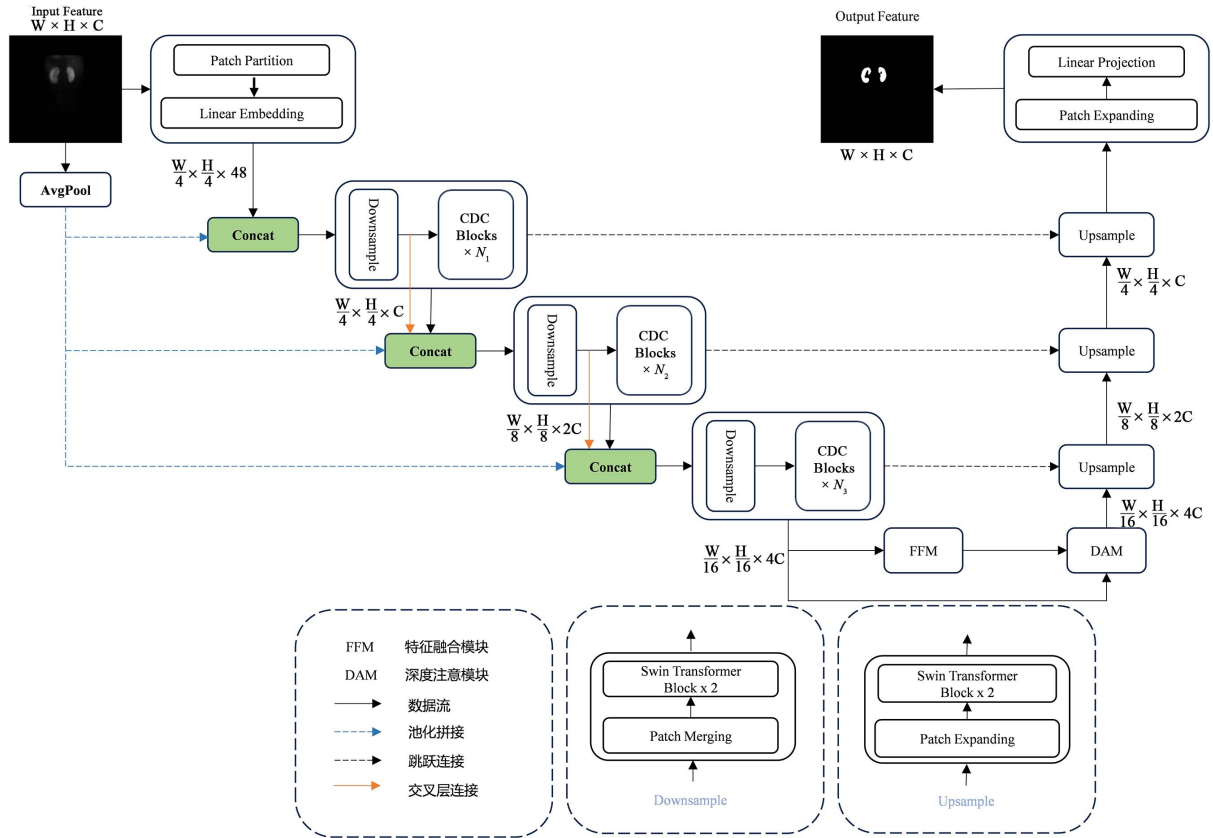


Figure 1. General framework of the proposed model
图 1. 所提出模型的总体框架

3.2. Swin-Transformer 模块

与传统的多头自聚焦(MSA)区块不同, Swin-Transformer 区块[12]是基于移位窗口构建的。图 2 给出了两个连续的 Swin-Transformer 块。每个 Swin-Transformer 块由一个 LayerNorm (LN)层、一个多头自聚焦模块、一个残差连接和一个具有 GELU 非线性的 2 层 MLP 组成。分别使用了基于窗口的多头自关注 (W-MSA)模块和基于移位窗口的多头自关注(SW-MSA)模块。基于这种窗口划分机制, 连续旋转变压器块可表示为:

$$\hat{z}^l = W - MSA(LN(z^{l-1})) + z^{l-1} \tag{1}$$

$$z^l = MLP(LN(\hat{z}^l)) + \hat{z}^l \tag{2}$$

$$\hat{z}^{l+1} = SW - MSA\left(LN\left(z^l\right)\right) + z^l \quad (3)$$

$$\hat{z}^{l+1} = MLP\left(LN\left(\hat{z}^{l+1}\right)\right) + \hat{z}^{l+1} \quad (4)$$

其中 \hat{z}^l 和 z^l 分别表示第 l 个块的(S)W-MSA 模块和 MLP 模块的输出。根据先前的研究[23] [24], 自我注意力的计算公式如下:

$$Attention(Q, K, V) = Softmax\left(QK^T / \sqrt{d} + B\right)V \quad (5)$$

其中 $Q, K, V \in R^{M^2 \times d}$ 表示查询矩阵、密钥矩阵和值矩阵。 M^2 表示一个窗口内的面片数量。 d 是查询或键的大小。 $B \in R^{M^2 \times M^2}$ 表示相对位置偏移, B 中的值取自偏移矩 $\hat{B} \in R^{M^2 \times M^2}$ 。

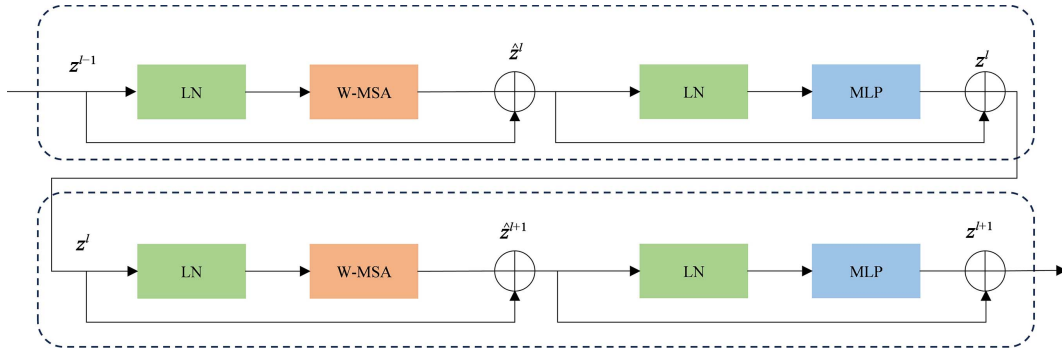


Figure 2. Swin Transformer block

图 2. Swin Transformer 模块

3.3. 连续膨胀卷积模块

如图 3 所示, 本文提出的连续膨胀卷积模块(CDC)利用扩张卷积来提取多尺度局部特征。与仅在网络的最后一层使用并行不同, 我们在网络的最后一层插入了多个连续的扩张卷积模块。不同的是, 我们在每个阶段都插入了几个连续的稀释卷积, 以实现充分的多尺度上下文聚合。给定一个二维信号 $x[i]$, 二维稀释卷积的输出 $y[i]$ 可以定义为二维扩张卷积的输出可定义为:

$$y[i] = \sum_{k=1}^K x[i+r \cdot k]w[k] \quad (6)$$

其中, $w[k]$ 为长度为 k 的滤波器, r 为用于卷积输入 $x[i]$ 的膨胀率。在标准非展开卷积中 $r=1$ 。通过使用扩展卷积, 网络可以保持输出特征映射的大小固定, 同时获得更大的感受野。考虑一个尺寸为 $H \times W \times C$ 的输入特征 X , CDC 模块输出 \hat{X} 为:

$$\hat{X} = X + Linear_G\left(Linear\left(BN\left(DDWConv_r(X)\right)\right)\right) \quad (7)$$

其中, $Linear_G$ 表示逐点卷积操作, 然后是 GELU [25] 激活。BN 为批归一化层, $DDWConv_r(\cdot)$ 为 3×3 深度膨胀卷积, 扩张速率为 r 。

3.4. 特征增强模块

如图 4 所示, 在医学图像分割领域, 特征表示的准确性直接影响到网络的分割和识别效果。为此, 我们引入特征增强模块, 旨在通过吸收补充信息来增强提取的特征, 从而增强网络的判别能力。特征增强模块在不增加参数数量的情况下, 使用了一个具有不同窗口大小(窗口大小 = 4, 6, 8, 12)的 Swin 空

间金字塔池化模块, 以捕捉多尺度表示。在我们的设计中, 较小的窗口尺寸旨在捕捉局部信息, 而较大的窗口尺寸则用于提取全局信息。然后, 生成的多尺度表征被输入到跨语境注意力模块, 以非线性技术融合并捕捉通用表征。

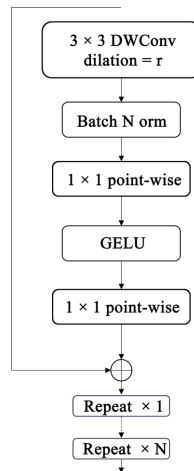


Figure 3. Continuous expansion convolution module
图 3. 连续膨胀卷积模块

在 Swin 空间金字塔池化之后, 由通道注意机制生成通道注意力图(Channel Attention Map, CAM)。该 CAM 基于通道间的关系, 使网络能够识别和优先考虑显著的特征通道, 从而在医学图像中聚焦相关结构, 同时减轻无关的背景噪声。连接的特征集通过 3×3 卷积和批处理归一化层进一步细化。通过 1×1 卷积处理与原始输入特征的加性融合来补充这一点, 确保原始特征完整性与增强特征集的保留和整合。最后, 特征集输入到空间注意机制(Spatial Attention Module, SAM), 可自适应地重新校准空间权重分布。这种机制在突出重要的空间区域和减少非必要背景元素的影响方面至关重要。通过进一步改进了特征表示, 提高了网络在医学图像分割任务中的性能和准确性。

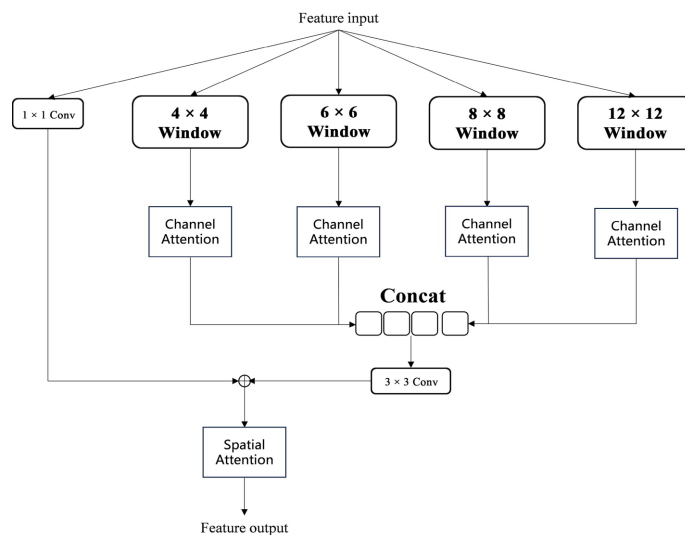


Figure 4. Feature enhancement module
图 4. 特征增强模块

3.5. 深度注意模块

如图 5 所示为深度注意模块的结构。该模块以原始模型中特征提取的最后一层和特征增强模块的特征为输入。这些特征是在点向卷积后添加的, 其中通道减少率设置为 4。经过 ReLU、点向卷积和 Sigmoid, 我们得到了与原始输入特征大小相同的注意力图。最后, 将提取的注意力与来自特征增强模块的特征相乘, 得到最终的注意力图。

对于利尿肾图, 存在较多的干扰, 目标与背景之间的对比度很低。因此, 为了融合多尺度的特征并迫使网络专注于突出的目标区域, 受[8]的启发, 我们使用了一种新的策略, 称为深度注意模块。深度注意模块将下采样的最后一层输出的特征和特征增强模块输出的特征为输入, 输出的融合注意力图 F_{out} , 我们将下采样的最后一层和特征增强模块输出的特征分别定义为 F_R 和 F_A 。为了减少网络模型的参数, 我们选择了点向卷积来提取特征, 用 PConv 表示。因此, 深度注意模块可以表示为:

$$F_{att} = \sigma_2 \left(PConv \left(\sigma_1 \left(PConv(F_R) + PConv(F_A) \right) \right) \right) \quad (8)$$

$$F_{out} = F_{att} \times F_A \quad (9)$$

其中 F_{att} 表示获得的注意力图。 $\sigma_1(\cdot)$ 表示 ReLU 激活函数, $\sigma_2(\cdot)$ 表示 Sigmoid 激活函数。 F_{out} 表示最终的注意特征图, 然后是全局自适应平均池化层和完全连接层, 输出特定疾病的预测概率。因此, 我们可以利用深度关注模块, 通过深度监督的方式来获得具有判别性的高级特征。

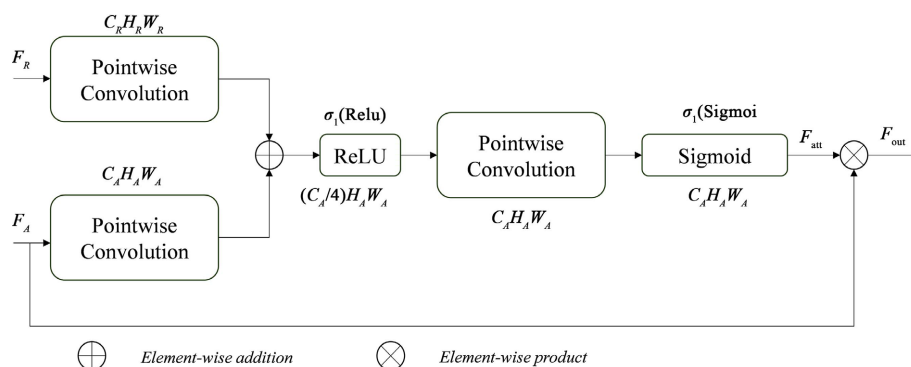


Figure 5. Feature fusion module

图 5. 特征融合模块

4. 实验

4.1. 实验设置

我们提出的模型使用我们精心收集的肾脏动态成像数据集进行了评估。训练和测试的仿真在配备 24 GB 内存和英伟达 RTX 3090 的强大 GPU 上使用 PyTorch 执行。在仿真训练过程中, 我们配置了特定的模型参数: 输入图像大小为 384×384 , 仿真训练过程跨度为 150 个 epoch, 批量大小为 4, 以提高学习效率。为了优化网络梯度和参数, 我们采用了 Adam 优化算法, 初始学习率为 0.0001。

4.2. 数据集

本研究使用的数据集包括从上海交通大学医学院附属新华医院获得的利尿肾图。数据集是在患者的临床诊断过程中使用飞利浦医疗公司的 SPECT 设备采集的。新华医院批准使用这些数据, 其中包括 980 个儿科患者的动态图像序列。原始数据的分辨率为 128×128 像素, 以 DICOM 格式保存。每次诊断持续

20 分钟, 共生成 136 幅原始图像。在第一分钟内, 共采集了 60 幅原始图像, 每幅图像以每秒一帧的频率拍摄; 在随后的 19 分钟内, 共采集了 76 幅原始图像, 每幅图像以每 15 秒一帧的频率拍摄。根据核医学医生准确诊断肾功能的要求, 我们选择了前 2 分钟的原始图像进行肾脏分割, 特别是前 64 帧。选定的原始图像经过合成和增强预处理后, 形成了分辨率为 384×384 像素的 980 幅图像数据集。图 6 显示了原始肾脏标签, 黑色代表肾脏外部, 白色代表肾实质。数据集被随机分成三个子集: 784 张图像分配到训练集, 98 张图像分配到验证集, 98 张图像分配到测试集。

所有原始图像的标注过程都是在一名经验丰富的核医学医生的指导下进行的。对利尿剂肾图中实质区域的识别经过了多次研究, 并与核医学医生进行了讨论。医生进行了多次标记尝试并进行了彻底审查, 反复修改, 以确定肾脏标记的精确准则。最终, 另一名核医学医生确认了所有图像标注, 以确保在利尿剂肾图中准确标注肾实质。

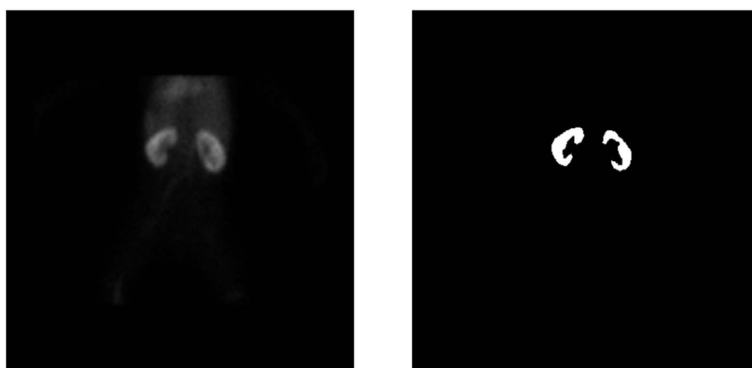


Figure 6. Diuretic renogram (raw data on the left, labels on the right)

图 6. 利尿肾图(左为原始数据, 右为标签)

4.3. 评价指标

分割结果使用四个指标进行评估: DSC、IOU、精确度和召回率。骰子相似系数(DSC)是一个区域级指标, 用于衡量预测结果与地面实况之间的重叠率。它对掩码的内部填充特别敏感。图像分割可以作为单个像素的二元分类任务来处理。通过定量分析真阳性(TP)、真阴性(TN)、假阳性(FP)和假阴性(FN), 我们可以有效地评估分割的性能。这些指标定义如下:

$$Dice = \frac{2TP}{2TP + FN + FP} \quad (10)$$

$$IoU = \frac{TP}{2TP + FN + FP} \quad (11)$$

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

4.4. 实验结果

在本节中, 我们通过与最先进的利尿肾图分割方法进行对比分析, 评估所提出的模型的功效。所有模型都是在与模型相同的参数和仿真环境条件下进行评估的。表 1 中的性能评估使用四个指标来评估效果, 展示了比较结果。在所有评估指标上, 我们的方法始终优于其他方法, 显示出卓越的效果。

表 1 全面比较了用于利尿肾图分割的不同分割方法。仿真实验结果清楚地表明, 所提出的方法在各种评价指标上始终优于其他模型, 展示了其卓越的功效。具体来说, 该方法的最佳 Dice 和 IoU 分数分别为 91.53% 和 83.29%, 比 Unet 有显著提高, Dice 和 IoU 分别提高了 2.7% 和 1.95%。此外, 本方法的 Dice 和 IoU 分别比 TransUnet 提高了 0.92% 和 0.95%, Recall 也提高了 1.4%。与 Swin-Unet 相比, 本方法的 Dice 和 IoU 分别提高了 1.19% 和 1.1%。此外, 在所有模型中, 本方法的仿真得分指标均为最高, 这些结果凸显了本方法与其他模型相比的优越性。

Table 1. Quantitative results on the diuretic renogram dataset

表 1. 利尿肾图数据集上的定量结果

方法	Dice (%)	IoU (%)	Precision (%)	Recall (%)
Unet	88.83	81.34	87.35	90.64
Att-Unet	89.61	82.15	88.24	92.13
Unet++	89.26	81.64	88.36	90.96
PSPNet	88.91	81.56	88.84	91.84
DeepLabV3+	90.11	82.42	91.89	90.15
Swin-Unet	90.34	82.19	91.74	92.34
TransUnet	90.61	82.34	92.14	92.58
Ours (本文方法)	91.53	83.29	92.86	93.18

图 7 直观展示了各种方法的分割结果, 提供了相应结果的定性可视化。前景中的白色区域表示需要分离的独特肾实质。我们选择了四个具有代表性的案例来研究不同网络的性能差异。在案例 1 中, 虽然在检测过程中肾脏以外的部分器官可能会出现在肾脏动态可视化中, 但由于肾功能更加明显, 白色区域得以突出显示, 因此最终预测结果更好。然而, 在肾功能受损的案例 2、3、4 中, 利尿剂肾图中肾实质的亮度明显降低。这给一些网络准确区分所有肾脏区域带来了挑战, 导致预测不完整, 并容易受到邻近器官的干扰。因此, 这导致了不理想的分割结果。与其他方法相比, 本文所提出方法的分割性能更好, 这主要体现在来自非肾脏区域的干扰较少, 网络更专注于对肾实质的学习。对于肾脏内差异不明显的区域识别准确率更高。总体而言, 这些可视化结果验证了我们提出的方法的有效性。

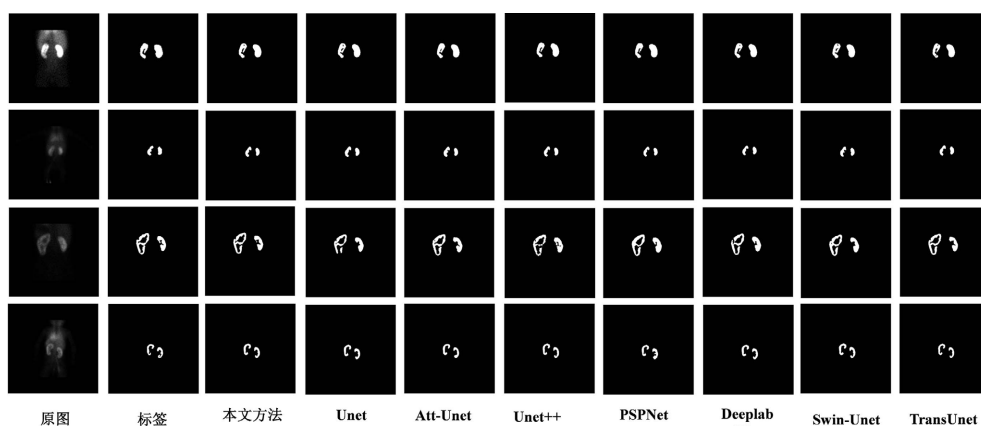


Figure 7. Visualisation of quantitative results from different methods

图 7. 不同方法的定量结果可视化

5. 结论

本文介绍了一种旨在自动进行肾脏分割的创新算法。我们的分割模型结合了 Swin-Transformer、深度注意模块和特征增强模块的优势。通过利用 Swin-Transformer 自我关注机制的先进远程上下文信息建模功能, 显著提高了不同医学图像中语义分割的准确性。特征增强模块在连接空间表征和通过空间金字塔池生成多尺度表征方面发挥了关键作用。此外, 连续扩张卷积能通过获取多尺度上下文聚合来扩大感受野。在我们的仿真实验中, 所提出模型在自动分割利尿肾图中的表现优于许多现有的医学图像分割模型。总之, 我们的模型通过提取深层语义特征来自动识别和分割肾实质, 在准确性上超越了现有模型。但是, 复杂的网络势必会增加更多的训练参数, 训练和推理的速度也会增加, 这将在下一步的工作中进行改进。

基金项目

国家自然科学基金(61703277)。

参考文献

- [1] Gordon, I., Piepsz, A. and Sixt, R. (2011) Guidelines for Standard and Diuretic Renogram in Children. *European Journal of Nuclear Medicine and Molecular Imaging*, **38**, 1175-1188. <https://doi.org/10.1007/s00259-011-1811-3>
- [2] Taha, A., Lo, P., Li, J. and Zhao, T. (2018) Kid-Net: Convolution Networks for Kidney Vessels Segmentation from Ct-Volumes. *Medical Image Computing and Computer Assisted Intervention-MICCAI 2018: 21st International Conference*, Granada, 16-20 September 2018, 463-471. https://doi.org/10.1007/978-3-030-00937-3_53
- [3] Koff, S.A., Thrall, J.H. and Keyes Jr., J.W. (1980) Assessment of Hydronephrosis in Children Using Diuretic Radionuclide Urography. *The Journal of Urology*, **123**, 531-534. [https://doi.org/10.1016/S0022-5347\(17\)56007-3](https://doi.org/10.1016/S0022-5347(17)56007-3)
- [4] Da Cruz, L.B., Araújo, J.D.L., Ferreira, J.L., Diniz, J.O.B., Silva, A.C., De Almeida, J.D.S., De Paiva, A.C. and Gattass, M. (2020) Kidney Segmentation from Computed Tomography Images Using Deep Neural Network. *Computers in Biology and Medicine*, **123**, Article ID: 103906. <https://doi.org/10.1016/j.combiomed.2020.103906>
- [5] Sharma, K., Rupprecht, C., Caroli, A., Aparicio, M.C., Remuzzi, A., Baust, M. and Navab, N. (2017) Automatic Segmentation of Kidneys Using Deep Learning for Total Kidney Volume Quantification in Autosomal Dominant Polycystic Kidney Disease. *Scientific Reports*, **7**, Article No. 2049. <https://doi.org/10.1038/s41598-017-01779-0>
- [6] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [7] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference*, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [8] Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B. and Rueckert, D. (2019) Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images. *Medical Image Analysis*, **53**, 197-207. <https://doi.org/10.1016/j.media.2019.01.012>
- [9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Red Hook, 4-9 December 2017, 6000-6010.
- [10] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G. and Gelly, S. (2020) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale.
- [11] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L. and Zhou, Y. (2021) Transunet: Transformers Make Strong Encoders for Medical Image Segmentation.
- [12] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. and Guo, B. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 10012-10022. <https://doi.org/10.1109/ICCV48922.2021.00986>
- [13] Halkar, R., Chrem, Y., Galt, J., Rothfus, B., Grant, S., Jones, M., Folks, R., Alazraki, N. and Taylor, A. (1996) Interoperator Variability in Quantitating the MAG3 Renal Uptake Based on Semiautomated and Manual Regions of Interest. *Journal of Nuclear Medicine*, **37**, 285-293.
- [14] Tomaru, Y., Inoue, T., Oriuchi, N., Takahashi, K. and Endo, K. (1998) Semi-Automated Renal Region of Interest Se-

- lection Method Using the Double-Threshold Technique: Inter-Operator Variability in Quantitating 99m Tc-MAG3 Renal Uptake. *European Journal of Nuclear Medicine*, **25**, 55-59. <https://doi.org/10.1007/s002590050194>
- [15] Inoue, Y., Yoshikawa, K., Yoshioka, N., Watanabe, T., Saegusa, S., Kaneko, Y., Yokoyama, I. and Ohtomo, K. (2000) Evaluation of Renal Function with 99mTc-MAG3 Using Semiautomated Regions of Interest. *Journal of Nuclear Medicine*, **41**, 1947-1954.
- [16] Sun, P., Mo, Z., Hu, F., Song, X., Mo, T., Yu, B., Zhang, Y. and Chen, Z. (2022) Segmentation of Kidney Mass Using AgDenseU-Net 2.5D Model. *Computers in Biology and Medicine*, **150**, Article ID: 106223. <https://doi.org/10.1016/j.compbiomed.2022.106223>
- [17] Jing, X., Zheng, X., Song, S. and Liu, K. (2017) A Computer Tomography Assisted Method for the Automatic Detection of Region of Interest in Dynamic Kidney Images. *Journal of Biomedical Engineering*, **34**, 928-933.
- [18] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., *et al.* (2018) Attention U-Net: Learning Where to Look for the Pancreas.
- [19] Wu, Z., Hai, J., Zhang, L., Chen, J., Cheng, G. and Yan, B. (2019) Cascaded Fully Convolutional DenseNet for Automatic Kidney Segmentation in Ultrasound Images. 2019 *2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, Chengdu, 25-28 May 2019, 384-388. <https://doi.org/10.1109/ICAIBD.2019.8836994>
- [20] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 801-818. https://doi.org/10.1007/978-3-030-01234-2_49
- [21] Hsiao, C.H., Lin, P.C., Chung, L.A., Lin, F.Y., Yang, F.J., Yang, S.Y., *et al.* (2022) A Deep Learning-Based Precision and Automatic Kidney Segmentation System Using Efficient Feature Pyramid Networks in Computed Tomography Images. *Computer Methods and Programs in Biomedicine*, **221**, Article ID: 106854. <https://doi.org/10.1016/j.cmpb.2022.106854>
- [22] Azad, R., Heidari, M., Shariatnia, M., Aghdam, E.K., Karimijafarbigloo, S., Adeli, E. and Merhof, D. (2022) Transdeeplab: Convolution-Free Transformer-Based Deeplab V3+ for Medical Image Segmentation. In: *International Workshop on Predictive Intelligence in Medicine*, Springer, Berlin, 91-102. https://doi.org/10.1007/978-3-031-16919-9_9
- [23] Hu, H., Gu, J., Zhang, Z., Dai, J. and Wei, Y. (2018) Relation Networks for Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 3588-3597. <https://doi.org/10.1109/CVPR.2018.00378>
- [24] Hu, H., Zhang, Z., Xie, Z. and Lin, S. (2019) Local Relation Networks for Image Recognition. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 3464-3473. <https://doi.org/10.1109/ICCV.2019.00356>
- [25] Hendrycks, D. and Gimpel, K. (2016) Bridging Nonlinearities and Stochastic Regularizers with Gaussian Error Linear Units.