

基于改进 UNet 模型的土壤 CT 图像分割方法^①

宋忠林, 闫娇, 马韞韬, 李保国, 周虎*

(中国农业大学土地科学与技术学院, 北京 100080)

摘要: X 射线断层扫描(CT)已成为土壤孔隙结构研究的主要手段, 而 CT 图像分割作为其中关键环节, 直接关系到孔隙分析的准确性。因传统的图像分割方法精度有限, 近年来深度学习技术被广泛应用于土壤 CT 图像分割中。NAM(Normalization-based attention module)是一种高效的注意力机制, 改进了空间注意力和通道注意力子模块的设计, 能够增强深度学习模型分割的效果。本研究采用 UNet、UNet++、DeepLabV3 以及加入了 NAM 注意力机制的 AttentionUNet 4 种深度学习模型, 对我国东北地区黑土的 CT 图像进行分割, 并同传统分割方法进行比较。结果表明, AttentionUNet 在交并比、F1 分数、精确度和召回率 4 个指标的评估值分别达到 88.08%、0.96、96.02%、96.64%, 精确度相对于 UNet 和 UNet++ 分别提升了 1.46 和 0.21 个百分点, 较传统 Watershed 方法的精确度和召回率分别提升 101.13% 和 33.54%。本文提出的 AttentionUNet 模型可有效提高土壤 CT 图像分割的效率和精度, 为土壤 CT 图像的孔隙分割提供一种更加高效的方法。

关键词: UNet; NAM 模块; CT 图像; 图像分割; 注意力机制

中图分类号: TP391.4 **文献标志码:** A

Soil CT Image Segmentation Based on Improved UNet Method

SONG Zhonglin, YAN Jiao, MA Yuntao, LI Baoguo, ZHOU Hu*

(College of Land Science and Technology, China Agricultural University, Beijing 100080, China)

Abstract: X-ray tomography (CT) has become the main means of soil pore structure research, and CT image segmentation, as one of the key links, is directly related to the accuracy of pore analysis. Traditional image segmentation methods have limited accuracy, and in recent years deep learning techniques have been widely used in soil CT image segmentation. NAM (Normalization-based attention module) is an efficient attention mechanism that improves the design of spatial attention and channel attention sub-modules and can enhance the effect of deep learning model segmentation. In this study, four deep learning models, UNet, UNet++, DeepLabV3, and AttentionUNet with the addition of the NAM attention mechanism, were used to segment CT images of northeastern black soil and compared with traditional segmentation methods. The results showed that AttentionUNet achieved evaluation values of 88.08%, 0.96, 96.02%, 96.64% in the four indexes of intersection and concurrency ratio, F1 score, precision, and recall, respectively, and the precision was improved by 1.46 and 0.21 percentage points relative to UNet and UNet++, respectively, and compared with the traditional Watershed method, the precision and recall rate were increased by 101.13% and 33.54%. In conclusion, AttentionUNet model proposed in this paper can effectively improve the efficiency and precision of soil CT image segmentation, and provide a more efficient method for pore segmentation of soil CT images.

Key words: UNet; NAM; Soil CT; Image segmentation; Attention mechanism

土壤孔隙结构对于水分和气体在土壤中的运动、植物根系的生长以及土壤生物的活性具有关键影响^[1], 准确获取土壤内部孔隙结构信息, 是深入理解和预测这些过程的基础^[2]。目前, X 射线计算机断层扫描(CT)

技术已成为分析土壤孔隙结构的重要工具^[3], 而图像分割是提取孔隙信息的核心步骤。然而, 土壤内部结构的复杂性带来了极大的分割挑战, 传统图像处理方法难以充分应对多变的土壤纹理与结构特征。

①基金项目: 国家重点研发计划项目(2022YFD1500701)和中央高校基本科研业务费专项资金项目(2024RC015)资助。

* 通信作者(zhouhu@cau.edu.cn)

作者简介: 宋忠林(1997—), 男, 河南商丘人, 硕士研究生, 主要从事土壤物理结构的研究。E-mail: szhonglin@foxmail.com

在土壤 CT 图像分割中,研究者主要采用区域生长法、分水岭法、阈值分割法、模糊 C 均值法等传统方法。这些方法在分割效率和精度上各具优劣:区域生长法^[4]基于相似的邻近像素进行分割,但处理多个分离的孔隙时效率低;分水岭法^[9-13]通过灰度分布来分割,但土壤图像中的细微灰度可能导致孔隙结构的过度分割^[14];阈值分割法^[15-16]依赖设定的阈值,适用于二值或多分类分割,但容易产生过度分割或者欠分割的问题;模糊 C 均值法(Fuzzy C-means, FCM)^[17-19]通过概率计算为图像提供更高的鲁棒性,但其对每个像素点的逐次迭代计算耗时较长。此外一些成熟的图像处理软件也集成了各种分割方法,如大津法(Otsu)、最大熵法(MaxEntropy)和阈值法等^[20],在处理前景和背景差异较大的图像时效果较好,但由于 CT 机器的部分容积效应,孔隙边界往往模糊^[21],这种模糊性使得这些软件在应对孔隙分割时可能导致过度分割或不足分割。在复杂土壤结构的 CT 图像中,需要考虑设计一种高精度、适应性强的分割方法,从而提升土壤孔隙结构信息的提取准确性。

近年来,深度学习分割方法通过自动学习大量图像特征,展示出更强的鲁棒性和泛化能力,为复杂图像分割问题提供了新的解决方案。尤其是 UNet 深度学习模型的引入,因其编码器-解码器结构和跳跃连接的设计,有助于在捕获图像上下文信息的同时保留细节信息,从而实现精确的像素级分割。Shelhamer 等^[22]在 2015 年提出的全卷积网络结构(Fully convolutional network, FCN)奠定了图像语义分割的基础,实现了像素级别的分类。而近年来,UNet 在土壤 CT 图像分割中的应用已取得了一定的成效。例如,韩巧玲^[23]在 2020 年使用 UNet 模型和简化的卷积神经网络(Convolutional neural networks, CNN),成功对土壤 CT 图像实现了较高质量的分割。然而,由于土壤 CT 图像的复杂性和多样性,进一步提高分割精度和效率仍然具有挑战性。为解决这一问题,本文在神经网络 UNet 模型的基础上,通过融合注意力机制的方式提高模型对有效信息的获取能力,从而提升孔隙结构的分割精度。通过构建适用于图像分割的土壤 CT 图像数据集,将该改进模型与传统分割方法(Watershed)以及多种深度学习分割模型(UNet++、DeepLabV3)进行效果对比,以全面评估该方法在不同形态孔隙上的分割性能和适用性。

1 神经网络结构与算法原理

神经网络 UNet 在土壤 CT 图像分割中的应用主

要体现在以下几个方面:首先,通过训练大量的土壤 CT 图像数据,UNet 模型能够学习到土壤内部结构的特征表示,从而实现对土壤图像的自动分割;其次,利用 UNet 模型的高分辨率输出特性,研究人员可以获得到更为精细的土壤结构信息,这对于分析土壤孔隙结构、评估土壤质量等具有重要意义;最后,神经网络 UNet 还具有良好的泛化性能,能够适应不同类型、不同质地的土壤图像分割需求,为土壤科学研究的广泛应用提供了便利。本研究采用 UNet、UNet++、DeepLabV3 以及加入了注意力机制的 AttentionUNet 共 4 种深度学习模型来开展土壤 CT 图像的分割。

1.1 UNet 网络结构

UNet 是 Ronneberger 等^[24]于 2015 年提出的一种对称的 U 型结构网络,左边部分编码器由多个卷积层和池化层组成,用于将输入图像降采样并提取特征;右侧的解码器由多个上采样层和卷积层构成,作用是将编码器输出的特征映射还原到与原始输入图像相同的分辨率,并逐步细化分割结果。为缓解信息丢失问题,编码器与解码器之间通过跳跃连接直接传递高分辨率特征,从而增强信息流动,提高分割精度和细节还原能力。这种结构使 UNet 在复杂背景下的小目标分割任务中表现出色。

1.2 UNet++模型

UNet++模型是 Zhou 等^[25]在 2018 年基于 UNet 改进提出的模型。在 UNet 的基础上,UNet++模型引入了更多的上采样节点和跳跃连接。该网络中的每个嵌套卷积块通过几个卷积层提取语义信息,通过共享一个编码器将不同深度的 UNet 有效集成,嵌套卷积块中的所有卷积层都通过密集的跳跃连接彼此相连,使得每一层都能够融合来自不同深度的语义信息,这些改进使网络性能得到明显提升。

1.3 DeepLabV3 模型

DeepLabV3 模型^[26]由谷歌的研究人员开发,通过增强模型的几个关键方面,特别是空洞卷积和空间金字塔池化,相比之前的版本(DeepLabV1 和 DeepLabV2)有了显著改进。空洞卷积(Atrous convolution)允许网络通过调整卷积核的视野大小来捕捉多尺度上下文信息,而无需增加参数数量或计算量。空洞空间金字塔池化(ASPP)使用多种不同膨胀率的空间卷积来有效地捕捉和整合多尺度特征,是一种改进的空间金字塔池化技术。DeepLabV3 可以基于先进的主干网络,如 ResNet、Xception 或 MobileNet 构建,这些网络有助于从图像中提取丰富的特征表示。通过全局平均池化纳入图像级特征,有助于捕捉全局上下文,对于解决

分割任务中的歧义至关重要，同时 DeepLabV3 将空洞卷积无缝集成到 ResNet 架构中，使其能够更有效地处理不同尺度的图像。

1.4 加入注意力机制的 AttentionUNet 模型

注意力机制^[27]包含通道注意力^[28]和空间注意力^[29] 2 个模块(图 1)，其核心在于引导网络模型聚焦于关键部分，是实现网络自适应注意的一种方法，从而提高对重要信息的捕捉能力。Chen 等^[30]提出的 TransUNet 将自注意力机制用于 U 型网络的底端以关

注图像的全局特征，在解码阶段通过 CNN 为全局特征补充局部空间特征增强细节信息，提高了模型获取信息的能力。NAM(Normalization-based attention module)^[31]作为一种高效、轻量级的注意力机制，采用 CBAM(Convolutional block attention module)的模块整合，对通道和空间注意力子模块进行了重新设计。将 NAM 模块集成到每个网络块的末端，对于残差网络，则嵌入在残差结构的末尾，以进一步提升特征提取的效率和效果。

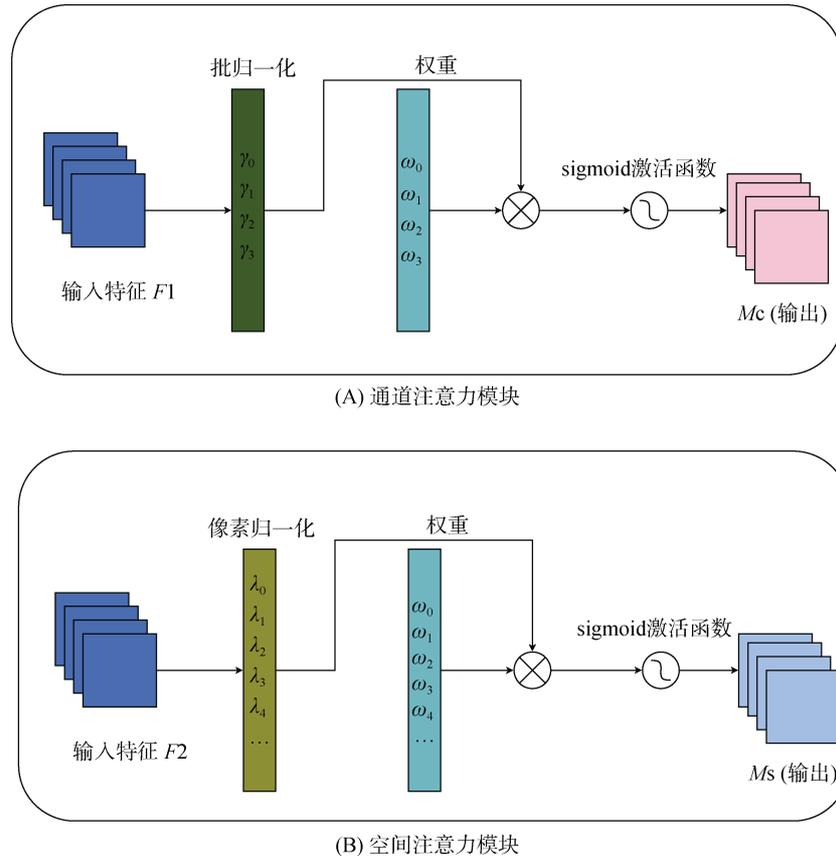


图 1 通道注意力(A)和空间注意力(B)示意图
Fig. 1 Channel (A) and spatial (B) attention

如图 1A 所示，通道注意力子模块输出的特征表示为 M_c 。其中，比例因子 γ 用于调整每个通道的权重，权值计算方法如公式(1)所示， γ_i 是第 i 个通道的缩放因子。

$$\omega_i = \frac{\gamma_i}{\sum_{j=0}^n \gamma_j} \quad (1)$$

在空间维度上，使用批归一化(Batch normalization, BN)的比例因子来衡量像素的重要性，这个过程被称为像素归一化。而对应的空间注意力子模块如图 1B 所示，输出被标记为 M_s ，其中的比例因子是 λ ，权值和损失值的计算方法如公式(2)和公式(3)所示：

$$\omega_i = \frac{\lambda_i}{\sum_{j=0}^n \lambda_j} \quad (2)$$

$$\text{Loss} = \sum_{(x,y)} l(f(x,W)) + p \sum g(\gamma) + p \sum g(\lambda) \quad (3)$$

式中： λ_i 是第 i 个像素位置的缩放因子； l 表示损失函数(Loss function)； g 表示正则化项的函数(Regularization function)； p 表示正则化项的权重； x 表示输入数据(Input data)； y 表示目标标签(Target label)； $f(x,W)$ 是模型的输出，表示模型根据输入数据 x 和网络权重 W 计算的预测结果。

本研究在 UNet 的上采样路径中，将注意力机制

模块 (NAM)分别加入到特征融合部分和跳跃连接部分(图 2), 用于增强解码器输出和编码器相应层输出之间的特征融合, 同时减少冗余的跳跃链接(Skip connection)。解码器的当前层输出和对应的编码器

层输出作为输入传递给注意力模块, 通过注意力模块调整后的编码层输出将允许网络专注于更重要的特征, 从而提高特征融合的质量, 改善多尺度特征融合。

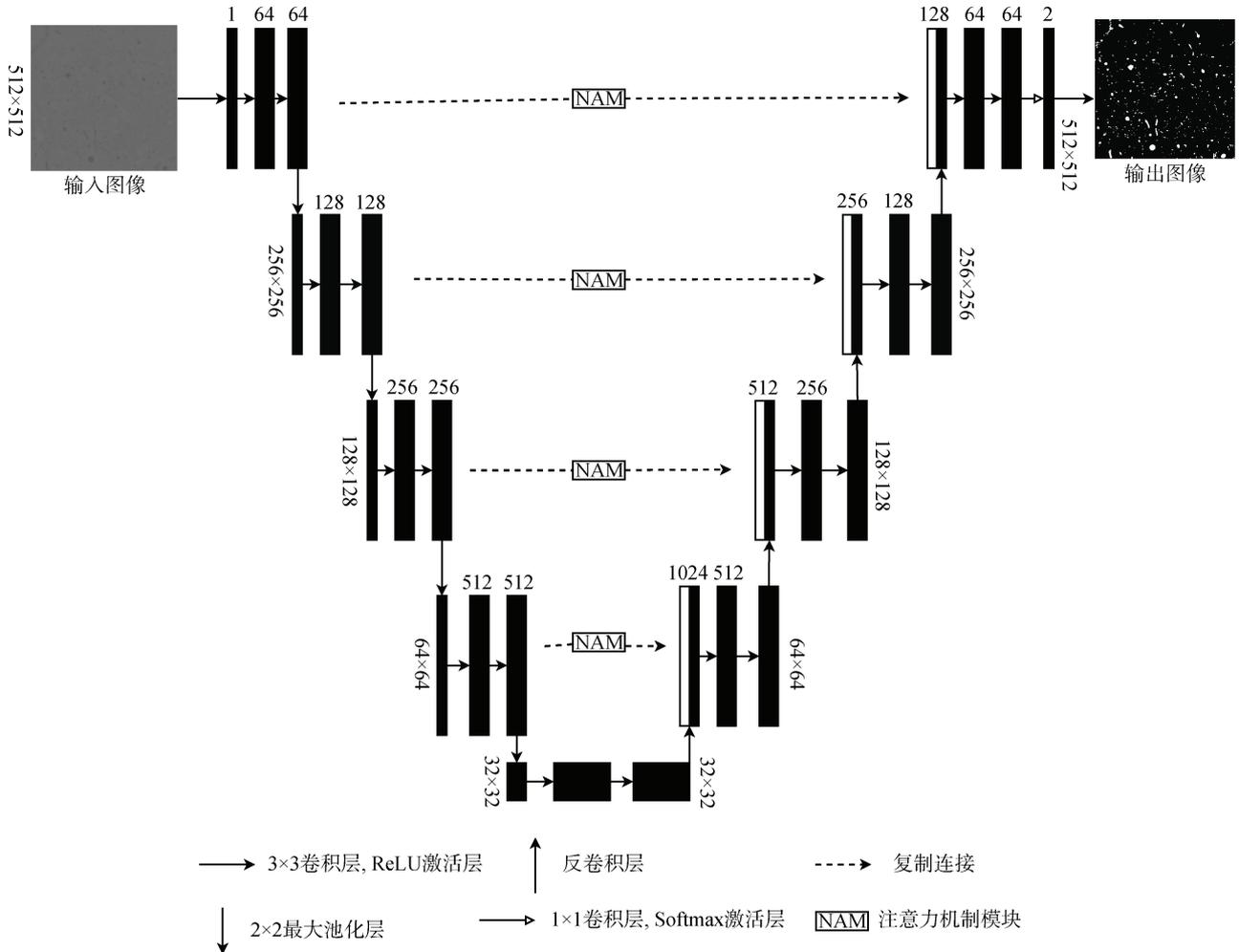


图 2 改进的 UNet 结构图
Fig. 2 Improved UNet structure diagram

在基本的 UNet 结构中, 前半部分是编码部分, 由两个 3x3 卷积层(ReLU)加上一个 2x2 Maxpooling 层组成下采样模块, 需经过 4 次操作; 后半部分是特征融合网络, 由一层反卷积+特征拼接 concat+两个 3x3 的卷积层(ReLU)反复构成, 同样经过 4 次操作, 与特征提取网络刚好相对应, 最后接一层 1x1 卷积, 降维处理, 即将通道数降低至特定的数量, 得到目标图。

UNet 架构中的跳跃连接设计旨在利用编码器中的高分辨率特征图来增强解码器中的特征表达。在跳跃连接中引入注意力机制, 可以让解码器优先关注对任务更重要的特征, 从而提升特征融合的有效性。这种机制不仅带来了更精细的上下文信息, 还能在处理细小或边缘模糊的目标时显著改善分割效果, 同时提

高网络的效率和泛化能力。

本研究的孔隙识别网络由训练模型和测试模型组成(图 3)。在训练阶段, 网络首先提取输入图像的特征, 并将其与人工标注的孔隙二值图像进行对比, 计算预测结果与标注图像之间的差异。根据这一差值, 通过反向传播优化网络参数, 不断调整模型以提升分割性能。训练完成后, 利用已优化参数的网络对测试图像进行分割, 最终生成孔隙的二值化图像。

2 材料与方法

2.1 软硬件环境

本研究在 GPU 环境中运行网络模型, 其具体配置如表 1 所示。开发工具选用集成开发环境 PyCharm,

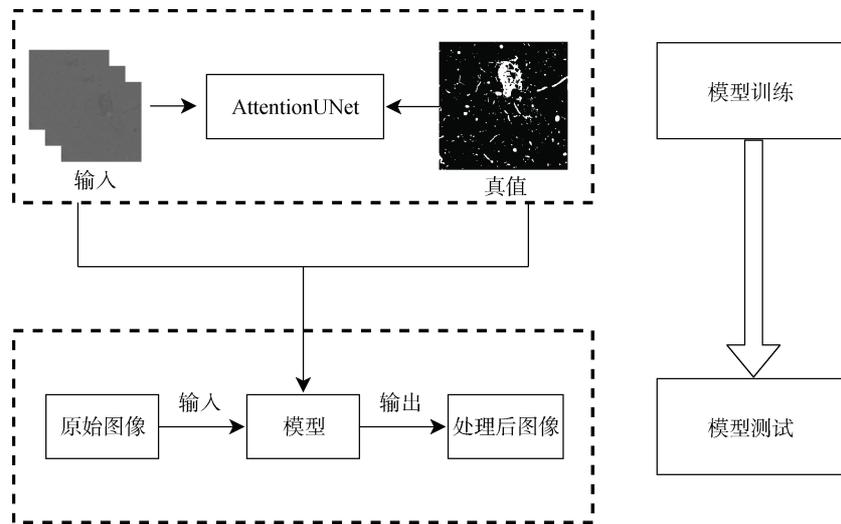


图 3 孔隙识别网络示意图

Fig. 3 Schematic diagram of pore identification network

表 1 平台硬件与软件配置

Table 1 Platform hardware and software configuration

平台硬件与软件配置	
CPU	Core i5-12490F
GPU	NVIDIA RTX4060
显存	8 GB
内存	1 TB
框架	Pytorch 2.1.2
CUDA	CUDA 12.1
语言	Python 3.9
操作系统	Windows 10 Professional

为提升训练效果和训练速度,使用 `resize_image` 函数调用图像插值算法将输入的图像尺寸调整为 512×512 像素,训练的迭代次数设为 200,批大小(batchsize)设为 4,训练过程中初始学习率(Init_lr)为 $1e-4$,最小学习率(Min_lr)为 $Init_lr \times 0.01$,优化器选择 Adam 以动态校正各网络参数的拟合速率,学习率衰减方式为余弦衰减(`lr_decay_type = 'cos'`),模型权值每 5 个训练轮次(epoch)保存一次,以确保模型训练的安全性和

可控性。

2.2 数据集

土壤样本采自中国东北黑土区中国农业大学梨树实验站,使用柱状圆筒采样器(内置有机玻璃管,内径 85 mm,长度 100 cm,厚度 2.5 mm)采集 7 个土柱,用保鲜膜包裹并用加宽胶带将两端封口。将每个土柱切割为大约 33 cm 长的小段,使用工业 CT(VI Tomo I x S240, Waygate Technologies, Germany)进行扫描,电压为 210 kV,电流为 $520 \mu A$,分辨率为 $83 \mu m$ 。使用 `Datos|x2 rec` 软件进行图像重建,然后利用 `VG Studio` 软件导出 16 位灰度图像,储存为 `tiff` 格式,图像大小为 $2\ 200 \times 2\ 200$ 像素。

本研究选取所采集图像中的 200 张 CT 图像作为数据集,对数据集以 8 : 2 的比例分为包含 160 张 CT 图像的训练集和包含 40 张 CT 图像的测试集。为避免图像中的损伤区域,在选取图片时裁取图像中央区域 600×600 像素(图 4),然后使用数据标注软件 `LabelMe`^[32]对裁剪图像进行手工标注,为神经网络模型提供准确的训练目标,确保模型的分割精度。

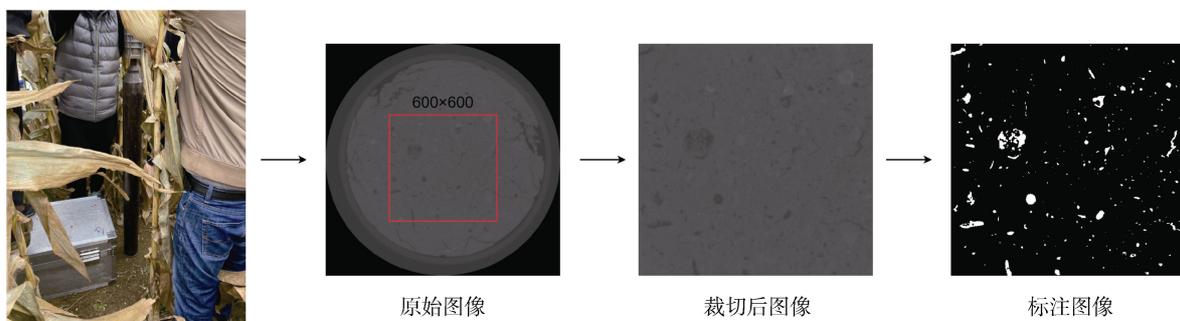


图 4 土壤 CT 图像采集与处理过程

Fig. 4 Soil CT image acquisition and processing process

2.3 模型评价指标

采用 4 种分割精度指标评估模型的性能,包括平均交并比(Mean intersection over union, MIoU)、精确率(Precision)、平均召回率(Recall)、F1 分数和百分比准确率(Percentage accuracy), 这些指标的值都在 0 ~ 1, 能够从不同的角度反映模型在土壤 CT 图像分割任务上的表现。

平均交并比:就是该数据集中每一个类的交并比的平均值,在语义分割的问题中,交并比就是该类的真实标签和预测值的交和并的比值。

$$MIoU = \frac{1}{2} \left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FN + FP} \right) \quad (4)$$

式中(下同):第一位 T/F,表示预测的对错(True/False);第二位 P/N,表示预测的结果(Positive/Negative); TP 表示实际为正样本且被正确预测为正样本的样本数, FP 表示实际为负样本但被错误预测为正样本的样本数, TN 表示实际为负样本且被正确预测为负样本的样本数, FN 表示实际为正样本但被错误预测为负样本的样本数。

精确率:表示被正确预测为正样本的像素数量在所有被预测是正样本的结果中所占的比例。

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

召回率:表示在真实值是正样本的所有结果中,模型预测对的比例。

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

F1 分数是一个综合评估模型性能的指标,是精确率和召回率的调和平均数,它提供了一种平衡考虑精确率和召回率的方法,F1 分数的取值范围在 0 ~ 1, 值越高表示模型的性能越好。当且仅当精确率和召回率都很高时, F1 分数才会达到较高的值。

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

准确度:是指模型正确分类的样本数占总样本数的比例。

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (8)$$

3 结果与分析

3.1 不同模型训练过程的损失值

如图 5 所示, 4 种网络模型在训练集上经过 200 次迭代的 Loss 变化曲线显示, 训练集的损失值随着训练轮次的增加逐渐下降, 表明了网络的可行性。本文提出的 AttentionUNet 通过在原始 UNet 基础上引入注意力机制, 不仅改进了网络结构, 还提升了模型的性能和收敛速度。具体而言, UNet 和 UNet++模型的损失值在训练 100 轮后收敛, 并最终稳定在 0.1 左右; 而改进的 AttentionUNet 模型在 60 轮次后损失值便趋于收敛, 最终稳定在 0.05 左右, 显著优于其他模型。

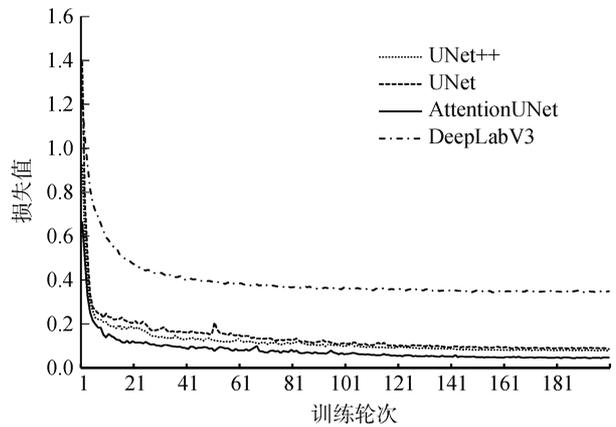


图 5 不同模型的训练 Loss 曲线图
Fig. 5 Training loss curves under different models

3.2 不同神经网络模型识别结果精度

在完成分割模型训练后, 使用不同算法模型在 40 组土壤 CT 图像测试集上进行验证, 计算孔隙识别的交并比、精确度、F1 分数和召回率及各项参数的方差, 结果见表 2。

加入注意力机制的 AttentionUNet 分割方法精确度达到了 96.02%, 说明准确识别孔隙结构的能力较高^[33]; AttentionUNet 模型的精确度指标相对于 UNet 模型(94.56%)提升了 1.46 个百分点, 相较于 UNet++ 模型(95.81%)提升 0.21 个百分点, 而 DeepLabV3 方

表 2 不同神经网络模型识别精度
Table 2 Recognition accuracies under different models

模型	精确度 (%)	精确度方差 (% ²)	召回率 (%)	召回率方差 (% ²)	交并比 (%)	交并比方差 (% ²)	F1 分数	F1 分数方差 (% ²)
UNet	94.56	0.77	92.59	0.83	88.46	0.46	0.94	0.06
UNet++	95.81	0.64	93.96	0.89	88.18	0.42	0.95	0.06
AttentionUNet	96.02	0.46	96.64	0.54	88.08	0.14	0.96	0.04
DeepLabV3	79.83	1.21	73.29	0.97	66.69	0.58	0.76	0.12

法的精确度为 79.83%，说明该方法对孔隙结构存在较大程度的过度识别。

AttentionUNet 的召回率为 96.64%，该数值越高表明孔隙结构被错误识别为固体颗粒的像素数越少。召回率指标相较于 UNet (92.59%) 和 UNet++ (93.96%) 分别提升 4.05 和 2.68 个百分点，说明 AttentionUNet 方法在土壤大孔隙和固相物质的分割方面能力是最强的。

AttentionUNet 的交并比为 88.08%，这衡量了模型预测的分割结果与真实标签之间的重叠程度。交并比不仅考虑了正确分类的像素数，还考虑了错误分类的像素数，从而提供了一个更全面的模型性能评估。在图像分割任务中，交并比越高，表示模型的分割结果与真实标签越一致，模型性能越好。UNet 的交并比最高，达到 88.46%，表明其分割区域与实际区域的吻合程度最好；UNet++ 和 AttentionUNet 的交并比相近，分别为 88.18% 和 88.08%，略低于 UNet，但差距极小。而 DeepLabV3 的交并比远低于其他模型，仅为 66.69%，表明其分割结果与实际情况有较大偏差。

AttentionUNet 的 F1 分数最高，为 0.96，反映了其在精确度和召回率上的出色表现。F1 分数是一个综合指标，平衡了精确度和召回率，是在样本类别不均衡情况下评估模型性能的一个很好的指标，高的

F1 分数意味着模型在精确度和召回率上都表现良好。UNet++ 的 F1 分数为 0.95，表现也非常接近 AttentionUNet；UNet 的 F1 分数为 0.94，虽然略低于 UNet++ 和 AttentionUNet，但依然出色。而 DeepLabV3 的 F1 分数最低，为 0.76，进一步体现了其在综合性能上的不足。

总的来说，AttentionUNet 在所有指标中都表现优异，特别是在精确度、召回率和 F1 分数上占据优势，各项指标的方差也相对较小，说明该模型在处理各种情况时都能较为稳定地提供高质量的结果。UNet++ 的整体表现接近 AttentionUNet，虽然在精确度和召回率上稍有不及，但仍然是一个非常强大的模型。UNet 的表现略逊色于前两个模型，但在交并比上稍有优势。DeepLabV3 明显落后于其他模型，精确度、召回率、交并比和 F1 分数都表现不佳，表明不太适合当前任务和数据集。

3.3 分割结果对比

图 6 展示 3 幅不同的黑土 CT 图像，将 4 种模型的分割结果与人工标注的标准图像进行对比。原始土壤 CT 图像中，孔隙以黑色为主，部分孔隙与固相物质的边界较为模糊。从 3 组分割实例可以看出，无论是基础的 UNet 模型还是改进后的 UNet 模型，都在土壤 CT 图像分割中展现出优异的性能，分割出的土壤孔隙能够完整地包含原始图像中的孔隙，孔隙轮廓

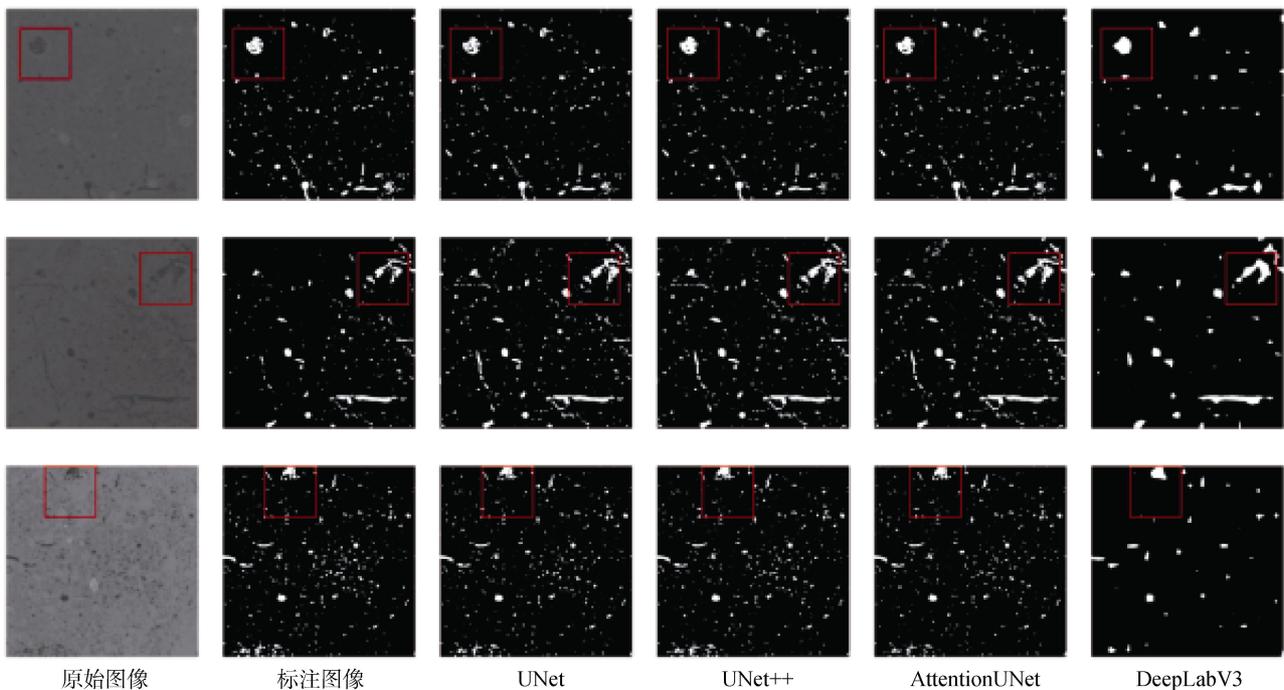


图 6 不同模型分割结果对比

Fig. 6 Comparison of segmented results under different models

光滑自然,与标准标注图像基本吻合。但是 UNet 方法在识别一些不规则的孔隙时如图 6 中红色方框标注部分,还是存在欠分割的问题,这可能是由于网络结构的欠拟合,对多尺度的特征融合缺少考虑;UNet++模型在面对这种情况时则会存在过度分割的问题,高估了孔隙结构;而 DeepLabV3 模型则存在明显的欠分割,对一些直径较小的孔隙识别能力较差。总的来说,AttentionUNet 分割精度和分割效果优于其他的神经网络方法。

为了评估 AttentionUNet 方法及 3 种传统的图像

处理方法(分水岭(Watershed)^[34]、模糊 C 均值法(Fuzzy C-means, FCM)^[35]以及大津法(Otsu)^[36])的分割性能,将其分割结果进行对比(表 3)。使用 Python 代码和 OpenCV 库来实现以上 3 种传统分割方法,对 CT 图像进行适当的预处理(如噪声去除和初步阈值分割),通过标记前景和背景,算法基于图像的梯度信息,将图像分割成多个区域,利用分割正确率、精确度和召回率 3 个指标定量评价孔隙分割结果,并根据多组图像测试得到的结果计算方差。对于 Watershed 和 FCM 方法,根据土壤孔隙、裂缝的结构特点,种子点数设为 4。

表 3 AttentionUNet 和传统分割方法效果对比

Table 3 Comparison of effectiveness of AttentionUNet and traditional segmentation methods

分割方法	正确率(%)	正确率方差(% ²)	精确度(%)	精确度方差(% ²)	召回率(%)	召回率方差(% ²)
Watershed	94.34	1.25	47.74	2.17	72.37	1.87
FCM	51.25	2.39	49.89	2.62	51.93	2.12
Otsu	85.01	1.92	46.17	2.74	90.36	1.98
AttentionUNet	97.16	0.52	96.02	0.19	96.64	0.36

从表 3 可以看出,AttentionUNet 的正确率最高(97.16%),表明其在整体分类上最为准确;Watershed 和 Otsu 的正确率也较高,但低于 AttentionUNet。Watershed 正确率高达 94.34%,说明该方法在大部分情况下能正确分类;Otsu 正确率 85.01%,说明该方法能较好地分类前景和背景;FCM 的正确率明显较低,表明其分类能力较弱。AttentionUNet 的精确度(96.02%)显著高于其他方法,表明其在细节和边界的处理上非常出色;FCM、Watershed 和 Otsu 的精确度较低,表明它们在处理复杂边界时的能力有限。AttentionUNet 的召回率最高(96.64%),说明其能够检测出几乎所有的目标区域,且误检率低;Otsu 的召回率为 90.36%,说明能找到大部分目标区域,但可能包含一些误检;Watershed 的召回率为 72.37%,说明能找到

一部分的目标区域,但仍有漏检;FCM 的召回率最低,进一步验证了其在目标区域检测上的不足。综合来看,AttentionUNet 的精确度和召回率相比于 Watershed 提高了 101.13% 和 33.54%,在所有性能指标上表现最优,其次是 Watershed 和 Otsu,FCM 的表现相对较差。

不同方法在大孔隙的分割中效果不同(图 7)。Watershed 在大孔隙的分割方面表现良好,但是存在一些小孔隙的欠分割现象,表明其泛化能力较差;而 FCM 和 Otsu 在分割大孔隙和小孔隙都存在过度分割的问题,这一现象在 Otsu 方法中尤为严重,一些分散的孔隙被错误地连接起来;而 AttentionUNet 方法未出现上述问题,分割的结果都更加接近人工标注的图像,这表明在多尺度孔隙共存的情况下,AttentionUNet 在分割时与传统方法相比有很大的优势。

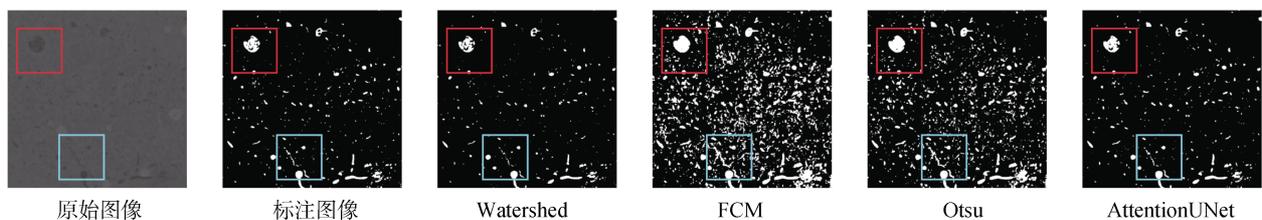


图 7 不同分割方法效果对比

Fig. 7 Comparison of effects of different segmentation methods

通过 ImageJ 软件生成基于不同分割方法的三维孔隙图像(图 8),对分割结果进行直观对比和分析。对于传统的 Watershed 方法,分割结果存在明显的欠分割现象,特别是在小孔隙的分割上存在较大不足,

部分小孔隙被忽略或连接在一起,无法准确反映土壤内部孔隙的真实分布和形态。相比之下,UNet 和 UNet++方法在分割较小孔隙时表现出一定的改进,但二者均存在不同程度的过度分割问题。这种过度分

割导致孔隙结构的边界模糊,尤其是小孔隙区域的分割结果显得杂乱无序,使得孔隙的连续性和脉络特征受到影响。改进的 AttentionUNet 方法显著优化了分割结果,其分割的三维孔隙图像表现出更高的准确性和连续性,模型能够更好地捕获孔隙的细节特征,避免了过度分割和欠分割现象的发生。具体而言,AttentionUNet 分割后的孔隙结构更为清晰,形态脉络更具逻辑性,能够准确还原土壤孔隙的真实三维结构,为后续土壤孔隙特征的量化分析奠定了坚实的基础。

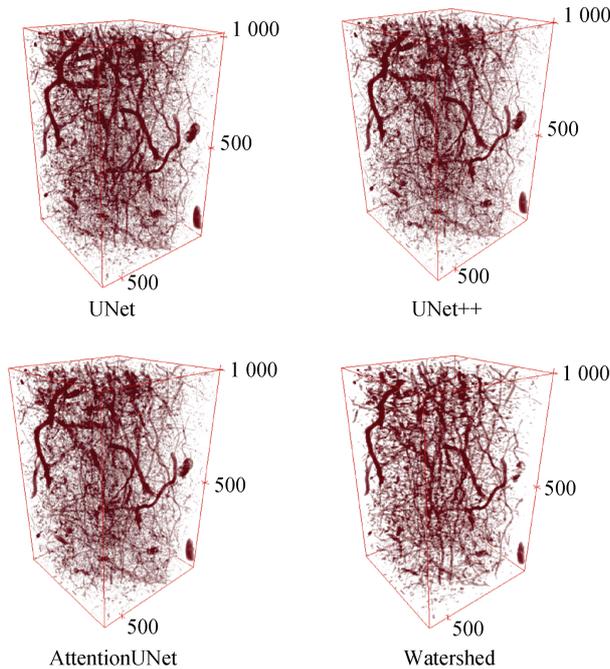


图 8 Watershed 和神经网络方法分割孔隙的三维结构
Fig. 8 3D structures of soil pores segmented by Watershed and neural network methods

4 结论与讨论

本研究提出了一种改进的 UNet 神经网络模型,旨在针对土壤 CT 图像分割任务的特点进行优化。在深入分析 UNet 网络结构和性能潜能的基础上,结合 NAM 注意力机制,对原始 UNet 模型进行了调整与优化。实验表明,深度学习在土壤 CT 图像分割中具有显著优势,尤其在复杂孔隙结构的处理上,明显优于传统方法。UNet 等神经网络通过结合真实标签与先验知识,能够自主学习并全面表达土壤孔隙特征,显著提升大孔隙结构的精细分割能力。改进后的模型(AttentionUNet)因引入注意力机制,能够模拟人类注意力,将关注聚焦于图像中关键的孔隙信息,能够更有效地捕获图像中的关键特征,提升在处理复杂土壤 CT 图像时的分割性能和稳定性。通过在实验数据集上的验证,改进的 AttentionUNet 在土壤 CT 图像分

割任务中表现出色,能够准确识别并分割出土壤孔隙,与传统 CNN 相比表现更优。与 UNet、UNet++ 以及 DeepLabV3 模型的对比实验显示,AttentionUNet 在精确率、召回率、平均交并比和 F1 分数等指标上均有提升;与传统的方法(如 Watershed)相比,改进模型在精确率和召回率方面实现了显著提升;为土壤 CT 图像分割提供了一种高效、可靠的深度学习解决方案。

尽管深度学习模型展现了较高的性能,但仍面临数据和计算资源方面的挑战,特别是在土壤样本多样性和特征异质性较高的情况下。本研究在分割精度上尚有提升空间,未来可从以下几个方向进一步优化。

1) 提升数据多样性与模型泛化能力。土壤类型、颗粒形态和孔隙率等差异显著影响图像特征分布,因此构建更广泛且多样化的数据集将有助于提高模型在不同土壤类型上的适应性和分割精度。同时,更丰富的数据可降低模型对未见样本的误差,提升其在实际应用中的鲁棒性。

2) 探索半监督与自监督学习。标注土壤 CT 图像耗时且成本高,利用未标注数据的半监督或自监督学习方法能够挖掘潜在特征,在标注数据有限的情况下提高模型性能。这类方法通过预训练或自学习挖掘数据中的潜在特征,提升模型对细微结构的识别能力,将成为未来土壤 CT 图像分割的重要发展方向。

3) 引入多模态数据融合。结合 CT 图像与其他数据类型(如光谱图像、土壤化学与物理参数)可为模型提供更全面的特征信息,帮助模型理解土壤的多层次特性,从而进一步提升分割效果。

参考文献:

- [1] 张靖, 陈琳, 周虎, 等. 基于数字图像技术的土壤孔隙结构定量研究进展[J]. 土壤, 2023, 55(1): 21–29.
- [2] Wildenschild D, Vaz C M P, Rivers M L, et al. Using X-ray computed tomography in hydrology: Systems, resolutions, and limitations[J]. Journal of Hydrology, 2002, 267(3/4): 285–297.
- [3] 周虎, 李文昭, 张中彬, 等. 利用 X 射线 CT 研究多尺度土壤结构[J]. 土壤学报, 2013, 50(6): 1226–1230.
- [4] Shih F Y, Cheng S X. Automatic seeded region growing for color image segmentation[J]. Image and Vision Computing, 2005, 23(10): 877–886.
- [5] 黄展鹏, 蒋世忠, 鲍苏苏, 等. CT 系列图像的三维层次化子块生长分割[J]. 计算机工程与应用, 2011, 47(11): 182–184.
- [6] Higo Y, Oka F, Morishita R, et al. Trinarization of μ X-ray CT images of partially saturated sand at different water-retention states using a region growing method[J]. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 2014, 324: 63–69.

- [7] 韩巧玲, 赵玥, 赵燕东, 等. 基于全卷积网络的土壤断层扫描图像中孔隙分割[J]. 农业工程学报, 2019, 35(2): 128–133.
- [8] Hashemi M A, Khaddour G, François B, et al. A tomographic imagery segmentation methodology for three-phase geomaterials based on simultaneous region growing[J]. *Acta Geotechnica*, 2014, 9(5): 831–846.
- [9] Fei Y H, She D L, Gao L, et al. Micro-CT assessment on the soil structure and hydraulic characteristics of saline/sodic soils subjected to short-term amendment[J]. *Soil and Tillage Research*, 2019, 193: 59–70.
- [10] Berg S, Saxena N, Shaik M, et al. Generation of ground truth images to validate micro-CT image-processing pipelines[J]. *The Leading Edge*, 2018, 37(6): 412–420.
- [11] Houston A N, Otten W, Falconer R, et al. Quantification of the pore size distribution of soils: Assessment of existing software using tomographic and synthetic 3D images[J]. *Geoderma*, 2017, 299: 73–82.
- [12] Avanthi Isaka B L, Ranjith P G, Rathnaweera T D, et al. Quantification of thermally-induced microcracks in granite using X-ray CT imaging and analysis[J]. *Geothermics*, 2019, 81: 152–167.
- [13] 刘勇, 胡霞, 李宗超, 等. 基于 CT 的青海湖流域芨芨草草地土壤大孔隙特征分析[J]. 土壤, 2017, 49(1): 184–188.
- [14] 赵玥, 刘雷, 韩巧玲, 等. 基于 CT 图像的土壤孔隙结构重构[J]. 农业机械学报, 2018, 49(S1): 401–406.
- [15] Kim H, Anderson S H, Motavalli P P, et al. Compaction effects on soil macropore geometry and related parameters for an arable field[J]. *Geoderma*, 2010, 160(2): 244–251.
- [16] Hapca S M, Houston A N, Otten W, et al. New local thresholding method for soil images by minimizing grayscale intra-class variance[J]. *Vadose Zone Journal*, 2013, 12(3): vzj2012.0172.
- [17] Verma H, Agrawal R K, Sharan A. An improved intuitionistic fuzzy c-means clustering algorithm incorporating local information for brain image segmentation[J]. *Applied Soft Computing*, 2016, 46: 543–557.
- [18] Ojeda-Magaña B, Quintanilla Domínguez J, Ruelas R, et al. Pore detection in 3-D CT soil samples through an improved sub-segmentation method[J]. *European Journal of Soil Science*, 2019, 70(1): 66–82.
- [19] Sun X L, Zhao Y G, Wang H L, et al. Sensitivity of digital soil maps based on FCM to the fuzzy exponent and the number of clusters[J]. *Geoderma*, 2012, 171: 24–34.
- [20] Manahiloh K N, Meehan C L. Determining the soil water characteristic curve and interfacial contact angle from microstructural analysis of X-ray CT images[J]. *Journal of Geotechnical and Geoenvironmental Engineering*, 2017, 143(8): 04017034.
- [21] Garbout A, Sturrock C, Armenise E, et al. TopCap: An ImageJ plugin to automatically determine and quantify complex surface topologies and associated sub-surface structures in X-ray Computed Tomography images[J]. *Vadose Zone Journal*, 2018, 17(1): 1–11.
- [22] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[C]//IEEE transactions on pattern analysis and machine intelligence. IEEE, 2015: 640–651.
- [23] 韩巧玲. 基于 CT 图像的黑土大孔隙精细分割与重构方法研究[D]. 北京: 北京林业大学, 2020.
- [24] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation[M]//Medical image computing and computer-assisted intervention – MICCAI 2015. Cham: Springer International Publishing, 2015: 234–241.
- [25] Zhou Z W, Rahman Siddiquee M M, Tajbakhsh N, et al. UNet++: A nested U-Net architecture for medical image segmentation[M]//Deep learning in medical image analysis and multimodal learning for clinical decision support. Cham: Springer International Publishing, 2018: 3–11.
- [26] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834–848.
- [27] Zagoruyko S, Komodakis N. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer[EB/OL]. 2016: 1612.03928. <https://arxiv.org/abs/1612.03928v3>.
- [28] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF conference on computer vision and pattern recognition. Salt Lake City, UT, USA: IEEE, 2018: 7132–7141.
- [29] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30.
- [30] Chen J N, Lu Y Y, Yu Q H, et al. TransUNet: Transformers make strong encoders for medical image segmentation [EB/OL]. 2021: 2102.04306. <https://arxiv.org/abs/2102.04306v1>.
- [31] Liu Y C, Shao Z R, Teng Y Y, et al. NAM: Normalization-based attention module[EB/OL]. 2021: 2111.12419. <https://arxiv.org/abs/2111.12419v1>.
- [32] Russell B C, Torralba A, Murphy K P, et al. LabelMe: A database and web-based tool for image annotation[J]. *International Journal of Computer Vision*, 2008, 77(1): 157–173.
- [33] Shirzadi A, Shahabi H, Chapi K, et al. A comparative study between popular statistical and machine learning methods for simulating volume of landslides[J]. *Catena*, 2017, 157: 213–226.
- [34] Bleau A, Leon L J. Watershed-based segmentation and region merging[J]. *Computer Vision and Image Understanding*, 2000, 77(3): 317–370.
- [35] Bezdek J C, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm[J]. *Computers & Geosciences*, 1984, 10(2/3): 191–203.
- [36] Xu X Y, Xu S Z, Jin L H, et al. Characteristic analysis of otsu threshold and its applications[J]. *Pattern Recognition Letters*, 2011, 32(7): 956–961.