

文章编号:1671-5896(2026)04-0998-10

基于预训练迁移学习的高原区域高分辨率水体提取

易志伟^{1,2}, 程鑫¹, 马经宇¹, 顾玲嘉², 邹波¹, 朱瑞飞¹

(1. 长光卫星技术股份有限公司 数据中心三室, 长春 130102; 2. 吉林大学 电子科学与工程学院, 长春 130012)

摘要: 针对为提升三江源区域高分辨率水体提取的细粒度问题, 通过对吉林一号亚米级遥感影像的研究与分析, 提出一种基于私有数据预训练的迁移学习方法。基于 ViT (Vision Transformer) 骨干网络进行自监督预训练, 并构建以 ViT-B 为编码器、UperNet 为解码器的水体迁移学习提取框架。实验结果表明, 该方法的 IoU (Intersection over Union) 达到 92.93%, 较未预训练模型提升 11.87%, 且优于其他开源预训练权重。最终提取水体图斑 25 621 个, 水体面积 1 544 km²。与 WorldCover 2021 对比分析显示, 该方法可有效检测出被遗漏的细小水体, 为高原湿地水资源分析提供了更精准的参考。

关键词: 高分辨率遥感影像; 预训练; 迁移学习; 三江源地区; 微小水体提取

中图分类号: TP75

文献标志码: A

High-Resolution Water Body Extraction in Plateau Regions Based on Pre-Trained Transfer Learning

YI Zhiwei^{1,2}, CHENG Xin¹, MA Jingyu¹, GU Lingjia², ZOU Bo¹, ZHU Ruifei¹

(1. Data Center 3, Chang Guang Satellite Technology Company Limited, Changchun 130102, China;

2. College of Electronic Science and Engineering, Jilin University, Changchun 130012, China)

Abstract: This study aims to improve the accuracy of high-resolution water body extraction in the Sanjiangyuan region. Using sub-meter Jilin-1 satellite imagery, a self-supervised pretraining approach based on the ViT (Vision Transformer) backbone is developed, followed by fine-tuning a water body extraction model with ViT-B as the encoder and UperNet as the decoder. The results demonstrate that the proposed method achieves an IoU (Intersection over Union) of 92.93%, outperforming non-pretrained models by 11.87% and surpassing other open-source pretrained weights. A total of 25 621 water body patches with an area of 1 544 km² are extracted. Comparative analysis with WorldCover 2021 reveals that the method effectively detects small water bodies missed by previous methods, providing a more accurate reference for water resource analysis in plateau wetlands.

Key words: high-resolution remote sensing images; pre-trained models; transfer learning; Sanjiangyuan region; water body extraction

0 引 言

卫星遥感数据是环境保护、资源管理、城市建设等领域的主要数据来源之一^[1-2]。三江源地区位于青藏高原, 是长江、黄河、澜沧江的发源地, 并且拥有众多湖泊, 被誉为“中华水塔”, 是我国重要的水源

收稿日期: 2025-05-21

基金项目: 长春市科技发展计划基金资助项目(2024WX02)

作者简介: 易志伟(1996—), 男, 江西萍乡人, 长光卫星技术股份有限公司工程师, 吉林大学博士研究生, 主要从事卫星遥感地表覆盖分类研究, (Tel)86-13077900011 (E-mail)yizw2018@radi.ac.cn; 顾玲嘉(1981—), 女, 江苏灌云人, 吉林大学教授, 博士生导师, 主要从事人工智能遥感技术与应用研究, (Tel)86-431-85155485 (E-mail)gulingjia@jlu.edu.cn; 通讯作者: 朱瑞飞(1986—), 男, 山西朔州人, 长光卫星技术股份有限公司正高级工程师, 博士, 主要从事卫星遥感数据挖掘与应用研究, (Tel)86-18686682764 (E-mail)zhuruiwei@jlu.cn。

涵养区, 其中唐北区域是世界上湿地海拔最高、面积最大、分布最集中区域, 是三江源重要组成和我国重要生态安全屏障, 因此精确掌握三江源唐北地区的地表水覆盖面积对三江源地区及中国的生态安全具有重要意义。然而, 目前有关该区域的水体分布研究主要集中于 10 m 等中分辨率产品, 导致一些细小的湖泊河流被忽略, 不能精确提取三江源唐北区域水体面积数据及小型湖泊的空间分布。因此有必要利用亚米尺度的超高分辨率影像准确提取三江源唐北区域的水体空间分布范围。

目前从遥感影像中进行大规模水体提取的准确性与有效性已有许多研究^[3]。早期遥感影像中水体的提取采用基于规则的方法, McFeeters 首先引入了归一化差分水指数 (NDWI: Normalized Difference Water Index)^[4], 利用 Landsat 多光谱数据的绿色波段和近红外波段检测水体, 其充分考虑了水体对绿波段反射较强, 对近红外波段反射较弱的光谱特性, 并已发展为最重要的遥感指标之一。但由于大气条件、云层覆盖和气溶胶随时间分布的变化, 同一区域图像的水体阈值变化很大, 不同区域的图像也会导致阈值的巨大差异, 尤其对于复杂的高原地区, 这使选择阈值变得十分困难。需要注意的是, 高原地区水体受冰川融水和含沙量影响, 存在大量微小水体, 其影像特征与稀疏草地及沙地相似, 极易混淆。因此, 基于规则的方法难以有效区分遥感影像中的水体与背景, 且适配超高分辨率影像效果较差。

近年来, 深度学习越来越多地被用于高分辨率遥感影像中识别水体和其他物体, 并获得了优异结果^[5-7]。与基于规则的水体提取算法相比, 深度学习可有效提取低层位置和高层语义信息, 具有强大的多尺度特征表示能力。一系列 CNN (Convolutional Neural Network) 模型, 包括 UNet^[8]、PSPNet^[9]、DeepLab^[10]、SegNet^[11] 和 RefineNet^[12] 等, 被用于遥感影像的信息提取。然而, 高质量的训练样本一直是深度学习提取水体的主要障碍和挑战^[13]。不同区域的水体特征可能存在显著差异, 使一个区域的水体样本不适用于其他区域。尤其高原地区的地表景观特征与其他区域有显著的区别, 导致现有的研究样本也难以直接使用。

为应对下游样本缺失与模型泛化性差的问题, 当前计算机视觉领域提出了许多以大规模预训练范式为核心的新方法。例如, 在 2012 年深度学习取得巨大成功后, 已有研究使用大规模图像分类数据集 ImageNet^[14] 做图像分类代理任务, 以进行模型预训练。He 等^[15] 于 2022 年发展了基于掩码重建学习 (MAE: Masked Autoencoders) 的代理任务, 可获得图像高质量表示。Oquab 等^[16] 于 2023 年采用多粒度对比学习策略在 1.4 亿张图片上进行预训练, 以获取图像的高质量表征。使用基于大规模预训练范式构建的通用视觉表征模型 DINOv2 (Distillation with No Labels version2), 依托 Transformer 架构的全局注意力机制, 在对图像进行图像编码的前提下, 其迁移学习能力得到大幅提升, 为下游任务的快速适配提供了理论保障与工程实践基础。但当前的视觉预训练模型多基于自然图像训练而成, 而对遥感图像表征效果不明, 因此有必要探索基于私有应用数据预训练加微调的方式提升下游应用的性能。

为克服三江源唐北区域水体提取应用的局限性, 笔者以吉林一号高分辨率卫星影像为数据源, 提出了一种基于预训练方式的深度学习高分辨率遥感影像水体提取方法, 并实现了三江源唐北区域水体空间分布制图。该方法的 IoU (Intersection over Union) 达到 92.93%, 较未预训练模型提升 11.87%, 且优于其他开源预训练权重。其核心包括先利用掩码预测任务与研究区域高分辨率遥感影像对深度学习模型骨干进行预训练; 然后利用公开 10 m 分辨率土地覆盖产品结合 NDWI 指数, 获取研究区域置信度比较高的水体提取结果, 并制作成水体提取样本; 最后对预训练的深度学习模型在水体样本上进行监督训练, 并将训练完成的模型应用于整个区域的水体提取。

1 研究区概况

研究区区位及影像概况如图 1 所示。三江源地区地处青海省西南部, 地理范围为北纬 32°26'4" ~ 36°16'49", 东经 89°24'6" ~ 99°6'46", 总面积 190 700 km², 平均海拔超过 4 700 m。该地区属于青藏高原气候大区, 冷暖两季且雨热同季, 多年平均气温为 -5.6 ~ 7.8 ℃。三江源地区水系发达, 是长江、黄河、澜沧江的发源地, 源区湖泊众多, 对中国乃至亚洲的生态安全具有重要意义。三江源唐北区域为世界上湿地海拔最高、面积最大、分布最集中区域, 其不仅是长江发源地和重要流域水源发育区, 还是三江源

重要组成部分和我国重要生态安全屏障。三江源唐北区域总面积约为 47 000 km²。

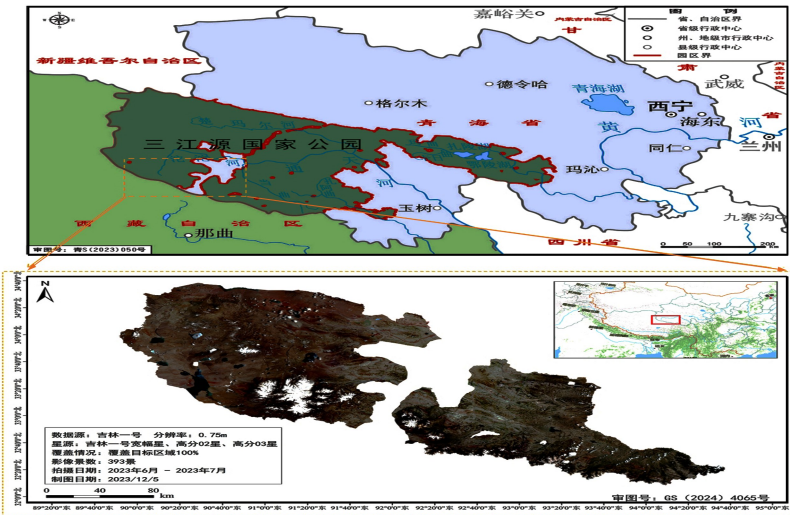


图1 研究区域概况

Fig. 1 Geographical overview of the research area

2 基于预训练模型的三江源地区水体提取框架

为在少量标签的情况下学习到水体的特征，笔者提出的私有化预训练的少样本条件下三江源水体提取框架如图 2 所示，其以 Vision Transformer^[17]加 UperNet^[18]作为水体分割基本框架。利用全域无标签数据对 Vision Transformer 进行重建掩码学习预训练，使骨干模型可学习到高原地区遥感影像的通用特征。然后，构建小批量水体标注数据集对分割模型进行微调。最后利用微调完成的模型对全域进行滑窗推理，获取三江源地区唐北区域水体分布范围。

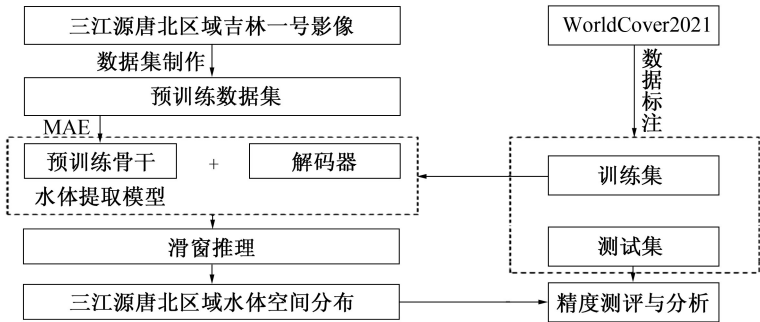


图2 总体技术流程

Fig. 2 Overall technical flowchart

2.1 数据

2.1.1 区域影像

笔者通过分析吉林一号亚米级卫星获取星座的高频次数据，并筛选了 2023 年 6 月-2023 年 7 月 393 景的云量小于 10% 的吉林一号卫星遥感影像，以覆盖三江源唐北区域。经影像预处理后获得覆盖唐北区域全域的一张图影像(见图 1)，数据整体空间分辨率为 0.75 m，影像详细信息如表 1 所示。

表 1 研究区吉林一号影像信息统计

Tab. 1 Summary of Jilin-1 image in the research region

卫星名称	分辨率/m	幅宽/km	成像模式	数量/景	卫星名称	分辨率/m	幅宽/km	成像模式	数量/景
宽幅 01A 星		136	推扫	337	高分 03 星		17	推扫	41
高分 02 星	0.75	40	推扫/立体	15					

2.1.2 自监督数据集

笔者将 393 景吉林一号遥感影像按 518×518 像素大小进行样本的裁切, 共获取到 778 660 个样本, 并对受云影像以及位于影像边缘的样本进行剔除, 将剩余的样本按空间均匀分布筛选, 最终构建了一个包含 500 000 张无标注的高质量样本的自监督预训练数据集, 如图 3 所示。自监督样本具备全面性, 基本覆盖大部分研究区域, 同时对高原湖泊、细长河流、细小水体等充分覆盖。

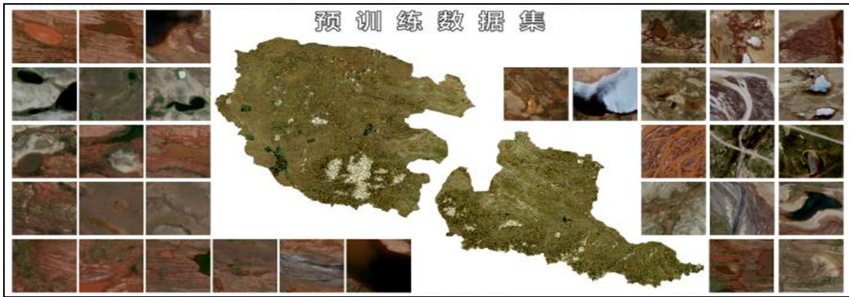


图 3 预训练数据示意图

Fig. 3 Example of pre-trained data

2.1.3 水体标注数据集

研究使用的水体标签来源于 WorldCover 2021^[19] 土地覆盖数据集, 该数据集基于 Sentinel-2 和 Sentinel-1 卫星影像, 包含水体在内的 11 种地表覆盖类型总体精度达 76.7%, 原始结果中存在大量错误的水体碎斑(见图 4), 若直接作为水体训练标签将给训练引入大量噪声。由于研究的主要模型是 patchsize 为 14 的 Vision Transformer 架构, 样本尺寸应为 14 的整数倍。为此, 笔者基于 WorldCover 中水体覆盖区域挑选了 5 000 个 518×518 像素大小区域进行样本标注, 其过程包括栅格矢量化、矢量平滑与人工修正, 保证水体标注严格贴合边界, 结果如图 5 所示。最终, 共制作 5 000 张 518×518 像素的图像标签对, 样本区域覆盖面积约为 754 km², 占研究区总面积的 1.55%。在后续模型训练与性能评估中, 将水体标注数据集随机按 8 : 2 进行训练集与测试集划分, 其中 4 000 张用于模型训练, 1 000 张用于模型性能评价。

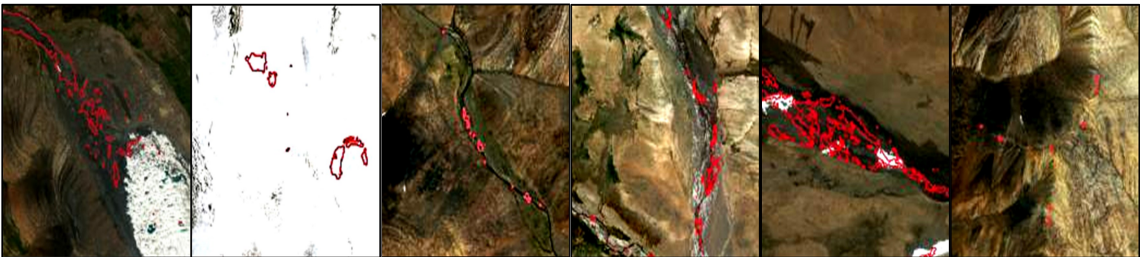
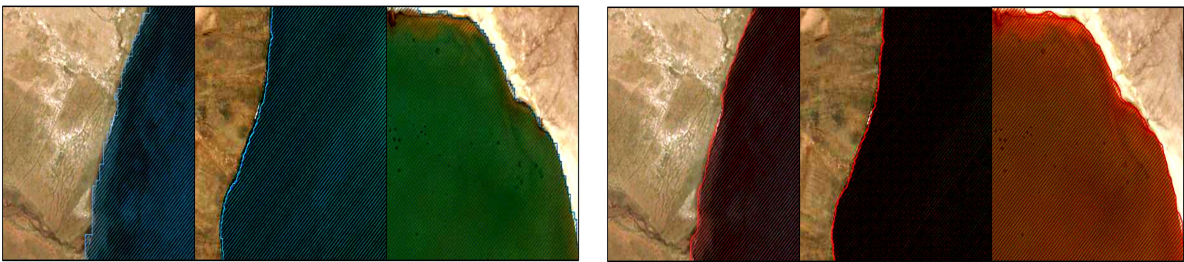


图 4 WorldCover 2021 水体图层示意图

Fig. 4 WorldCover 2021 water body illustration



a 未平滑

b 平滑

图 5 水体样本示意图

Fig. 5 Water body sample schematic

2.2 Vision Transformer 骨干模型

ViT(Vision Transformer)是一种基于纯自注意力机制的图像处理架构，其突破了传统卷积神经网络(CNN)的局部感受野限制。ViT 将输入图像分割为固定尺寸的序列化图像块(patches)，通过线性投影映射为嵌入向量，并添加可学习的位置编码以保留空间信息。在嵌入序列前预置分类标记(class token)后，输入多层 Transformer 编码器处理，每层包含多头自注意力(MHSA: Multi-Head Self-Attention)和前馈网络(FFN: Feedforward Network)模块。自注意力机制通过全局交互建模图像块间长程依赖，生成上下文感知的特征表示，用于下游应用。ViT 摒弃了 CNN 的归纳偏置，依赖大规模数据预训练实现卓越性能，验证了 Transformer 在视觉任务中的普适性。当前众多视觉基础模型(如 DINOv2、MAE)均以 ViT 为核心架构，其核心优势在于全局上下文捕捉能力和灵活的可扩展性。笔者选择 ViT-Base 模型作为基础视觉骨干架构，ViT-Base 模型的参数选择包括：窗口划分大小 Patch Size 为 14，隐藏层维度为 768，具体架构如图 6 所示。

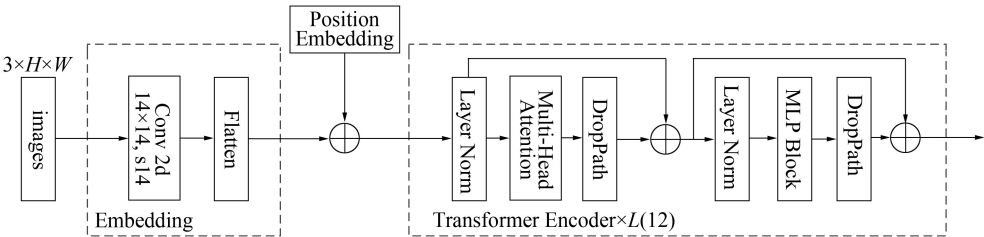


图 6 Vision Transformer 网络结构示意图

Fig. 6 ViT schematic architecture

2.3 预训练算法

笔者借鉴了图像自监督领域广泛应用的掩码预测任务对 ViT 骨干模型进行预训练。预训练对样本进行75%面积的掩码，以 518×518 像素输入为例，首先在模型输出 14×14 特征图上随机选取 75% 的像素，根据像素位置，确定其在输入图上对应的区域，并将该区域像素值替换为 0，如图 7 所示。将掩码影像输入到骨干网络中，获取到影像特征后，利用简单的线性层将特征重建为输入影像大小，并利用掩码部分原始像素结合均方误差损失函数对重建的影像进行监督训练。

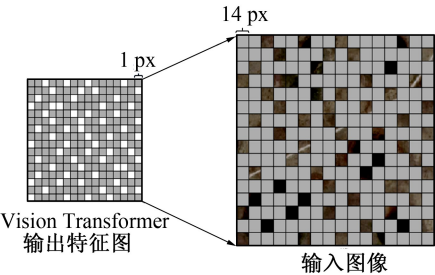


图 7 输入图像随机掩码示意图

为提升预训练效率，笔者借鉴了 GreenMIM^[20] 架构 Fig. 7 Illustration of random masking on input images 的训练方法，采用分组窗口注意力与稀疏卷积的方式，在保持出色性能的同时，显著提高了预训练过程的速度达 2.7 倍，同时减少了 70% 的 GPU (Graphics Processing Unit)内存的使用量。如图 8 所示。

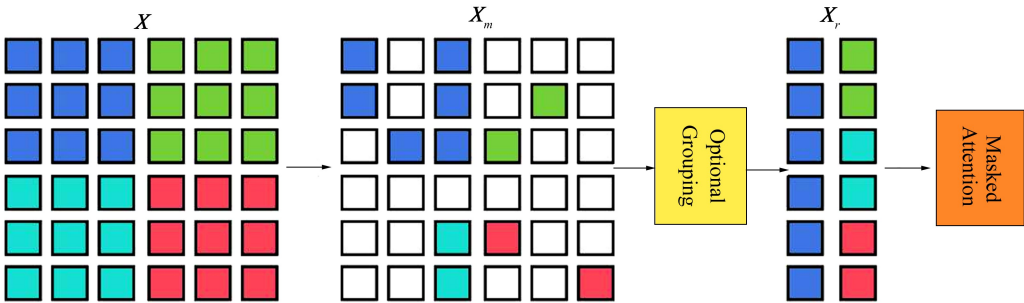


图 8 分组窗口注意力示意图

Fig. 8 Schematic diagram of grouped window attention

分组窗口注意力机制通过将具有不同数量可见图块的局部窗口聚集成几个大小相等的组, 在每组内进行掩码自注意力计算, 从而降低了计算复杂度。稀疏卷积则利用稀疏数据的优势, 减少了无效计算, 进一步提升了运算速率。

实验所涉及的程序均使用 Pytorch+Python 框架实现, 表 2 为自监督模型训练的超参数信息。

表 2 自监督模型训练的超参数信息

Tab. 2 Hyperparameter configuration for model pretraining

批量大小	优化器	迭代轮数	权重衰减	初始学习率	学习率更新策略	线性预热/轮
224	AdamW	300	0.05	0.001	Cosine Annealing	40

2.4 基于预训练骨干的水体分割模型微调

利用预训练 Vision Transformer 模型作为影像的特征提取骨干, UperNet^[20] 模型做为解码器, 如图 9 所示, 在水体数据集上进行训练。UperNet 同时对骨干模型输出的 4 个尺度特征进行交互与融合, 使解码器能充分学习到图像不同尺度的信息。

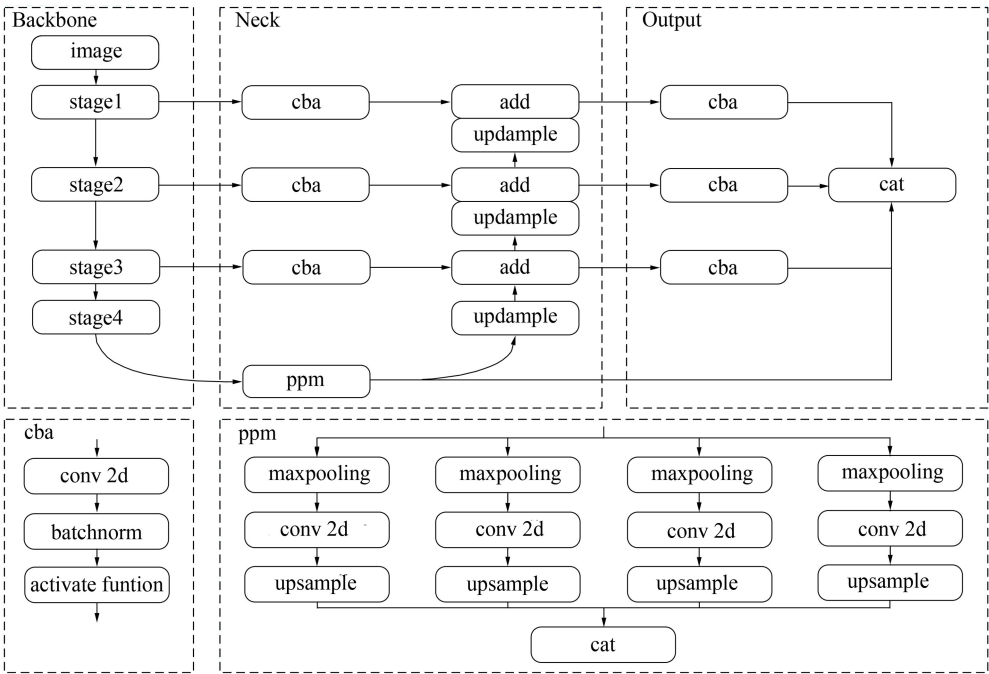


图 9 Vision Transformer+UperNet 水体提取模型
Fig. 9 ViT-UperNet water body extraction model

经预训练的模型骨干可对图像进行较好表征, 下游微调仅需对解码器进行少量微调训练。因此, 前 80 轮冻结骨干, 只对解码器部分进行训练, 后 20 轮解冻骨干参数, 对整个水体模型进行训练, 预训练微调超参数如表 3 所示。为探究下游水体提取应用中基于应用数据重新预训练的必要性, 与其他 4 种经过不同方法预训练再初始化权重微调训练的水体提取 ViT 模型进行应用对比分析, 分别是随机初始化的骨干 (ViT_Random_Water)、ImageNet 有监督训练骨干 (ViT_ImageNet_Water)、开源 DINOv2 骨干 (ViT_DINOv2_Water) 和开源 MAE (ViT_MAE_Water) 预训练骨干。

表 3 模型微调超参数配置

Tab. 3 Hyperparameter configuration for model fine-tuning

输入图	批量	优化器	优化器	迭代	权重	初始学	学习率更	线性预	数据	损失
像/像素	大小		动量	轮数	衰减	习率	新策略	热/轮	增强	函数
518×518	128	AdamW	(0.9, 0.999)	100	0.05	0.000 1	Cosine Annealing	10	RandomResized Crop, RandomFlip	CrossEntropy

2.5 精度评价

为验证笔者提出的水体提取框架的准确性与有效性,采用水体像素的交并比(IoU, I)、 F_1 值(F_1 -Score, F_1)、召回率(Recall, R)和准确率(Precision, P)作为评价标准。如下:

$$I = \frac{T_p}{T_p + F_p + F_N}, \tag{1}$$

$$R = \frac{T_p}{T_p + F_N}, \tag{2}$$

$$P = \frac{T_p}{T_p + F_p}, \tag{3}$$

$$F_1 = 2 \times \frac{RP}{R + P}, \tag{4}$$

其中 T_p (True Positive)表示实际预测为水体且预测为水体的水体像素, F_p (False Positive)表示实际为非水体但预测为水体的像素, F_N (False Negative)表示实际为水体但预测为非水体的像素。在相同的 1 000 张图片测试集上,对笔者提出的水体算法 ViT_MAEPRI_Water 与 ViT_Random_Water、ViT_ImageNet_Water、ViT_DINOv2_Water 和 ViT_MAE_Water 进行精度评价。

3 实验与分析

3.1 预训练实施效果

预训练完成后,在部分样本上利用预训练模型进行了掩码恢复测试,结果如图 10 所示。从测试样本中可看到模型具备一定恢复复杂背景的能力,对三江源地区的水域、湿地、草地和冰川等典型场景均可做到较好的恢复,说明骨干网络学习到了研究区域影像的通用特征,可用于后续水体提取应用。

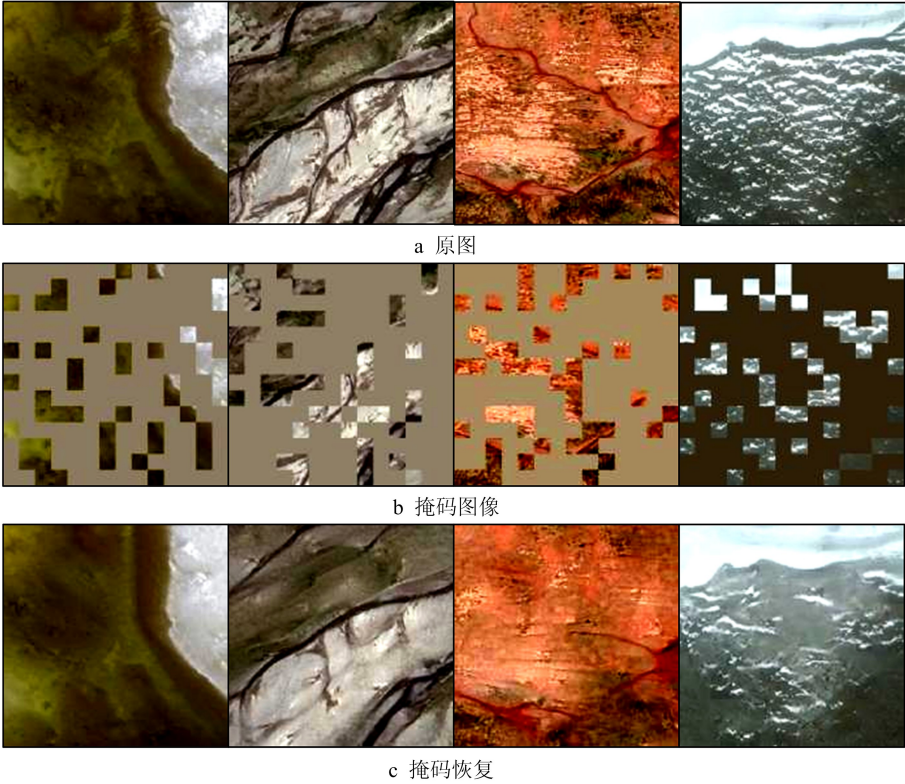


图 10 预训练模型测试

Fig. 10 Testing of image masked reconstruction based on pre-trained models

3.2 模型性能评测

基于不同预训练模型初始化微调进行水体提取应用的精度评估结果如表 4 所示。从表 4 可看出, 采用三江源目标区域全域数据进行 MAE 预训练进行水体提取的方法 ViT_MAEPRI_Water 精度最高, IoU 为 92.93%, F_1 值为 96.33%, 召回率为 95.13% 及准确率为 97.57%, 均优于其他预训练模型微调的水体提取算法。采用预训练模型初始化的方法均显著优于未利用预训练模型初始化的 ViT_Random_Water, 整体精度提升超过 7%。上述结果表明, 采用预训练加微调方式进行下游水体提取应用效果优于随机初始化结果, 采用私有数据支撑进行预训练的下流水体提取应用效果优于当前开源的基于自然图像的预训练初始化的水体提取模型。

表 4 模型精度评估结果

Tab.4 Model accuracy evaluation results %

模型	IoU	F_1 -score	Recall	Precision	模型	IoU	F_1 -score	Recall	Precision
ViT_Random_Water	81.06	79.54	78.43	80.67	ViT_DINOv2_Water	89.42	89.27	88.26	90.32
ViT_ImageNet_Water	88.23	88.61	87.59	89.65	ViT_MAEPRI_Water	92.93	96.33	95.13	97.57
ViT_MAE_Water	88.64	88.86	87.96	89.78					

对于收敛速度, 经过预训练的模型 ViT_MAEPRI_Water 能较快实现收敛, 而未做预训练的模型训练 100 轮仍然未完全收敛(见图 11)。同时, 对于效率, 经预训练的模型先微调解码器可在快速提升性能的同时保持高效率, 骨干模型解冻后仍然能使性能有小幅提升, 整个微调过程较未经过预训练的模型在效率上提升约 2 倍。

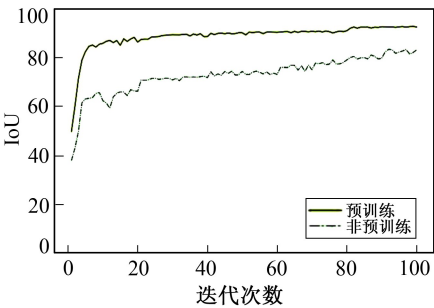


图 11 模型训练 IoU 曲线图

Fig.11 IoU curve of model training

3.3 三江源唐北区域水体分布制图

利用 ViT_MAEPRI_Water 水体提取模型, 对图 1 所示研究区域吉林一号镶嵌影像进行水体提取。模型使用单张 GeForce RTX 3090 对研究区域影响进行推理, 其推理效率约为 5 200 km²/h, 共计用时约 10 h 完成全域推理。推理滑窗大小为 518×518 像素, 滑窗间保持 20% 像素的重叠, 推理结果如图 12 所示。研究区域共提取水体图斑 25 621 个, 水体总面积 1 544 km², 占研究区域总面积约 3.2%。图 12 给出了本研究与世界覆盖 2021 的提取结果对比, 可视化效果显示本研究基于 0.75 m 分辨率影像的预训练加微调的深度学习方法, 可有效提取细小河流与湖泊, 水体提取结果更加细致, 且边界效果更好, 而 WorldCover 2021 采用多时相 10 m 分辨率的哨兵系列影像, 存在较多细小水体遗漏。

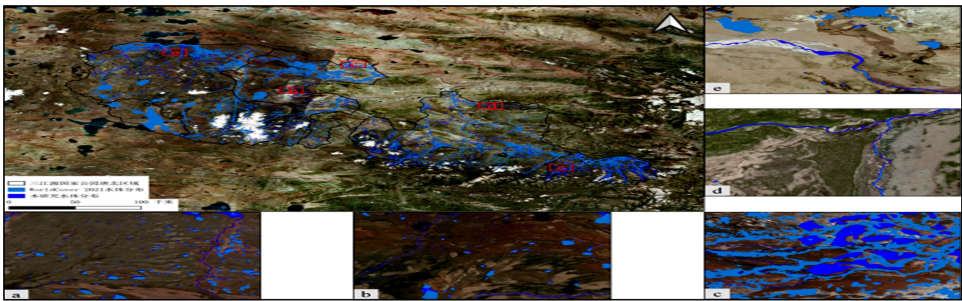


图 12 三江源地区唐北区域水体分布图

Fig.12 Distribution map of water bodies in the Northern Tanggula Mountain region of the Sanjiangyuan area

为有效评估笔者提出的基于私有数据预训练模型高分辨率影像提取三江源地区水体的优越性, 笔者将提取的 25 621 个图斑与世界覆盖 2021 的水体结果进行空间叠置分析, 统计并未与世界覆盖 2021 相交的水体图斑及其面积分布结果如表 5 所示。结果表明, 在 25 621 个水体图斑中, 有 13 502 个图斑并未与世界覆盖 2021 相交, 未相交图斑数占比 52.69%。其中无法由 WorldCover 2021 获取的水体区域

呈现面积越小比例越多的趋势,小于1 000 m²的图斑中其占比为90.51%,1 000~5 000 m²中存在52.33%。当图斑面积大于5 000 m²时,无法由WorldCover 2021获取的水体区域比例减少为10.98%,至10 000 m²时,其比例减少至3.9%。上述结果表明,基于高分辨率影像的三江源地区唐北区域水体提取较已有结果有明显细粒度提升,而中分辨率水体提取方法对于细小湖泊无法准确提取,其可为后续高原湿地地区水资源分析提供参考。

表5 水体提取结果与 WorldCover 2021 叠置分析

Tab.5 Overlay analysis of water body extraction results with WorldCover 2021

面积分布/m ²	0-1 000	1 000-5 000	5 000-10 000	>10 000	总计
未相交图斑数	2 087	5 401	2 446	159	13 502
提取图斑数	2 209	6 064	3 213	4 013	25 621
百分比/%	90.51	52.33	10.98	3.9	52.69

4 结 语

笔者提出了一种基于私有数据预训练加下游应用微调的高分辨率遥感影像水体提取方法,有效实现了三江源唐北区水体空间分布精细化提取。采用ViT-B骨干网络,基于目标区域393景吉林一号亚米遥感影像进行自监督预训练。针对区域小规模水体标注样本集,构建以ViT-B为编码器、UperNet为解码器的水体提取模型。实验结果表明,该方法IoU达92.93%,较未预训练基线提升11.87%,且性能优于其他开源预训练权重。最终,笔者获取了研究区域水体空间分布结果,总计提取水体图斑25 621个,水体面积1 544 km²。将其结果与WorldCover 2021叠置分析可知,提取水体图斑中52.69%被WorldCover 2021遗漏,小于1 000 m²的微小水体中其遗漏占比高达90.1%,因此其为高原湿地地区水资源分析提供了更精准的参考。综上,基于预训练迁移学习的高分辨率水体提取方法可提取并发现已有分辨率难以发现的微小水体,为高原区域水资源监测提供更精细的数据基础。未来,该方法有望在更广泛的地理环境 and 应用场景中发挥更大的作用,为全球湿地保护和水资源管理提供更精准、更高效的技术支持。

参考文献:

[1]陈静,王晓轩,吴宇静,等. 基于CNN的零样本城市遥感影像场景分割算法[J]. 吉林大学学报(信息科学版), 2023, 41(4): 739-745.
CHEN J, WANG X X, WU Y J, et al. Zero-Sample Urban Remote Sensing Image Scene Segmentation Algorithm Based on Convolutional Neural Network [J]. Journal of Jilin University (Information Science Edition), 2023, 41(4): 739-745.

[2]李王波,范昕桐,顾玲嘉. 基于星载被动微波的中国东北森林雪深反演[J]. 吉林大学学报(信息科学版), 2023, 41(5): 914-921.
LI W B, FAN X T, GU L J. Snow Depth Retrieval for Forest Area in Northeast China Based on Spaceborne Passive Microwave [J]. Journal of Jilin University (Information Science Edition), 2023, 41(5): 914-921.

[3]BANGIRA T, ALFIERI S M, MENENTI M, et al. Comparing Thresholding with Machine Learning Classifiers for Mapping Complex Water [J/OL]. Remote Sensing, 2019, 11(11): 1351 [2025-04-20]. <https://doi.org/10.3390/rs11111351>.

[4]都金康,黄永胜,冯学智,等. SPOT卫星影像的水体提取方法及分类研究[J]. 遥感学报, 2001(3): 214-219.
DU J K, HUANG Y S, FENG X Z, et al. Research on Water Extraction Method and Classification of SPOT Satellite Images [J]. Journal of Remote Sensing, 2001(3): 214-219.

[5]陈文倩,丁建丽,李艳华,等. 基于国产GF-1遥感影像的水体提取方法[J]. 资源科学, 2015, 37(6): 1166-1172.
CHEN W Q, DING J L, LI Y H, et al. Water Extraction Method Based on Domestic GF-1 Remote Sensing Images [J]. Resource Science, 2015, 37(6): 1166-1172.

[6]张晗涛,胡荣明,姜友谊,等. 改进DeeplabV3+模型的河流水体提取[J]. 遥感信息, 2023, 38(3): 146-152.
ZHANG H T, HU R M, JIANG Y Y, et al. River Water Extraction with Improved DeeplabV3+Model [J]. Remote Sensing Information, 2023, 38(3): 146-152.

[7]杨笑天,鱼昕,刘铭,等. 基于UPCBAM-RYOLO V5的光学遥感舰船小目标检测[J]. 吉林大学学报(信息科学版), 2024, 42(6): 1048-1057.

- YANG X T, YU X, LIU M, et al. Optical Remote Sensing Ship Small Target Detection Based on UPCBAM-RYOLO V5 [J]. Journal of Jilin University (Information Science Edition), 2024, 42(6): 1048-1057.
- [8] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C] // Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer, 2015: 234-241.
- [9] ZHAO H, SHI J, QI X, et al. Pyramid Scene Parsing Network [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 6230-6239.
- [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, 2018, 40(4): 834-848.
- [11] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [12] LIN G, MILAN A, SHEN C, et al. RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 1925-1934.
- [13] LI J, MENG Y, LI Y, et al. Accurate Water Extraction Using Remote Sensing Imagery Based on Normalized Difference Water Index and Unsupervised Deep Learning [J/OL]. Journal of Hydrology, 2022, 615: 128202 [2025-04-20]. <https://doi.org/10.1016/j.jhydrol.2022.128202>.
- [14] DENG J, DONG W, SOCHER R, et al. ImageNet: A Large-Scale Hierarchical Image Database [C] // 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2009: 248-255.
- [15] HE K, CHEN X, XIE S, et al. Masked Autoencoders Are Scalable Vision Learners [C] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2022: 16000-16009.
- [16] OQUAB M, DARCET T, MOUTAKANNI T, et al. DINOv2: Learning Robust Visual Features without Supervision [DB/OL]. (2023-04-14) [2025-03-15]. <https://arxiv.org/abs/2304.07193>.
- [17] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale [DB/OL]. (2020-10-22) [2025-03-15]. <https://arxiv.org/abs/2010.11929>.
- [18] XIAO T, LIU Y, ZHOU B, et al. Unified Perceptual Parsing for Scene Understanding [C] // Proceedings of the European Conference on Computer Vision. Cham: Springer, 2018: 418-434.
- [19] OSGOUEI P E, SERTEL E, KABADAYI M E. Assessing the Accuracy of the Esa Worldcover 2021 for the Local Region of Lalapasa/Edirne, Turkey and Recommending Possible Accuracy Improvement Strategies [C] // 2023 11th International Conference on Agro-Geoinformatics. Piscataway, NJ, USA: IEEE, 2023: 1-4.
- [20] HUANG L, YOU S, ZHENG M, et al. Green Hierarchical Vision Transformer for Masked Image Modeling [J]. Advances in Neural Information Processing Systems, 2022, 35: 19997-20010.

(责任编辑:刘俏亮)