

基于人工智能模型的小流域沟道漂木识别方法

刘海涛^{1,2}, 陈剑刚¹, 陶紫琴^{1,2}, 李丹丹^{1,2}, 王金水^{1,2}, 王辰元^{1,2}

(1. 中国科学院、水利部成都山地灾害与环境研究所, 山地自然灾害与工程安全重点实验室, 四川成都 610213; 2. 中国科学院大学, 北京 100049)

摘要: [目的] 介绍一种基于人工智能模型的漂木图像分割方法, 为该模型在漂木灾害调查与评估方面的应用提供理论依据。[方法] 选取西藏自治区昌都市贡觉县则巴沟为研究区, 基于人工智能图像分割大模型(segment anything model, SAM), 通过引入轻量级适配器、简化掩码解码器、设计多任务损失函数以及添加辅助分类器, 构建一种针对漂木图像的分割方法(large wood SAM, LWSAM)。训练时冻结原始图像编码器和提示编码器的参数, 以低训练成本提升漂木分割性能, 在构建的漂木相机(LW_CAM_dataset)和无人机(LW_UAV_dataset)两个数据集上对模型进行训练与测试, 并与当前先进图像分割模型进行对比。[结果] ①多任务损失函数能从不同角度优化分割质量, 有效解决了漂木识别中前景稀疏和类别不平衡的问题, 提高了模型对多种漂木形态的适应能力; ②相较于 SAM 方法, 在采用点提示的情况下, LWSAM 在 LW_CAM_dataset 数据集上的 MDice, MIoU 和 F₁ 分数分别提升 15.9%, 15.9% 和 10.0%, 在 LW_UAV_dataset 数据集上的 MDice, MIoU 和 F₁ 分数分别提升 21.6%, 29.6% 和 16.7%; ③漂木分割效果受数据集质量影响, 高质量数据集模型分割结果更好。[结论] 采用 LWSAM 对漂木图像进行分割是可行的, 且在实际运用中表现出较高的精度和较强的鲁棒性, 能够准确分割漂木图像, 可应用于小流域漂木灾害调查。

关键词: 漂木; 泥石流; 分割一切模型; 适配器; 损失函数; 掩码解码器

文献标识码: B

文章编号: 1000-288X(2025)06-0158-11

中图分类号: TP3, X43

文献参数: 刘海涛, 陈剑刚, 陶紫琴, 等. 基于人工智能模型的小流域沟道漂木识别方法[J]. 水土保持通报, 2025, 45(6): 158-168. Liu Haitao, Chen Jiangang, Tao Ziqin, et al. Identification method for large wood in small watershed channels based on segment anything model [J]. Bulletin of Soil and Water Conservation, 2025, 45(6): 158-168.

DOI: 10.13961/j.cnki.stbctb.2025.06.026

CSTR: 32312.14.stbctb.2025.06.026

Identification method for large wood in small watershed channels based on segment anything model

Liu Haitao^{1,2}, Chen Jiangang¹, Tao Ziqin^{1,2}, Li Dandan^{1,2}, Wang Jinshui^{1,2}, Wang Chenyuan^{1,2}

(1. Key Laboratory of Mountain Hazards and Engineering Resilience,

Institute of Mountain Hazards and Environment, Chinese Academy of Sciences,

Chengdu, Sichuan 610213, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: [Objective] An image segmentation method for large wood based on the segment anything model (SAM) was introduced, in order to provide theoretical support for its application in investigating and assessing large wood disasters. [Methods] The Zebagou area in Gongjue County, Chamdo City, Xizang Autonomous Region was selected as the study area, based on SAM, a segmentation method for large wood images—large wood SAM (LWSAM)—was developed by introducing a lightweight adapter, simplifying the mask decoder, designing a multi-task loss function, and adding an auxiliary classifier. During training, the parameters of the original image encoder and the prompt encoder were frozen to improve large wood segmentation performance at a low training cost. The model was trained and tested on two datasets, LW_CAM_dataset and LW_UAV_dataset, and compared with current state-of-the-art image segmentation models. [Results] ① The proposed multi-task loss function could

收稿日期: 2025-06-27

修回日期: 2025-08-18

采用日期: 2025-08-21

资助项目: 国家重点研发计划项目“岩土与生物措施协同的泥石流治理关键技术”(2024YFC3012700); 国家自然科学基金项目“地震易发区山洪泥石流形成演进机制与动态监测预警”(U21A2008)

第一作者: 刘海涛(2001—), 男(汉族), 吉林省松原市人, 硕士研究生, 研究方向为山地灾害防治与工程安全。Email: liuhaitao@imde.ac.cn。

通信作者: 陈剑刚(1982—), 男(汉族), 河北省邢台市人, 博士, 研究员, 主要从事山地灾害防治与工程安全研究。Email: chenjg@imde.ac.cn。

optimize segmentation quality from different perspectives, effectively address the issues of sparse foreground and class imbalance in large wood recognition, and enhance the model's adaptability to various large wood morphologies. ② Compared with the SAM method, under point prompt conditions, LWSAM achieved improvements of 15.9%, 15.9%, and 10.0% in MDice, MIoU, and F1 score, respectively, on the LW_CAM_dataset, and improvements of 21.6%, 29.6%, and 16.7% on the LW_UAV_dataset, respectively. ③ The performance of large wood segmentation was influenced by dataset quality, with models trained on higher-quality datasets achieving better segmentation results. [Conclusion] Using LWSAM for large wood image segmentation is feasible, and it demonstrates high accuracy and strong robustness in practical applications, enabling accurate segmentation of large wood images. This approach can be applied to large wood disaster investigations in small watersheds.

Keywords: large wood; debris flow; segment anything model; adapter; loss function; mask decoder

地震、滑坡、山火等因素导致大量树木被破坏并散落在沟道内形成漂木,漂木被定义为长度大于1 m直径大于10 cm的树木枝干^[1]。漂木灾害是植被覆盖率高的山区的一种常见次生灾害,通常伴随山洪或泥石流的发生而发生,会加剧山洪泥石流运动过程的复杂性^[2-4]。在漂木运动过程中,一方面直接冲击桥梁等建筑物造成破坏,另一方面在障碍物处形成堵塞体产生回水,造成受灾面积扩大^[5]。随后山洪泥石流的持续冲击作用与静水压力增大,不稳定的堵塞体易发生溃决,造成流量的急剧增大,加剧对沟岸及建筑物基础侵蚀和冲刷^[6]。此外,天然沟道内漂木堵塞体往往不止一个,尤其存在于发生山火的沟道内,一旦发生级联溃决,将会造成更严重的致灾效应,增大下游村镇房屋、公路铁路等基础设施的受灾风险。

当前针对漂木的补给、搬运、堵塞、溃决过程开展了大量试验和数值模拟研究,建立了河段尺度漂木补给公式,并针对漂木起动后的运动过程进行了详细研究。Ruiz等^[7]通过野外调查结合数值模拟,提出密度、长度、直径、流体流量是影响漂木搬运能力的重要因素。Pina等^[8]通过水槽试验结合数值模拟,探究单个桥墩处的漂木堵塞概率,得出影响堵塞概率的相应关键参数并提出了预测堵塞概率的公式。Schalko等^[5]开展了一系列水槽试验,认为来流的弗劳德数、细颗粒含量、漂木迎流角度和堆积体形状是影响上游回水高度的重要因素。杨华铨等^[9]通过对四川木里县项脚沟“7·5”特大型泥石流的野外调查研究,认为沟道内漂木造成的泥石流堵溃是造成灾害规模扩大的主要原因。当前研究大多聚焦于漂木的动力过程以及致灾效应,而对灾害发生前漂木的定量评估鲜有研究,导致对沟道内漂木致灾能力的低估。因此,在漂木灾害发生前进行调查,探清其分布与数量,对于漂木灾害的评估和管理具有重要意义。

漂木调查的方法随着技术手段的创新而不断发展,早期的调查方法主要是通过野外调查进行人工测量,获取漂木的大小、数量和分布^[10]。此种方法具

有操作简单的优点,但因其工作量巨大,往往耗时较长。随着科技的发展,通过摄像记录和标记追踪的漂木调查方式逐渐兴起^[11],相对于人工测量,此方法可以进行实时观测,但受限于设备条件,适用范围有限,具有较大误差,且成本较高。上述调查方法均有各自的局限性,针对特定情况下的漂木调查不能普遍适用,因此需要一种能够普适、高效、准确的方法对漂木进行识别与量化。

近年来,计算机视觉大模型的兴起为实现漂木自动化、量化的评估提供了新思路。目前在计算机视觉领域中,已经形成诸多广泛应用的大模型,例如 Mask R-CNN^[12], ViT^[13] (vision transformer), SAM^[14] (segment anything model),这些视觉大模型以其庞大的参数量、丰富的训练数据和优秀的零样本迁移能力被应用到诸多领域。因此,以视觉大模型为基础,通过有限的标注数据对其进行二次开发和参数调优,能够为解决漂木识别问题提供有效的方案。

为实现高效、全面、准确的漂木识别和分割,本研究基于SAM图像分割大模型,通过引入轻量级适配器、简化掩码解码器、设计多任务损失函数以及添加辅助分类器,构建一种针对漂木图像的分割方法(LWSAM),并选用野外沟道采集的漂木图像数据构建漂木数据集对方法进行测试,研究结果可为漂木灾害的评估和管理提供参考。

1 材料与方法

1.1 SAM图像分割模型

SAM是Meta公司在2023年发布的计算机视觉大模型,是计算机视觉实例分割领域第一个通用人工智能模型,使用SA-1B数据集进行分割训练,包含 1.10×10^7 张图像,超过 1.00×10^9 个掩码。SAM的模型架构由图像编码器、提示编码器和掩码解码器3部分组成。图像编码器采用MAE预训练的ViT对输入的图像进行特征编码,提示编码器可以处理不同

类型的提示(点、框、文本),掩码解码器基于Transformer解码器块和动态掩码预测头设计,能接收图像嵌入和提示嵌入,高效生成掩码。其中,SAM模型的零样本迁移能力使模型能通过适当提示输出高质量掩码,从而解决各种下游任务,具有零样本灵活转化新任务、新领域的的能力。

SAM模型的结构及工作流程如图1所示,其任务是通过交互式提示,对输入的图像生成预测掩码。首先,将输入尺寸为 $H_0 \times W_0$ 的图像 I 在图像编码器内被划分为 N 个 $P \times P$ 大小的图像块(patch),每个patch向量化并通过同一个线性层映射到 D 维特征:

$$x_p^i = W_e \text{vec}(I_p^i) + b_e \quad (i = 1 \cdots N) \quad (1)$$

式中: x_p^i 为第 i 个patch在投影后的 D 维特征向量; W_e 为线性投影的权重矩阵;vec为向量化算子; I_p^i 为

原始输入图像被切分后的第 i 个patch; b_e 为线性投影的偏置向量。

在添加可学习的位置编码后,得到初始序列 $X^0 \in \mathbb{R}^{N \times D}$ (\mathbb{R} 代表实数集,右上角标为其维度),随后将 X^0 输入 L 层ViT,每层依次做多头自注意力捕捉全局上下文,再通过前馈网络进行通道间信息融合,最终输出重塑为空间特征图的深度特征张量 $F_{\text{enc}} \in \mathbb{R}^{H \times W \times D}$ 。提示编码器将输入的提示坐标线性映射为与特征图通道一致的提示嵌入,然后与图像特征按位相加,在末尾添加输出令牌,组成提示序列 T^0 。在解码阶段,将图像特征 F_{enc} 与提示序列 T^0 输入掩码解码器,经过若干层交叉注意力Transformer解码单元,提示与图像特征进行自注意力和交叉注意力计算,以实现信息交互与融合,最后根据置信度返回最终掩码。

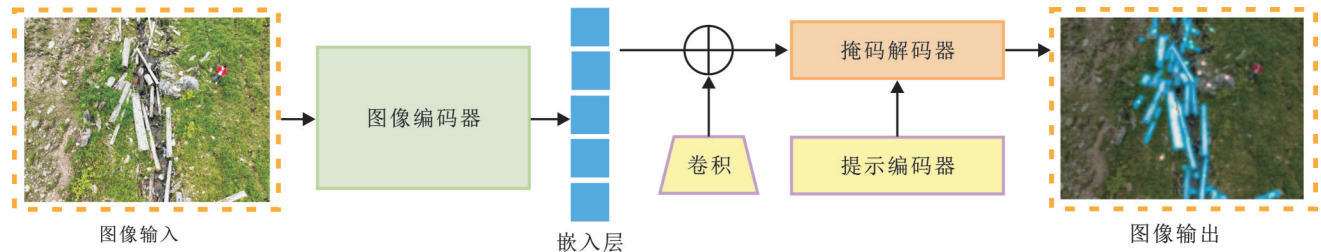


图1 SAM模型结构及工作流程
Fig.1 SAM model structure and workflow

SAM模型在许多自然图像分割任务中表现出色,适用于大多数图像分割任务。但针对特定类别的图像分割任务可能仍无法提供足够精确的分割结果,因此通过对SAM模型进行微调和改进,使其适应特定的任务成为当前研究的热点。目前,基于SAM图像分割大模型的数字图像处理技术已在多个领域广泛应用,在遥感影像处理方面,周洁等^[15]借助SAM模型,实现多尺度标签优化的半监督学习遥感目标检测。在工业检测方面^[16-20],SAM模型主要应用于获取堆石料的颗粒级配曲线和进行工业材料表面缺陷的检测工作。在医学图像分割领域,刘娜等^[21]通过对SAM轻量级微调构建了一种结肠息肉分割方法,刘复昌等^[22]通过对SAM模型的微调实现提示式三维牙齿分割,未来还有望将SAM模型应用于CT扫描、核磁共振等医学图像。

针对漂木图像分割中,由于漂木环境的复杂性超出SAM的预训练经验范围,背景中碎石和泥沙堆积、植被覆盖复杂,目标识别难度增加,且漂木可能相互重叠形态不一,同时存在较严重的样本不均衡,原始的SAM模型同样无法进行精准的分割。此外,漂木图像分割的精度也受采集设备和方式的制约,近景数据中前景占比高,且背景干扰大,光照、角度

条件也不确定,远景数据中存在地形阴影、坡面植被等干扰,进一步影响SAM的特征提取稳定性。因此,需要针对漂木图像的特定需求对SAM模型进行适当改进,使其适用于漂木图像分割任务,提升分割的准确性和鲁棒性。

1.2 漂木识别模型构建

1.2.1 模型结构

以SAM为主干,通过对模型的调整与训练,得到针对漂木分割任务的LWSAM模型,结构如图2所示。主要由图像编码器、适配器、提示编码器和一个分割预测模块组成。选择ViT-B作为模型的基本配置,在训练阶段将SAM图像编码器的所有参数全部冻结,预训练权重在整个训练过程中保持不变,不参与梯度更新,在一定程度上利用其预训练能力,同时减少训练过程中参数更新的计算量。提示编码器采用点和提示框两种方式进行实现,在训练过程中,根据Ground Truth掩码采样出前景点或生成一个提示框,将这些坐标通过线性映射到与特征通道相同的维度,然后与特征图相加,实现提示掩码信息的融合。最后,分割预测模块由掩码解码器和辅助分类器构成,在处理图像和提示信息的融合特征的同时,对网络进行辅助监督,最终输出分割掩码。

1.2.2 适配器结构设计

为更好地适配特定的漂木分割场景,在图像编码器输出后添加一个轻量化的适配器模块,进行降维与升维的线性投影,并保留残差连接。其本质相当于一个简单的多层感知机结构,可以在不破坏原特征的前提下对SAM输出的特征进行后处理调整。输入的漂木图像在经过图像编码器提取特征后输入适配器,首先经过下投影,使用简单的MLP层将给定的嵌入压缩为较低维度,公式为

$$F_d = F_{enc} * W_{down} + b_{down},$$

$$W_{down} \in \mathbb{R}^{1 \times 1 \times \hat{d} \times d}, b_{down} \in \mathbb{R}^d \quad (2)$$

式中: F_d 为经过下投影后输出的特征张量; F_{enc} 为输入到适配器的原始特征张量; W_{down} 为下采样投影的权重矩阵; \mathbb{R} 代表实数集,右上角标为其维度; b_{down} 为下采样投影的偏置向量; \hat{d} 为原始特征维度; d 为瓶颈维度。

再经过GELU(Gaussian error linear unit)函数,为网络引入平滑的非线性,提升模型的表达能力和训练稳定性。GELU的定义为

$$GELU(x) = x \cdot \Phi(x) = x \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \approx$$

$$0.5x \left[1 + \tanh \left(\sqrt{\frac{2}{\pi}} (x + 0.044715x^3) \right) \right] \quad (3)$$

$$F_g = GELU(F_d) \quad (4)$$

式中: F_g 为经过GELU函数后输出的特征张量。

最后将GELU激活后得到的 F_g 使用另一个MLP层将压缩嵌入扩展为其原始维度并进行残差连接得到融合特征,既保留预训练模型得到的通用视觉表示,又融合适配器针对漂木分割任务所做的细微调整,公式为

$$F_u = F_g * W_{up} + b_{up}, W_{up} \in \mathbb{R}^{1 \times 1 \times d \times \hat{d}}, b_{up} \in \mathbb{R}^{\hat{d}} \quad (5)$$

$$F_{pad} = F_{enc} + F_u \quad (6)$$

式中: F_u 为经过上投影后输出的特征张量; W_{up} 为上采样投影的权重矩阵; b_{up} 为上采样投影的偏置向量; F_{pad} 为经过适配器后输出的融合特征。

1.2.3 分割预测模块结构设计

分割预测模块由掩码解码器和辅助分类器构成(图2)。由于SAM的图像编码器中已经采用Transformer架构,可以有效捕捉全局信息,并且LWSAM是针对漂木特定分割任务的模型,使用原始的掩码解码器的收益可能不明显,反而会让训练变得更复杂,因此将LWSAM的掩码解码器简化为一个带有跳跃连接的卷积+上采样的CNN轻量化结构,可以有效地恢复细节并降低模型的复杂度。

在提示信息输入掩码解码器之前,首先与适配器输出的图像特征进行融合:

$$F_{feat} = F_{pad} + T_{mask} \quad (7)$$

式中: F_{feat} 为图像特征与提示信息融合后的特征张量, T_{mask} 为提示信息特征。

随后对输入特征进行 3×3 卷积,降低通道数,经过ReLU激活函数保持空间分辨率不变的同时,提取更多局部特征,在进行上采样之后,重复 3×3 卷积和ReLU扩大空间分辨率。为保留更多原始特征信息,使网络更好地恢复细节,解码器会对最初的输入特征做 1×1 卷积上采样,并通过跳跃连接与当前分辨率的特征相加,将相加后的特征再进行上采样,使用 1×1 卷积调整通道数。最后将掩码插值回原图大小,得到与输入图像等尺寸的最终掩码。

LWSAM的辅助分类器是一个与掩码解码器并列的结构,在接收图像和提示信息的融合特征后,首先对特征图做一次 3×3 卷积,降低通道数降并ReLU激活,随后对整张特征图进行平均池化,提取全局信息。最后将全局特征向量展平后通过全连接层输出。在训练过程中,掩码解码器和辅助分类器相互协同,分割损失梯度与分类损失梯度在共享特征空间中进行联合优化,实现多任务学习。分类任务驱动网络学习全局语义一致的特征表示,通过残差连接影响像素级预测,使得局部分割决策与全局语义理解保持一致,帮助提升像素级分割性能,实现多尺度监督互补,在局部和全局方面优化模型参数。

1.2.4 损失函数

在语义分割任务中,常见的损失函数通常围绕像素级的准确率、重叠度、边界匹配等指标展开。然而,针对特定的下游场景或多任务需求,单一的分割损失往往难以充分挖掘网络潜力。为此,许多工作将多种损失函数组合使用,既能提高网络对目标概念的区分能力,也能在一定程度上改善分割的精确度和鲁棒性。本研究针对漂木分割任务设计一种多任务损失函数:

$$L = L_{seg} + \alpha \cdot L_{cls} \quad (8)$$

式中: L_{seg} 为分割损失; L_{cls} 为辅助分类损失; α 为权重系数,设置为0.5。

(1) 分割损失。分割损失由Dice损失函数(Dice Loss)和IoU损失函数(IoU Loss)组成。

Dice系数本质上是衡量预测区域与真实区域重叠程度的指标,取值范围 $0 \sim 1$ 。预测和真实标签完全重合,则Dice系数达到1;如果二者完全不重叠,则Dice系数为0。Dice损失函数^[23]通过最大化预测结果和真实标签的重叠区域,有效地提升模型在处理不平衡数据时的表现。Dice损失函数的计算公式为:

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N (p_i \cdot g_i)}{\sum_{i=1}^N p_i + \sum_{i=1}^N g_i + \epsilon} \quad (9)$$

式中: N 表示像素总数; p_i 为第 i 个像素的预测值; g_i 为第 i 个像素的真实标签; ϵ 是一个很小的平滑项,防止分母出现零。

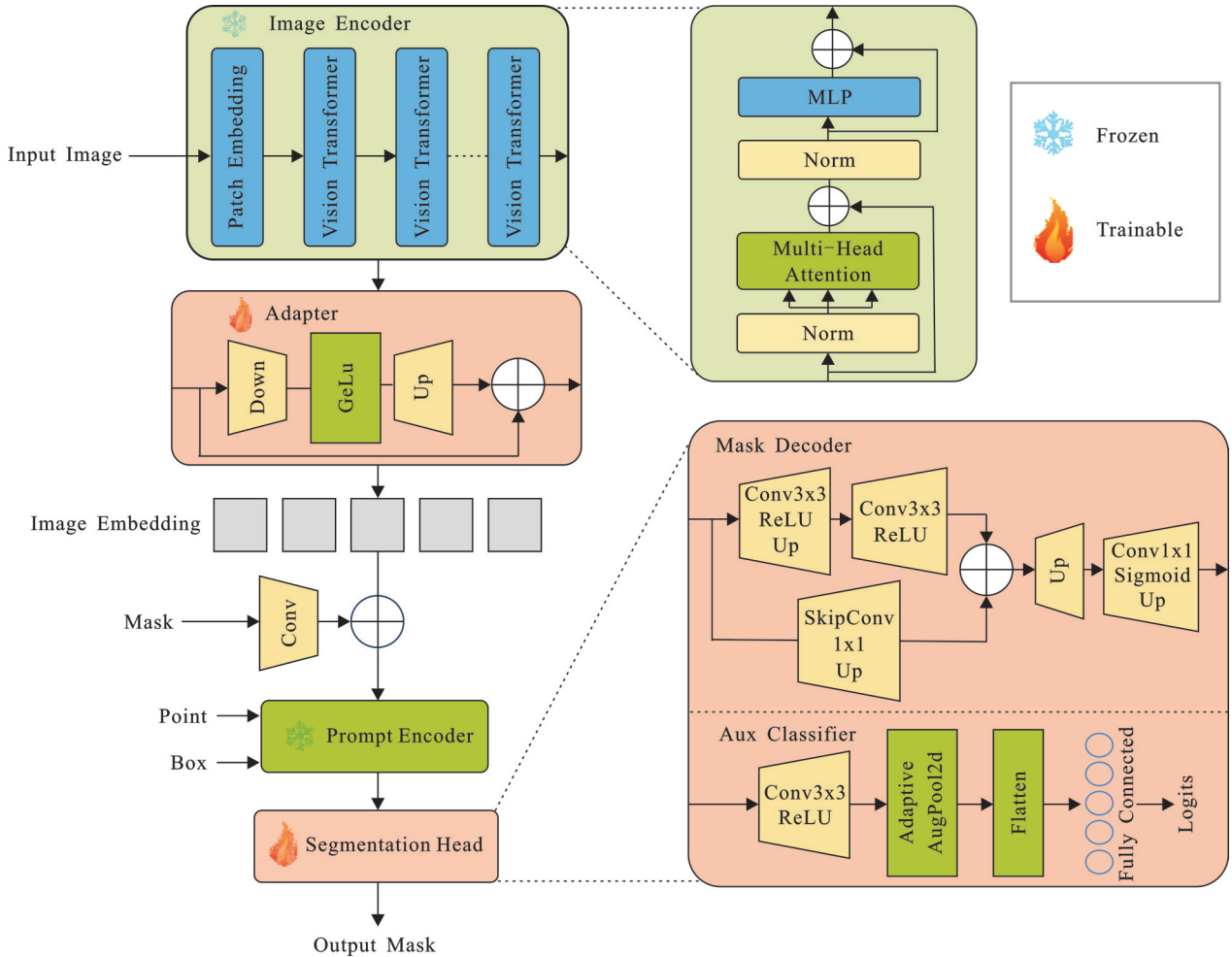


图 2 LWSAM 模型结构

Fig.2 LWSAM model structure

IoU 损失函数^[24]是应用于图像分割任务基于区域交并比的损失函数,交集表示预测前景和真实前景,同时为前景的像素部分集合,即模型正确预测漂木的区域。并集则表示预测前景和真实前景至少有一个为前景的像素部分集合,即模型预测为漂木区域与真实漂木区域的总和。IoU 损失函数的计算公式为

$$L_{\text{IoU}} = 1 - \frac{\sum_{i=1}^N (p_i \cdot g_i)}{\sum_{i=1}^N p_i + \sum_{i=1}^N g_i - \sum_{i=1}^N (p_i \cdot g_i) + \epsilon} \quad (10)$$

式中: N 表示像素总数; p_i 为第 i 个像素的预测值; g_i 为第 i 个像素的真实标签; ϵ 是一个很小的平滑项,防止分母出现零。

分割损失使用 Dice Loss 和 IoU Loss 结合,综合 Dice 对小目标敏感、对不平衡鲁棒的优势,以及 IoU 对全局区域覆盖更直接,与通用评估指标一致的特点,二者互补能够从不同角度衡量预测掩码与真实掩码的重叠程度,使网络在前景稀疏和不平衡场景中也能保持稳定,提升分割质量。

(2) 辅助分类损失。本研究在传统的分割网络基础上,增加辅助分类器对整张图像进行二分类,帮助网络更快捕捉到类间差异,从而提升分割精度。使用交叉熵损失函数^[25](cross entropy loss)作为模型辅助分类函数,引导网络在特征层面区分漂木,其针对二分类问题的计算公式为

$$L_{\text{CE}} = -\frac{1}{N} \sum_{i=1}^N [g_i \ln(p_i) + (1 - g_i) \ln(1 - p_i)] \quad (11)$$

式中: N 表示像素总数; g_i 为第 i 个像素的真实标签; p_i 为其对应的预测概率值。

因此本模型的联合损失函数表达式为

$$L = L_{\text{seg}} + \alpha \cdot L_{\text{cls}} = L_{\text{Dice}} + L_{\text{IoU}} + \alpha \cdot L_{\text{CE}} \quad (12)$$

1.3 试验验证

1.3.1 数据描述

选取西藏自治区昌都市贡觉县则巴沟为研究区(图3),则巴沟主沟全长7.96 km,沟道内有大量漂木,存在沟道堵塞溃决隐患。则巴沟漂木呈细长圆柱形,边界较为平整,颜色从浅灰到深棕色变化,表面粗糙导致光反射不均匀,存在较严重的阴影和光照干扰,空间分布从单根独立到多根交错重叠均有

存在。本研究数据分为两部分,分别为由大疆无人机(DJI Mavic 3)以固定距离和角度对漂木拍摄的正射影像和由尼康相机(NIKOND810)以随机距离和角度对漂木采集的图像。通过水平翻转、90度旋转、裁剪和色彩抖动的数据增强方式建立漂木的相机数据集(LW_CAM_dataset)和无人机数据集(LW_UAV_dataset),共包含图像1380张,增强后样本的类别分布、形态多样性保持原始数据的特征,包含不同自然条件下的漂木分布场景。所有采集的影像数据均为高分辨率(无人机:5280×3956,相机:7360×4912)的RGB彩色图像,采集日期均为2024年8月10日。

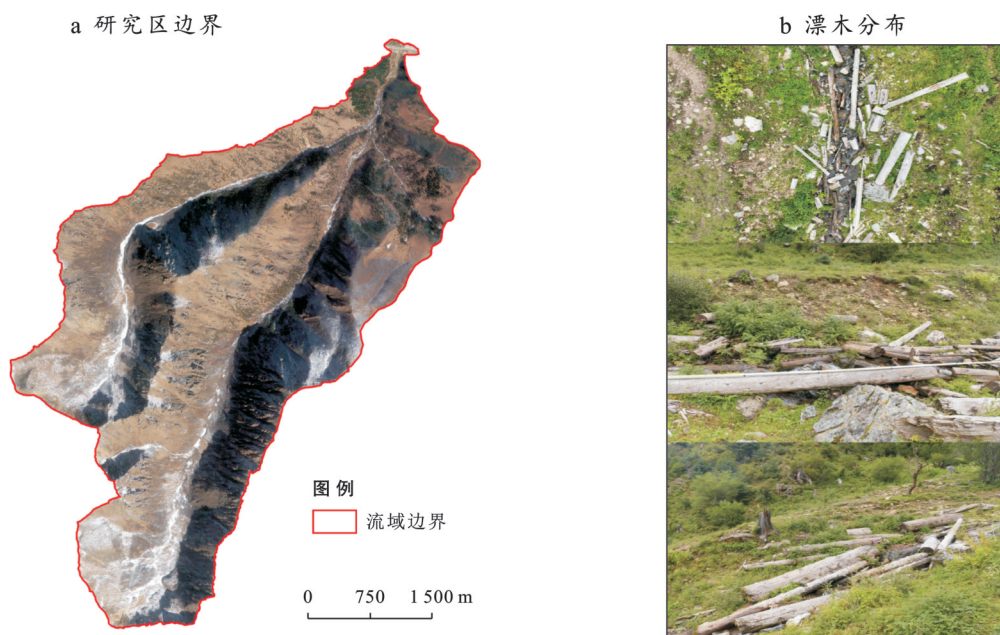


图3 则巴沟边界和漂木分布

Fig.3 Boundary of Zeba gully and distribution of large wood

1.3.2 试验设置

试验在Ubuntu 20.04操作系统环境下进行,深度学习框架PyTorch版本为2.4.1,Python版本为3.8。计算环境采用NVIDIA GeForce RTX 4090 D GPU,24 GB内存,并使用CUDA 11.8进行加速。在模型训练中选择Adam优化器,学习率设定为0.000 01, batch size设定为8, epoch设定为100。

为评估模型分割漂木任务的可行性和有效性,在自定义漂木数据集上对模型进行训练,训练集和测试集的比例为8:2。在训练过程中,所有图像均进行归一化处理,保持原始纵横比不变的情况下调整输入尺寸为1024×1024,边缘区域不足1024像素的填充黑色像素。为保证试验的可重复性,设置固定的随机种子。模型训练完成后,在测试集上进行

相应的定量精度评价。

1.3.3 评价指标

采用平均骰子系数(mean dice)、平均交并比(mean IoU)、 F_1 分数(F_1 score)作为评价指标来评估模型性能,计算公式见表1。

2 结果与分析

2.1 不同模型对比试验

选取当前图像分割领域主流的UNet^[26], TransUNet^[27], DeepLabv3⁺^[28], SAM, FastSAM^[29]模型与本研究改进后的LWSAM模型进行对比试验,为保证公平性,所有模型采用与LWSAM相同的冻结策略,在LW_CAM_dataset和LW_UAV_dataset上进行重新训练与测试。

表 1 模型性能评价指标

Table 1 Model performance evaluation metrics

评价指标	计算公式
平均骰子系数	$MDice = \frac{1}{n} \sum (Dice_i)$ $Dice_i = \frac{2TP_i}{2TP_i + FP_i + FN_i}$
平均交并比	$MIoU = \frac{1}{n} \sum (IoU_i)$ $IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i}$
F ₁ 分数	$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$ $Precision = \frac{TP}{TP + FP}$ $Recall = \frac{TP}{TP + FN}$

注: TP表示真正类,即正确分割漂木像素数目; FN表示假负类,即错误分割非漂木像素数目; FP表示假正类,即错误分割漂木像素数目。

由试验结果(表2)可知,对于SAM类的方法,无论采用Point还是Box的提示信息,LWSAM模型在LW_CAM_dataset和LW_UAV_dataset数据集上的分割性能均较其他模型均有大幅提升。与FastSAM方法相比,在采用点提示的条件下,在LW_CAM_dataset上的MDice, MIoU和F₁分数分别提升54.7%, 70.4%和35.0%,在LW_UAV_dataset上的MDice, MIoU和F₁分数分别提升69.0%, 101.9%和44.0%;在采用提示框提示的条件下,在LW_CAM_dataset上的MDice, MIoU和F₁分数分别提升94.6%, 125.3%和27.0%,在LW_UAV_dataset上的MDice, MIoU和F₁分数分别提升90.0%, 133.3%和68.1%。与原始的SAM方法相比,在采用点提示的条件下,在LW_CAM_dataset上的MDice, MIoU和F₁分数分别提升15.9%, 15.9%和10.0%,在LW_UAV_dataset上的MDice, MIoU和F₁分数分别提升21.6%, 29.6%和16.7%;在采用提示框提示的条件下,在LW_CAM_dataset上的MDice, MIoU和F₁分数分别提升3.4%, 0.3%和1.3%,在LW_UAV_dataset上的MDice, MIoU和F₁分数分别提升24.3%, 28.8%和19.1%。

对于当前图像分割领域主流的非SAM类的方法,LWSAM方法同样展现出卓越的分割性能,在采用点提示的条件下将LWSAM方法与UNet, TransUNet和DeepLabv3⁺方法进行对比,结果表明

在LW_CAM_dataset上MDice分别提升3.3%, 3.3%和8.2%, MIoU分别提升5.5%, 4.8%和12.9%, F₁分数分别提升3.9%, 3.9%和8.9%;在LW_UAV_dataset上MDice分别提升5.3%, 5.3%和12.6%, MIoU分别提升7.7%, 7.7%和19.2%, F₁分数分别提升5.9%, 5.9%和13.3%,表明LWSAM方法可以较好的提升SAM的分割能力。

表 2 不同模型对比试验结果

Table 2 Comparative experiment results of different models

数据源	模型	提示信息	MDice	MIoU	F ₁
LW_CAM_dataset	LWSAM	Point	0.789	0.656	0.794
		Box	0.790	0.658	0.795
	FastSAM	Point	0.510	0.385	0.588
		Box	0.406	0.292	0.450
	SAM	Point	0.681	0.566	0.722
		Box	0.764	0.656	0.785
	UNet		0.764	0.622	0.764
	TransUNet		0.764	0.626	0.764
	DeepLabv3 ⁺		0.729	0.581	0.729
	LW_UAV_dataset	LWSAM	Point	0.840	0.727
Box			0.840	0.728	0.842
FastSAM		Point	0.497	0.360	0.587
		Box	0.442	0.312	0.501
SAM		Point	0.691	0.561	0.724
		Box	0.676	0.565	0.707
UNet			0.798	0.675	0.798
TransUNet			0.798	0.675	0.798
DeepLabv3 ⁺			0.746	0.610	0.746

注:加粗字体表示最优结果。

综合来看,LWSAM模型在LW_CAM_dataset和LW_UAV_dataset数据集上的分割性能均优于其他模型,但相比于LW_CAM_dataset数据集,在LW_UAV_dataset数据集上的分割效果要更好,这是由于自然界内漂木的几何形态通常是细长的,且一张照片内的前景数目通常小于背景数目,背景对模型的干扰较大。LW_UAV_dataset数据集固定拍摄角度和距离,使得图像内的漂木可以较为清晰的展现。而LW_CAM_dataset数据集由于拍摄的角度和距离不同,造成光照和背景噪声的干扰,增加分割任务的难度,因此分割效果略差于LW_UAV_dataset数据集。同时LWSAM的结构在轻量化设计上具有明显优势。LWSAM的可训练参数仅有4.77 M,相比之下,SAM, FastSAM, UNet, TransUNet和DeepLabv3⁺的可训练参数分别为358.40, 68.00,

31.04, 105.28, 163.04 M, 模型的参数量和训练成本大幅降低;在计算复杂度方面, LWSAM 的 FLOPs 为 60.40 G, 而原始的 SAM 模型 FLOPs 高达 210.0 G, 表明 LWSAM 的计算复杂度降低约 70%, 能够更快地进行推理;在推理时间方面, LWSAM 的平均推理时间为 0.489 s, 而 SAM 的平均推理时间为 0.592 s, 推理时间大幅缩短, 响应速度更快, 表明 LWSAM 的轻量化设计可减少训练成本, 能够在资源有限的环境中应用。

2.2 可视化结果

选取 LWSAM 和 SAM 模型分别在 LW_CAM_dataset 和 LW_UAV_dataset 数据集上进行测试。由图 4 和图 5 可知, 在 LW_CAM_dataset 上对于单根漂木且背景干扰相对较小的数据, 无论采用点提示还是提示框提示的方法, LWSAM 和 SAM 都能够生成较精确的分割结果, 能够有效的将漂木从图像中分割出来。但是在多根漂木交错重叠且背景干扰较大的情况下, LWSAM 的点提示方法生成的掩码质量明显高于 SAM 采用相同提示方法生成的掩码质量, 在此情况下, SAM 方法在一次点击的条件下存在明显的目标遗漏, 不能完整地分割图片中所有的漂木, 而 LWSAM 方法能够分割出图像中大多数漂木, 且分割掩码接近于真实掩码。

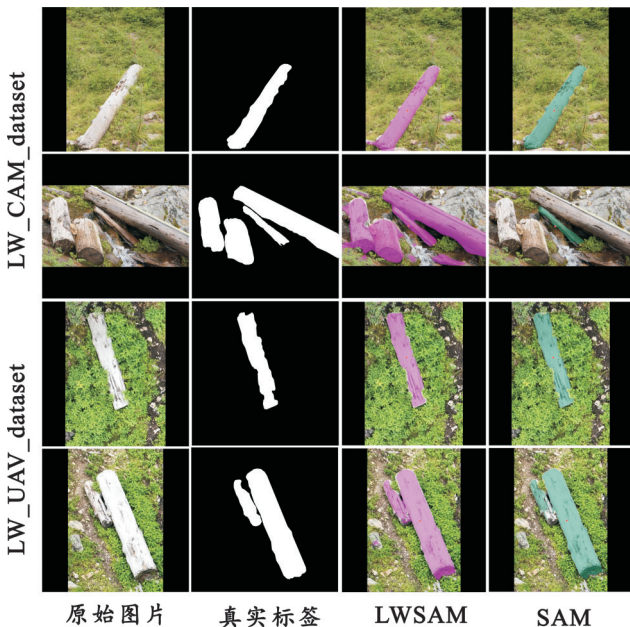


图 4 2 种模型对不同数据集的点提示预测结果对比
Fig.4 Comparison of prediction results of two models on different datasets by point prompt

在采用提示框方法的条件下, SAM 方法虽然能够分割出图片中更多的漂木, 但分割质量相较于 LWSAM 方法大大降低, 错误分割图中的较多背景

部分, 表明 SAM 模型在受到背景噪点的干扰时, 分割结果存在较大偏差。

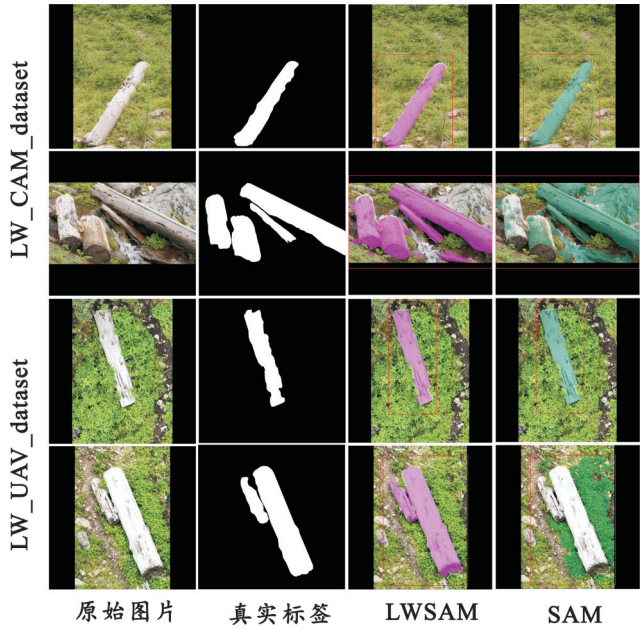


图 5 2 种模型对不同数据集的提示框提示预测结果对比
Fig.5 Comparison of prediction results of two models on different datasets by box prompt

对于 LW_UAV_dataset 数据集, LWSAM 模型的表现整体上仍优于 SAM 模型, 对于采用点提示的单根漂木, 2 个模型均能正确分割出前景目标, 但在一次点击的多根漂木条件下, LWSAM 模型能够正确分割出图片中所有目标, 而 SAM 模型对于多个目标的分割精度不高, 边界不够清晰。对于采用提示框提示的单根漂木, 2 个模型仍能正确分割出正确漂木, 但对于多个目标 LWSAM 相对于 SAM 模型展现出较好的优越性, LWSAM 模型可以正确地分割出图像中所有漂木且分割掩码位置和形状接近于真实掩码, 并能够有效处理复杂结构与背景的干扰, 相比之下, SAM 模型未能正确分割出图像中的漂木目标。

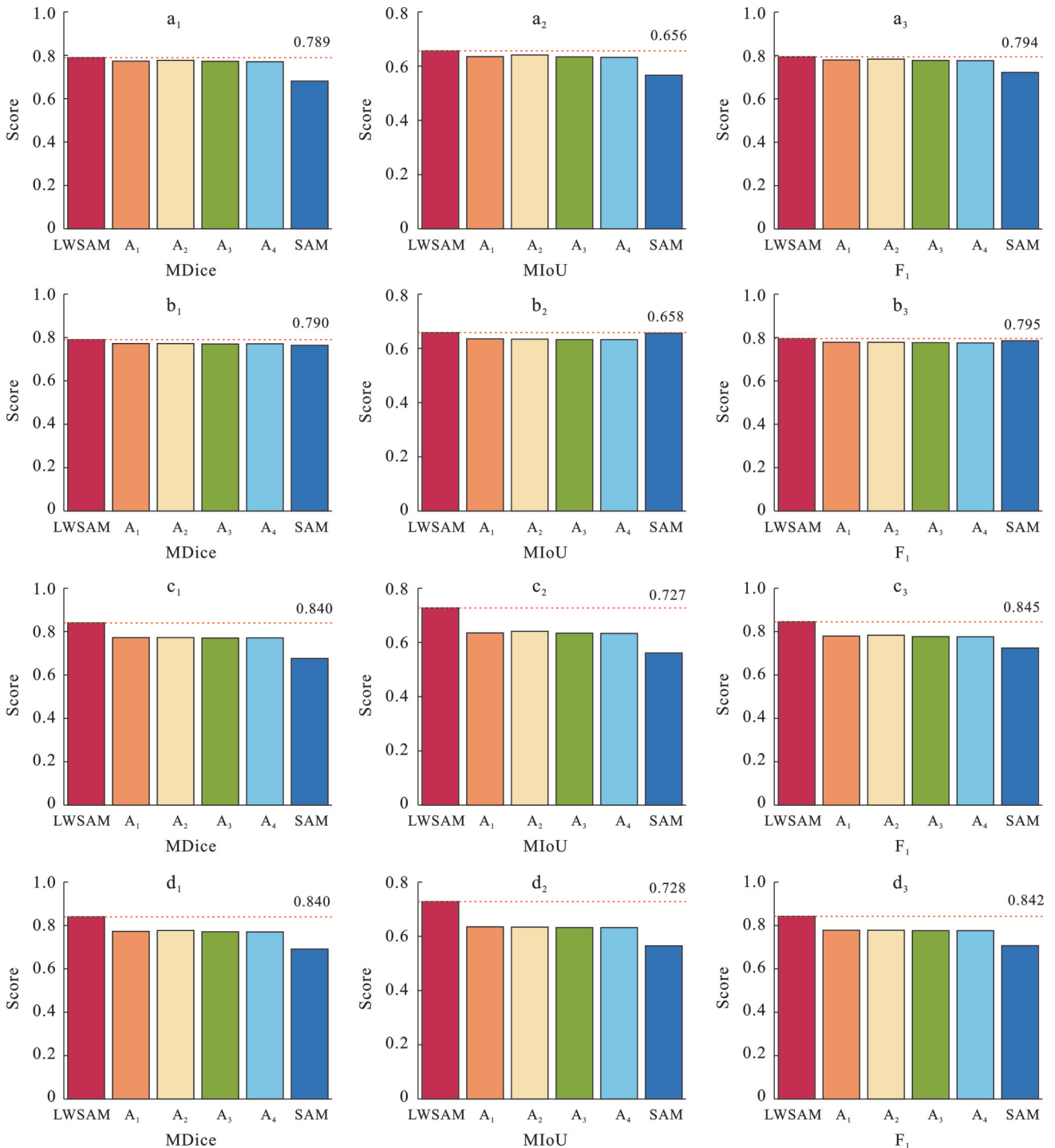
3 讨论

为提高 SAM 模型针对漂木的分割能力, LWSAM 模型引入适配器、辅助分类器、分类损失和掩码解码器的跳跃连接结构, 对比试验与图像分割的可视化结果已证明 LWSAM 模型对于漂木图像分割的优越性, 通过消融试验可评估模型各结构对 SAM 分割效果的影响。

图 6a 为 LW_CAM_dataset 数据集上的点提示消融试验结果, 图 6b 为 LW_CAM_dataset 数据集上的提示框提示消融试验结果, 图 6c 为 LW_UAV_dataset 数据集上的点提示消融试验结果, 图 6d 为

LW_UAV_dataset 数据集上的提示框提示消融试验结果,其中 SAM 表示原始 SAM 模型,LWSAM 为本

文方法,A₁-A₄ 分别为去除适配器、辅助分类器、分类损失和跳跃连接后的模型。



注:分图a—b分别为LW_CAM_dataset上的点、框提示消融试验结果;c—d分别为LW_UAV_dataset数据集上的点、框提示消融试验结果。其中SAM表示原始SAM模型,LWSAM为本文方法,A₁-A₄分别表示去除适配器、辅助分类器、分类损失和跳跃连接后的模型。

图 6 不同组件消融试验结果
Fig.6 Ablation experiment on different components

由消融试验结果可知,LWSAM方法整体上显著优于SAM方法,与对比试验结果契合,同时在去除相应模块后评价指标整体趋势仍优于SAM方法,表明改进后的方法的有效性和对于漂木分割的准确

性,因此引入适配器、辅助分类器、分类损失和跳跃连接结构能够使得算法更加适用于漂木分割场景。从图6可以看出,适配器是最关键的组件,其移除导致性能下降最明显,说明针对漂木特征的适应性调

整至关重要。同时,在 LW_UAV_dataset 数据集上加入相应模块后,模型评价指标的提升率高于 LW_CAM_dataset,表明模型对于数据质量的敏感性,固定角度和距离、少背景噪点的数据会进一步提高模型的分割精度。

综上所述,LWSAM 模型相对于其他模型在漂木分割任务中展现出优越性,受到背景噪点的干扰较小,分割精度更高,分割结果更加完整,呈现出在漂木灾害防治中的潜力。

随着极端天气和地质灾害的增多,漂木灾害频发,严重威胁沿线居民的生命财产安全,对漂木灾害的防灾减灾要求也愈发严峻。尽管本研究在小流域漂木识别模型方面进行初步探索,但仍存在一定局限性。未来可不断丰富漂木图像样本数据库,提升模型的识别精度和泛化能力;同时将模型进行更深入的轻量化设计,提升模型的可操作性,使其能够适应不同运行环境。流域内漂木的自动识别有助于区域尺度的漂木灾害风险评估,可为防灾减灾提供技术支撑。

4 结论

针对漂木的自动化识别问题,本研究提出了一种基于 SAM 方法的漂木图像分割模型(LWSAM)。通过冻结图像编码器并添加轻量级适配器,在保留预训练能力的同时大幅减少计算成本。简化的掩码解码器和跳跃连接结构使模型能够高效恢复漂木细节信息,适合在资源有限的环境中应用。通过 Dice 损失、IoU 损失和分类损失的组合,从不同角度优化分割质量,有效解决漂木识别中前景稀疏和类别不平衡的问题,提高模型对多种漂木形态的适应能力。在 LW_CAM_dataset 和 LW_UAV_dataset 数据集上的试验结果表明,LWSAM 方法能够高效全面地分割漂木图像,与当前主流 SAM 类和非 SAM 类方法的对比均表现出明显优势,消融试验进一步证明模型结构设计的有效性。试验结果表明,模型分割效果与数据质量有关,在 LW_UAV_dataset 固定角度和距离拍摄的数据集上,模型效果明显优于随机角度和距离的 LW_CAM_dataset。

参考文献(References)

- [1] Comiti F, Lucia A, Rickenmann D. Large wood recruitment and transport during large floods: A review [J]. *Geomorphology*, 2016, 269: 23-39.
- [2] 陈剑刚,费高高,王喜安,等.漂木对山洪泥石流运动致灾影响研究进展[J]. *水利水电科技进展*, 2022, 42(3): 104-111.
Chen Jiangan, Fei Gaogao, Wang Xi'an, et al. Advances on disaster effects of drift wood in flash flood debris flows [J]. *Advances in Science and Technology of Water Resources*, 2022, 42(3): 104-111.
- [3] Fei Gaogao, Wang Xiekang. A review of large wood dynamics relevant to hazard characteristics for built structures [J]. *Geomorphology*, 2024, 453: 109152.
- [4] Chen Jiangan, Liu Wenrun, Zhao Wanyu, et al. Magnitude amplification of flash floods caused by large woody in Keze gully in Jiuzhaigou National Park, China [J]. *Geomatics, Natural Hazards and Risk*, 2021, 12(1): 2277-2299.
- [5] Schalko I, Follett E, Nepf H. Impact of lateral gap on flow distribution, backwater rise, and turbulence generated by a logjam [J]. *Water Resources Research*, 2023, 59(10): e2023WR034689.
- [6] Schalko I, Lageder C, Schmocker L, et al. Laboratory flume experiments on the formation of spanwise large wood accumulations: Part II. Effect on local scour [J]. *Water Resources Research*, 2019, 55(6): 4871-4885.
- [7] Ruiz-Villanueva V, Piégay H, Gaertner V, et al. Wood density and moisture sorption and its influence on large wood mobility in rivers [J]. *Catena*, 2016, 140: 182-194.
- [8] De Cicco P N, Paris E, Solari L, et al. Bridge pier shape influence on wood accumulation: Outcomes from flume experiments and numerical modelling [J]. *Journal of Flood Risk Management*, 2020, 13(2): e12599.
- [9] 杨华铨,柳金峰,孙昊,等.四川木里县项脚沟“7·5”特大型泥石流特征及发展趋势分析[J]. *中国地质灾害与防治学报*, 2024, 35(1): 100-107.
Yang Huaquan, Liu Jinfeng, Sun Hao, et al. Analysis of the characteristics and development trends of the “7·5” catastrophic debris flow in Xiangjiao gully, Muli County, Sichuan [J]. *The Chinese Journal of Geological Hazard and Control*, 2024, 35(1): 100-107.
- [10] May C L, Gresswell R E. Processes and rates of sediment and wood accumulation in headwater streams of the Oregon Coast Range, USA [J]. *Earth Surface Processes and Landforms*, 2003, 28(4): 409-424.
- [11] MacVicar B, Piégay H. Implementation and validation of video monitoring for wood budgeting in a wandering piedmont river, the Ain River (France) [J]. *Earth Surface Processes and Landforms*, 2012, 37(12): 1272-1289.
- [12] He Kaiming, Gkioxari G, Dollár P, et al. Mask R-CNN [C] // 2017 IEEE International Conference on Computer Vision (ICCV). October 22-29, 2017, Venice, Italy. IEEE, 2017: 2980-2988.
- [13] Han Kai, Wang Yunhe, Chen Hanting, et al. A survey on vision transformer [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(1): 87-110.

- [14] Kirillov A, Mintun E, Ravi N, et al. Segment anything [C] // 2023 IEEE/CVF International Conference on Computer Vision (ICCV). 2023, Paris, France. IEEE, 2024:3992-4003.
- [15] 周洁, 方振宇. 基于 SAM 多尺度标签优化的半监督学习遥感目标检测[J/OL]. 微电子学与计算机, 2024:1-10. (2024-12-24) [2025-06-27]. <https://kns.cnki.net/kcms/detail/61.1123.TN.20241224.0839.002.html>.
Zhou Jie, Fang Zhenyu. Semi-supervised learning remote sensing target detection based on SAM multi-scale label optimization [J/OL]. *Microelectronics & Computer*, 2024: 1-10. (2024-12-24) [2025-06-27]. <https://kns.cnki.net/kcms/detail/61.1123.TN.20241224.0839.002.html>.
- [16] 张振伟, 蔡可天, 高轩, 等. 基于 SAM 图像处理的堆石料级配计算方法及验证[J]. 水力发电, 2025, 51(2):80-86.
Zhang Zhenwei, Cai Ketian, Gao Xuan, et al. Calculation method and verification of rockfill gradation based on SAM image processing [J]. *Water Power*, 2025, 51(2):80-86.
- [17] 张鸿, 杨俊雅, 刘可心, 等. 基于 Stone-SAM 的便携式粗集料级配智能检测[J]. 建筑材料学报, 2025, 28(6):581-590.
Zhang Hong, Yang Junya, Liu Kexin, et al. Portable intelligent detection of coarse aggregate gradation based on stone-SAM [J]. *Journal of Building Materials*, 2025, 28(6):581-590.
- [18] 付立群, 金峰, 张喜喜, 等. 结合 Mask R-CNN 和 SAM 获取堆石混凝土坝堆石级配曲线[J]. 水电能源科学, 2024, 42(11):7-11.
Fu Liqun, Jin Feng, Zhang Xixi, et al. Obtaining particle size distribution curves for rock-filled concrete dams by combining mask R-CNN and SAM [J]. *Water Resources and Power*, 2024, 42(11):7-11.
- [19] 马小川, 付佳, 王李廷煜, 等. SAM 特征引导的主动学习在缺陷检测中的应用[J]. 电子机械工程, 2025, 41(3):80-86.
Ma Xiaochuan, Fu Jia, Wang L, et al. Application of active learning to defect detection guided by SAM feature [J]. *Electro-Mechanical Engineering*, 2025, 41(3):80-86.
- [20] 陶攀, 方宇, 王欣, 等. 基于改进 SAM 模型的多任务轨道缺陷检测方法[J]. 南京大学学报(自然科学), 2024, 60(5):776-784.
Tao Pan, Fang Yu, Wang Xin, et al. Multi-task track defect detection method based on improved SAM model [J]. *Journal of Nanjing University (Natural Sciences)*, 2024, 60(5):776-784.
- [21] 刘娜, 封筠, 霍一儒, 等. SAMCP: 一种轻量级微调 SAM 的结肠息肉分割方法[J/OL]. 计算机应用, 2025:1-14. (2025-02-27). <https://kns.cnki.net/kcms/detail/51.1307.TP.20250227.1119.002.html>.
Liu Na, Feng Jun, Huo Yiru, et al. SAMCP: Lightweight SAM fine-tuning method for colon polyp segmentation [J/OL]. *Journal of Computer Applications*, 2025:1-14. (2025-02-27). <https://kns.cnki.net/kcms/detail/51.1307.TP.20250227.1119.002.html>.
- [22] 刘复昌, 蔡煜晨, 缪永伟, 等. 基于预训练 SAM 的提示式三维牙齿分割方法[J]. 浙江大学学报(理学版), 2025, 52(1):59-69.
Liu Fuchang, Cai Yuchen, Miao Yongwei, et al. Prompt-based three-dimensional tooth segmentation method based on pre-trained SAM [J]. *Journal of Zhejiang University (Science Edition)*, 2025, 52(1):59-69.
- [23] Li Xiaoya, Sun Xiaofei, Meng Yuxian, et al. Dice loss for data-imbalanced NLP tasks [C] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online. Stroudsburg, PA, USA: ACL, 2020:465-476.
- [24] Rahman M A, Wang Yang. Optimizing intersection-over-union in deep neural networks for image segmentation [C] // Advances in Visual Computing. Cham: Springer, 2016:234-244.
- [25] Mannor S, Peleg D, Rubinstein R. The cross entropy method for classification [C] // Proceedings of the 22nd International Conference on Machine Learning. 2005, Bonn, Germany. ACM, 2005:561-568.
- [26] 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.
- [27] Castro R, Ramos L, Román S, et al. U-net vs. transunet: performance comparison in medical image segmentation [C] // International Conference on Applied Technologies. Cham: Springer Nature Switzerland, 2022:212-226.
- [28] Chen L C, Zhu Yukun, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C] // Computer Vision-ECCV 2018. Cham: Springer, 2018:833-851.
- [29] Vaka I R, Sundharakumar K B. Comparative Analysis for SAM, FastSAM, EfficientSAM, Detectron 2 for Semantic Segmentation in Self Driving Cars [C] // International Conference on Computer Vision and Image Processing. Cham: Springer Nature Switzerland, 2024:281-294.