引文格式：张永年. 基于无人机光学影像和改进 U-Net 模型的河流轮廓提取[J]. 地理空间信息, 2024, 22(12): 30-33. doi:10. 3969/j.issn. 1672-4623.2024.12.007

基于无人机光学影像和改进U-Net模型的河流轮廓提取

张永年[1]
（1. 福建省卫星数据开发有限公司，福建 福州 350002）

摘　要：针对复杂背景下河流主干及细小支流轮廓无法被精准地识别和提取的问题，利用高分辨率无人机光学影像构建了一种新的语义分割模型UNetMformer。通过对比UNetMformer模型与经典UNet、DeepLabV3+模型在多种实际场景下对河流轮廓提取结果，以验证UNetMformer模型的优势。结果表明：UNetMformer模型凭借其独特的全局和局部注意力机制，在河流轮廓提取方面展现出了显著的优势，其对不同尺度的河流轮廓提取最为完整，错分漏分现象最少，边缘地区提取更加精细，在各个精度评价指标上也取得了最优值，在多种场景下的平均F1-score为96.06%、IoU为92.49%、Kappa系数为95.74%，充分证明了UNetMformer模型在应对不同尺度河流提取时的广泛适用性和高效性。该方法不仅为河流提取提供了新的视角，也为相关领域的研究和应用提供了有力的技术支撑和参考。

关键词：深度学习；全局-局部注意力机制；无人机光学影像；河流轮廓提取
中图分类号：P231　　**文献标志码：**B　　**文章编号：**1672-4623（2024）12-0030-04

River Contour Extraction Based on UAV Optical Image and Improved U-Net Model

ZHANG Yongnian[1]
(1. Fujian Satellite Data Development Co., Ltd., Fuzhou 350002, China)

Abstract: The contours of river trunks and small tributaries cannot be accurately recognized and extracted under complex backgrounds. Aiming to this problem, we constructed a new semantic segmentation model UNetMformer based on high-resolution UAV optical images. By comparing the river contour extraction results of UNetMformer model with those of classic U-Net and DeepLabV3+ in various actual scenarios, we verified the advantages of UNetMformer model. The results show that the UNetMformer model, by virtue of its unique global and local attention mechanism, has significant advantages in river contour extraction. UNetMformer exhibits the most complete river contour extraction at different scales, fewer misclassifications and omissions, and more delicate extraction of the edge areas. The model achieves the optimal values in different evaluation indexes, with an average F1-score of 96.06%, the IoU of 92.49%, and the Kappa coefficient of 95.74% in various scenarios, which proves the wide applicability and high efficiency of UNetMformer model in coping of river extraction. This method not only provides a new perspective for river extraction, but also provides a strong technical support and reference for research and application in related fields.

Key words: deep learning, unique and local attention mechanism, UAV optical image, river contour extraction

随着21世纪计算机算力的显著提升和遥感数据的爆炸式增长，基于机器学习的方法如支持向量机[1-2]、神经网络[3-4]、卷积神经网络[5]以及各种语义分割模型如U-Net[6-7]和DeepLab系列[8]等相继被应用于流域水体信息的提取。这些深度学习方法能够自适应地提取遥感影像中不同层级的特征[9-10]，并通过特征融合高效挖掘丰富的语义信息。然而高分辨率图像中河流的形态不连续、形状和尺寸多变、背景环境复杂，或者是背景与河流水体相似度较高时，模型会出现错误识别。当河流的细小支流较多时，有些融入背景较深的细小支流无法被完整提取并且在河流边缘的提取也不够精细。因此，如何利用高分辨率的无人机影像精确提取流域水体信息[11]，尤其是在复杂背景下保持高精度，仍是当前亟待解决的问题。

针对上述问题，本文基于无人机光学影像和UNetMformer模型精准高效提取多尺度河流轮廓。利用无人机遥感高分辨率的特点，结合全局和局部注意力机制，在复杂背景下能够较为完整地提取出河流主干及细小支流轮廓。结合深度学习技术，高分辨率的无人机光学影像在遥感影像语义分割领域展现出巨大的潜力，为河流轮廓的精确提取提供了创新手段。

1　研究区及数据预处理

1.1　研究区域概况

九龙江位于福建省南部，是漳州市第一大河，

收稿日期：2024-05-28。
项目来源：福厦泉国家自主创新示范区海陆界面复杂环境立体监测技术协同创新平台资助项目（3502ZCQXT2023002）。
作者简介：张永年（1979—），本科，高级工程师，从事地理信息系统平台建设及应用分析工作，E-mail：13635251755@139.com。

其干线全长达258 km，流域面积达14 741 km²。研究区域选取了九龙江的部分河段，包括河流主干、若干支流以及入海口附近。研究区域中河流水体的最小和最大宽度分别约为5 m和660 m，河流水体的最短和最长长度分别约为1 800 m和9 000 m。研究区域的周边环境十分复杂，包括池塘养殖区、河滩地、城镇、桥梁以及船只等各种大量的干扰因素。在这样的背景下，进行河流轮廓提取研究对模型的分类和精细化提取能力提出了极大的挑战。

1.2　无人机数据获取及预处理

采用大疆精灵4 RTK无人机，该无人机搭载3轴云台，角度抖动量±0.02°，相机镜头为FOV 84°，有效像素为2 000万，拍摄图片格式为.JPG，拍摄时无人机飞行高度为100 m，速度50 km/h，航向和旁向重叠度分别为70%和50%。通过PHOTOSCAN和航天远景EPT软件，我们对不同河段位置的高分辨率无人机原始影像数据进行色彩平衡及影像拼接融合等处理。

1.3　样本数据集制作

样本数据集制作包括河流轮廓标签制作、样本数据集划分以及训练样本数据图像增强处理（图1），具体过程如下：①标签制作：结合0.04 m空间分辨率的无人机光学影像和现场拍摄图片，利用ArcGIS软件平台进行河流轮廓目视解译，河流轮廓标记为1（白色），背景标注为0（黑色）；将人工目视提取的河流轮廓数据进行栅格化，将无人机影像和对应河流轮廓标签数据进行裁切，裁切成大小为512×512像素大小的样本数据对，样本数据如图1所示；②数据集划分：利用简单随机抽样的方法将70%的样本数据对作为训练数据集，30%的样本对作为验证数据集，得到训练集样本10 000张，验证集样本数量4 285张；选取3张无人机影像作为测试集，尺寸大小分别为163 653×80 901、46 713×30 362、39 953×13 812；③图像增强：利用旋转、镜像、高斯模糊、亮度调节等方法对训练集中的样本数据对进行数据扩增。



无人机影像　　河流水体标签

图1　河流轮廓样本数据集

2　研究方法

2.1　UNetMformer模型

UNetMformer模型由编码器、解码器和跳跃连接构成。编码器部分由多个MultiRes[11]模块堆叠而成，这些模块在下采样过程中能够细致捕捉图像的空间特征信息；在解码器部分，我们引入了创新的Mtransformer模块，并结合了多输出特征融合结构。Mtransformer模块通过其独特的全局和局部注意力机制，能够同时提取全局和局部的特征信息，从而有效捕捉不同尺度河流存在的特征。此外，多输出的上采样操作使得模型能够充分利用底层的特征信息，进一步增强网络的细化能力，实现对河流主干及其细小支流的精细化提取。

1）下采样结构由4个MultiRes Block组成，每个MultiRes Block以2的比例因子对特征图进行下采样。MultiRes Block代替一般卷积层从而扩大感受野，实现了对目标的多尺度分析，从而提高模型多分辨率分析能力。MultiRes Block包括一个卷积组和一个1×1卷积层组成。卷积组由多个串联的轻型卷积层组成，每个轻型卷积层是由一至多个相同的轻型卷积并联组成。每个轻型卷积层的输出进行拼接，然后与1×1卷积层的输出进行融合，实现多尺度特征的融合，融合后的结果作为当前MultiRes Block的输出传输至下一阶段。

2）上采样结构由4个Mtransformer模块和多输出特征融合结构组成，Mtransformer模块包括全局-局部注意力模块、批归一化层、多层感知以及加和操作。Mtransformer模块与transformer模块区别在于前者将全局-局部注意力模块替换了多头自注意力模块，从而实现全局-局部多尺度上下文结构特征提取。全局-局部注意力模块由全局分支模块和局部分支模块构成，将全局分支模块和局部分支模块的输出进行融合、深度卷积、批归一化以及1×1卷积操作后作为全局-局部注意力模块的输出。全局-局部注意力模块的输出经过批归一化层、多层感知以及加和操作后的结果作为Mtransformer模块的输出。

3）多输出的上采样操作，获取底层特征信息以增强网络细化能力。在上采样的前3个Mtransformer模块均设有输出通道，输出3个不同维度的特征图，通过上采样将其转变为相同大小后进行特征融合，将融合后的特征图8倍上采样后与网络最后一层的输出进行融合从而得到模型的最终输出。

2.2　评价指标

为检验不同模型在不同测试区域的提取结果做出客观的精度评价，本文使用到的评价指标主要包括总精度（*OA*），F1分数（F1-score），交并比（IoU），卡帕系数（Kappa）。这些指标均为语义分割中常用的指数，能够有效评判模型语义分割结果的性能，上述评价指标的表达式分别如下：

1）*OA*指标

$$OA=\frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

式中，*OA*为指分类结果中正确分类的像素占所有像素的百分比。

2）F1-score指标

$$\begin{cases}\text{Precision}=\dfrac{TP}{TP+FP}\\ \text{Recall}=\dfrac{TP}{TP+FN}\\ \text{F1}-\text{score}=\dfrac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}\end{cases} \tag{2}$$

Precision为所有被提取的河流轮廓中被正确分类的河流轮廓比例；Recall为被正确提取的河流轮廓与影像实际河流轮廓的比例；F1-score为精确率和召回率的调和平均数，同时考虑了精确率和召回率，能够更加客观地评价模型的性能。

3）IoU指标

$$\text{IoU}=\frac{TP}{TP+FP+FN} \tag{3}$$

IoU通常用于衡量研究目标实际分布和模型提取结果之间的相似度。

4）Kappa指标。Kappa用于比较模型的预测结果与实际分类结果的一致性程度。

$$\text{Kappa}=\frac{P^o-P^e}{1-P^e} \tag{4}$$

式中，*TP*为提取对应实际河流轮廓；*TN*为提取背景对应实际背景；*FP*为提取河流轮廓对应实际背景；*FN*为提取背景对应实际河流轮廓；P^o、P^e均根据混淆矩阵进行计算。

3 实验结果与分析

本文基于PyTorch框架进行模型的搭建，使用NVIDIA GeForce RTX3090 GPU对模型训练及测试过程进行加速。在本文中训练批次设置为24，输入图像大小为512×512，迭代次数为200，初始学习率为0.005，权重衰减为0.000 。

为测试各个模型在不同背景环境和河流长度下的提取精度，本文选取了3个不同长度的河流段作为测试区域。根据河流的长度和平均宽度进行分类，a为区域河流长度和宽度分别约为8 000 m和150 m，为大型河流；b为区域河流长度和宽度分别约为3 000 m和50 m，为中型河流；c为区域河流长度和宽度分别约为1 800 m和15 m，为小型河流。测试区域的地物环境包括池塘养殖区、河滩地、城镇、桥梁以及船只等干扰因素，在这样的背景环境下，能够充分检验各个模型的提取精度。将测试数据集分别输入训练完成的UNet、DeepLabV3+以及UNetMformer模型中，通过对比河流轮廓提取结果的空间分布和评价指标从定性和定量角度评价各个模型的提取精度。

3.1 不同提取方法定性分析

如图2所示，图中第1列代表无人机影像，第2列为各个区域河流轮廓的真实分布情况，第3到5列依次为UNet、DeepLabV3+、UNetMformer模型分类结果的空间分布。通过对比可知，在a区域中，UNet模型在河段下游船只密集处的分类结果最差，提取结果存在较大范围破碎的情况，DeepLabV3+模型同样也存在相同的问题，但是完整度略高于UNet。河流轮廓提取最为完整地是UNetMformer模型，与真实分布情况十分相似，对于河内干扰物的区分也十分准确，能很完整的提取出河流轮廓。在b区域能够发现UNetMformer模型在细小支流的提取上效果更佳。因为在b区域的左上位置存在一条细小支流，UNet和DeepLabV3+模型均未将其完整提取，但UNetMformer模型可以提取的较为完整。在c区域中，UNet模型在中段河流出现了大面积破碎现象，而DeepLabV3+和UNetMformer模型的分类结果均与真实分布较为接近。综上可知，UNetMformer模型在3区域的河流轮廓提取结果最佳。

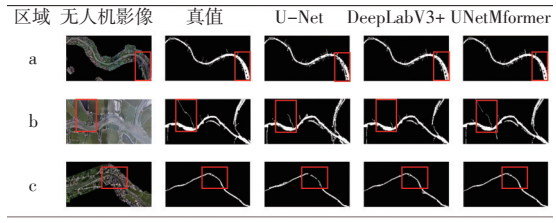


图2 不同区域各个模型的分类结果

为更为直观地对比各个模型在河流中存在干扰物边缘、河流主干流、河流细小分支处的分类精度，本文选取了6个典型场景进行对比，如图3所示。在a、b场景的红框处，可以发现在遇到干扰物时，UNetMformer模型能够准确区分河流轮廓、船只以及边缘陆地，U-Net和DeepLabV3+模型在这些场景均出现了错分现象，将停靠船只边缘和陆地部分提取为水体；在c、d场景的红框处，可以发现在对于主干流的提取上，U-Net和DeepLabV3+模型存在漏分和错分现象，尤其是U-Net模型，将水中的绿色植被和河滩分类为水体；在e、f场景下对于细小支流的提取上，U-Net模型存在漏分现象，在支流末端提取不完整，边缘提取不够细腻，而UNetMformer模型能够准确精细提取出河流轮廓，区分植被、船只以及浸水陆地。

综上可知，从定性角度进行流域水体提取结果对比，具备结合全局和局部特征能力的UNetMformer模型在3个大区域和6个经典场景中的表现均为最优，

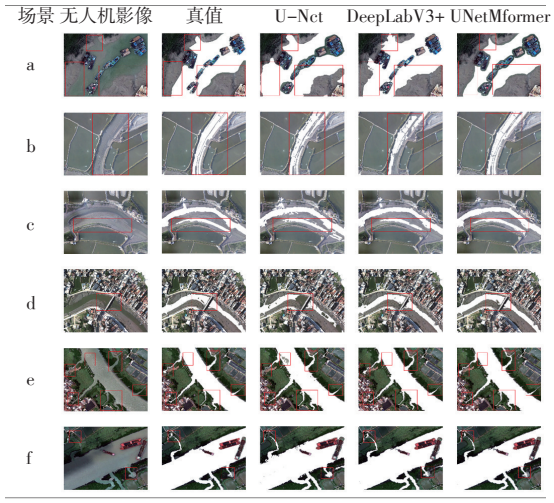


图3　不同典型场景各个模型分类结果

在复杂背景下能够较为完整地提取出河流主干及细小支流，并且在水体与陆地的边缘处提取更为精细。

3.2　不同提取方法定量分析

为验证改进模型UNetMformer的有效性，本文根据混淆矩阵计算精度评价指标以评估各个模型对河流轮廓的提取精度，各个模型的精度评价结果如表1所示，其中下划线和加粗部分代表在该区域该评价指标的最佳精度水平。对比4个精度评价指标的结果可知，从定量角度分析，UNetMformer模型在a、b和c区域的提取精度均为最优。

表1　不同区域河流轮廓提取精度评价/%

区域	模型	*OA*	*F*1－score	IoU	Kappa
a	U-Net	98.90	91.06	83.59	90.48
	DeepLabV3+	99.54	96.34	92.94	96.09
	UNetMformer	99.58	96.74	93.68	96.52
b	U-Net	96.92	83.98	72.39	82.28
	DeepLabV3+	98.36	90.62	82.85	89.73
	UNetMformer	98.79	93.51	87.81	92.84
c	U-Net	99.23	85.60	74.83	85.21
	DeepLabV3+	99.68	94.21	89.05	94.04
	UNetMformer	99.89	97.94	95.96	97.88

4　结　语

本研究提出了一种新的语义分割模型UNetMformer，通过对比UNetMformer模型与经典U-Net、DeepLabV3+在多种实际场景下对河流轮廓提取结果，分析其空间分布特征，从定性和定量的角度进行精度评估。研究结果表明，UNetMformer模型凭借其独特的全局和局部注意力机制，在河流轮廓提取方面展现出了显著的优势。相较于其他2个模型，UNetMformer在3个测试区域的提取精度最高，而且在处理过程中消耗的时间也更为短暂。这一结果充分证明了UNetMformer模型在应对不同尺度河流提取任务时的广泛适用性和高效性。

参考文献

[1] 邱凤婷，过志峰，张宗科，等. 基于改进SVM分类法的SAR图像水体面积提取研究[J]. 地球信息科学学报，2022，24(5): 940-948

[2] Gong W，Liu T，Jiang Y，et al. Applicability of the Surface Water Extraction Methods Based on China's GF-2 HD Satellite in Ussuri River，Tonghe County of Northeast China [J]. Nature Environment & Pollution Technology，2020，19(4): 1 537-1 545

[3] Li L，Yan Z，Shen Q，et al. Water Body Extraction from Very High Spatial Resolution Remote Sensing Data Based on Fully Convolutional Networks [J]. Remote Sensing，2019，11(10):1 162

[4] Long J，Shelhamer E，Darrell T. Fully Convolutional Networks for Semantic Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2015，39(4): 640-651

[5] Chen Y，Fan R，Yang X，et al. Latif a Extraction of Urban Water Bodies from High-resolution Remote-sensing Imagery Using Deep Learning [J]. Water，2018，10(5):585

[6] 何红术，黄晓霞，李红旮，等. 基于改进U-Net网络的高分遥感影像水体提取[J]. 地球信息科学学报，2020，22(10): 2 010-2 022

[7] 厉芳婷，胡菡，刘韬，等. 基于深度学习的遥感影像地表水时空监测与分析[J]. 地理空间信息，2023，21(9):52-56

[8] Chen L C，Papandreou G，Kokkinos I，et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected Crfs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2014(1):1 412

[9] Ronneberger O，Fischer P，Brox T. U-net: Convolutional Networks for Biomedical Image Segmentation[C]//Medical Image Computing and Computer-assisted Intervention-MICCAI 2015: 18th International Conference，Munich，Germany，October 5-9，2015，Proceedings，Part III 18. Springer International Publishing，2015: 234-241

[10] 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，2017，40(4): 834-848

[11] Ibtehaz N，Rahman M S. Multi Res U-Net: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation [J]. Neural Networks，2020，121:74-87