



www.sciengine.com/CSD

ISSN 2096-2223

CN 11-6035/N



文献 CSTR:

32001.14.11-6035.ncdc.2025.0112.zh



文献 DOI:

10.11922/11-6035.ncdc.2025.0112.zh

数据 DOI:

10.57760/sciencedb.j00001.01517

文献分类: 地球科学

收稿日期: 2025-06-04

录用日期: 2026-01-09

发表日期: 2026-02-28

2023 年青藏高原冰湖编目数据集

田志娟¹, 姚晓军^{1,2*}, 常辉¹, 段红玉^{1,3}

- 1. 西北师范大学地理与环境科学学院, 兰州 730070
- 2. 甘肃省绿洲资源环境与可持续发展重点实验室, 兰州 730070
- 3. 甘肃煤田地质局一四九队, 兰州 730020

摘要: 冰湖是冰川消融过程中形成的典型水文地貌单元, 近年来在全球高山和极地地区呈现快速扩张趋势。冰湖编目通过系统记录冰湖的空间位置、几何参数等信息, 对理解冰川演化过程、气候变化响应、淡水资源评估及冰湖溃决洪水风险预警具有科学价值。针对冰湖潜在分布区划定尚存争议、冰湖遥感提取依赖人工目视解译等挑战, 本研究提出了一种融合山系单元和冰川边界的冰湖潜在分布区界定新方法, 基于 Sentinel-2 MSI 遥感影像和 Attention DeepLab V3+ 深度学习模型, 实现了青藏高原全域高精度冰湖自动化提取与系统编目。结果表明, 2023 年青藏高原面积 $\geq 0.005 \text{ km}^2$ 的冰湖共 29089 个, 总面积 2020.65 km^2 。冰湖主要分布于东南部的念青唐古拉山、东喜马拉雅山和横断山。在规模结构上, 小型冰湖 ($0.01 - 0.05 \text{ km}^2$) 数量占比最高 (62.66%), 但面积贡献仅 20.39%; 大型冰湖 ($>1 \text{ km}^2$) 数量稀少 (0.68%) 却占总面积的 24.22%。在类型上, 以非冰川接触湖为主 (73.88%), 冰碛湖次之 (24.27%), 冰面湖最少 (1.85%)。由遥感影像空间分辨率造成的冰湖面积误差为 134.88 km^2 (6.68%)。本数据集全面反映了青藏高原冰湖资源现状, 为该地区冰湖溃决风险评估、水电站建设等研究提供了数据支撑。

关键词: 冰湖编目; Sentinel; 深度学习; 青藏高原

数据库(集)基本信息简介

数据库(集)名称	2023 年青藏高原冰湖编目数据集
数据作者	田志娟, 姚晓军, 常辉, 段红玉
数据通信作者	姚晓军 (yaoxj_nwnu@163.com)
数据时间范围	2023年
地理区域	青藏高原 (25°N–40°N, 74°E–104°E)
空间分辨率	10 m
数据量	81.71 MB
数据格式	ESRI Shapefile (压缩为*.zip)
数据服务系统网址	https://doi.org/10.57760/sciencedb.j00001.01517 https://www.ncdc.ac.cn/portal/metadata/fe4fb623-3ffe-44a8-a4c9-4f6346f972ec

* 论文通信作者

姚晓军: yaoxj_nwnu@163.com

基金项目	西北师范大学2023年度研究生科研资助项目（2023KYZZ-S181）；绿洲科学战略行动计划（NWNNU-LZKX-202301）。
数据库（集）组成	本数据集包含1个文件：2023年青藏高原冰湖编目数据集，命名为：Tibetan_Plateau_glacial_lakes_inventory_2023，冰湖（ $\geq 0.005 \text{ km}^2$ ）记录数量29089个。

引言

作为冰川动力过程的直接产物，冰湖的形成与演化与冰川变化密切相关^[1-3]。观测数据表明，自1990年以来，全球冰湖数量与面积显著增长，且冰湖溃决洪水事件（GLOFs）每年均有发生^[4-5]。冰湖一方面储存了大量冰川融水，对维系高山区淡水资源具有重要作用；另一方面因坝体溃决导致湖水在短时间内涌入下游排水系统，形成极具破坏力的洪水或泥石流^[6-8]。因此，准确获取冰湖的位置、面积及其动态变化信息，对深入理解气候—冰川—冰湖相互作用机制及山区防灾减灾具有重要意义。

目前全球现代冰川面积约 $15 \times 10^6 \text{ km}^2$ ，占陆地总面积的 11%。根据冰川学气候分期，当前处于现代间冰期，冰川规模呈相对萎缩趋势。而在末次冰期（约 2.6 – 1.15 万年前），全球冰川面积曾高达 $38 \times 10^6 \text{ km}^2$ ^[9-10]。末次冰期冰盛期（Last Glacial Maximum, LGM）是距今最近、冰川覆盖范围最广的全球气候阶段。理论上，冰湖可定义为“末次冰期冰盛期以来，在冰川侵蚀形成洼地中积水而成的天然水体”^[11]。但在现有冰湖编目中，学者通常采用以现代冰川边界向外扩展 10 km 的缓冲区来划定冰湖潜在分布区^[12-14]。该方法虽然操作简便，但随着气候变暖导致冰川消融退缩甚至完全消失，易遗漏大量因古冰川作用形成但现代冰川已消失区域的冰湖。更重要的是，单一的 10 km 距离阈值缺乏充分的冰川学依据。

湖泊遥感提取方法经历了从依赖专家经验的人工目视解译，到结合阈值与指数的半自动化提取，再到基于机器学习与深度学习的全自动提取的技术演进过程。人工目视解译虽能实现高精度的水体提取，但存在效率低下、一致性差等不足。我国现有冰湖编目数据集多采用人工目视解译方法^[12-14]，这在一定程度上限制了数据更新速率与应用广度。水体指数法（如 NDWI、MNDWI、CWI、CIWI 和 EWI 等）是基于多光谱遥感影像的半自动水体提取方法，通过波段运算增强水体信号，结合阈值分割显著提高了水体目标提取效率，但其精度受阈值设定影响较大，且易受山体阴影、积雪或薄云等环境噪声干扰。全自动水体提取方法仅需设定聚类类别数，但分类结果常存在过度分割或漏提问题。基于深度学习的语义分割方法通过监督训练可实现较高提取精度，已成为湖泊遥感领域的研究热点^[15-16]。

目前已有许多针对特定区域的冰湖编目研究。例如，Nie 等^[17]利用 348 景 Landsat 卫星影像绘制了 1990–2015 年喜马拉雅冰湖分布图，结果显示该阶段冰湖面积扩大了 14.1%；Gardelle 等^[18]绘制了 1990 年、2000 年和 2009 年喜马拉雅山、兴都库什山和喀喇昆仑山地区的冰湖地图，发现不同区域冰湖面积差异显著；Wang 等^[14]提取了 1990 年与 2018 年亚洲高海拔地区冰湖数据，表明 28 年间冰湖面积平均扩张率为 15.2%；Zhang 等^[19]研制了第三极地区冰湖数据集，指出 1990–2010 年间该区域冰湖面积年均扩张 6.43 km^2 ；Shugar 等^[4]基于 Landsat 系列影像绘制了全球冰湖分布图，数据集涵盖 1990–2018 年间 5 个多年期平均值，估算出该时段全球冰湖体积增加约 48%。尽管冰湖编目研究已取得一定进展，但在冰湖分布区界定与高效提取方面仍面临挑战：一方面，冰湖潜在分布区的划定缺乏统一且科学的标准；另一方面，传统人工目视解译方式耗时耗力，难以满足大范围、高精度的冰湖编目需求，限制了对冰湖动态变化的深入认识。

因此,本研究以青藏高原为研究区,提出一种兼顾山系单元和冰川边界的新方法,用于界定冰湖潜在分布区;结合 Sentinel-2 MSI 遥感影像与 Attention DeepLab V3+深度学习模型,实现 2023 年青藏高原冰湖边界的自动提取与系统编目,从而为探究冰川演变规律、评估气候变化影响、管理淡水资源及预警冰湖溃决洪水风险提供数据支撑与科学依据。

1 数据采集和处理方法

1.1 研究区概况

青藏高原是全球海拔最高(平均海拔>4000 m)的高原,其独特的下垫面对区域乃至全球气候与生态系统具有深远影响,并孕育了全球最为重要的冰冻圈系统。青藏高原现存冰川 7.35×10^4 条,总面积 $8.51 \times 10^4 \text{ km}^2$,主要分布于喜马拉雅山、念青唐古拉山、昆仑山和喀喇昆仑等主要山系(图 1)。这些冰川不仅是长江、黄河、澜沧江、恒河、印度河等亚洲重要河流的发源地,更是维系“亚洲水塔”生态功能的关键要素。在全球变暖背景下,近几十年青藏高原冰川普遍呈现退缩趋势,而湖泊则表现出明显的扩张态势,这种“冰川退缩-湖泊扩张”的耦合变化过程,不仅深刻影响着区域水资源平衡,也对下游生态系统和人类活动产生深远影响。

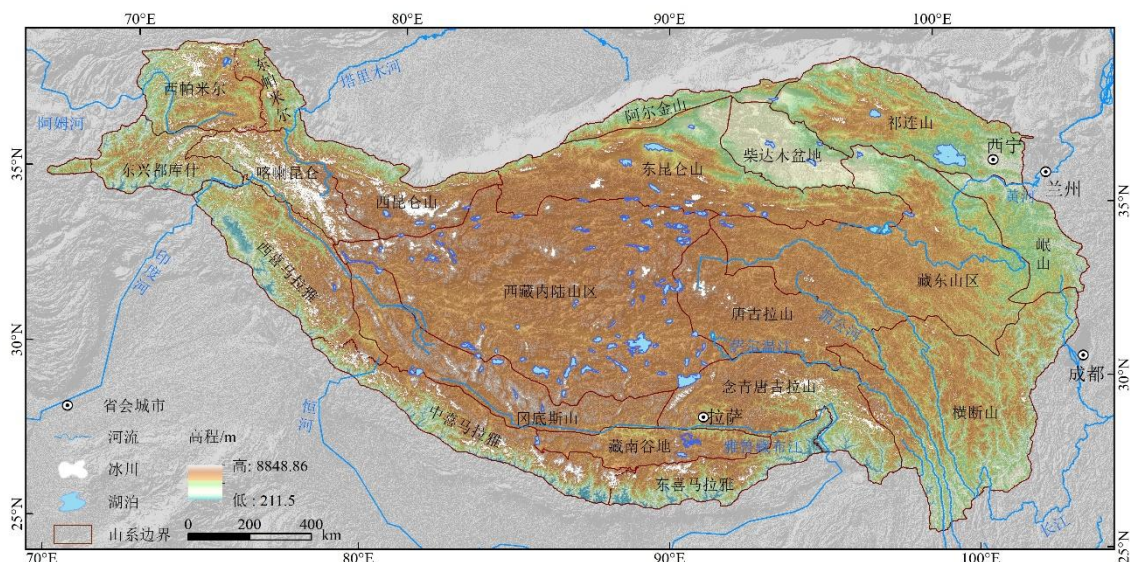


图 1 研究区

Figure 1 The study area

1.2 数据源

本数据集所采用的数据源主要包括:

(1) Randolph Glacier Inventory (RGI) V7.0 数据。该数据源自全球陆地冰监测计划 (Global Land Ice Measurements from Space, GLIMS), 从美国国家雪冰数据中心 (National Snow and Ice Data Center, NSIDC) 获取 (<https://doi.org/10.5067/f6jmovy5navz>)。最新发布的 RGI V7.0 数据集包含冰川轮廓、冰川复合体、冰川相交产品及冰川中心线四类数据产品。本研究选用其中的冰川复合体产品, 以界定冰湖潜在分布区域并辅助判别冰湖类型。

(2) Sentinel-2 MSI 影像。本研究基于 Google Earth Engine 云平台 (<https://earthengine.google.com/>) 获取的 296 景 2023 年 Sentinel-2 MSI Level-2A 级遥感影像进行冰湖提取。影像已完成辐射定标、几

何精校正和大气校正等预处理, 选用空间分辨率为 10 m 的蓝、绿、红及近红外 4 个波段。绝大多数影像时间为 6–10 月, 并配合时序中值合成算法有效消除了云层干扰, 平均云量<15%。

(3) FABDEM V1-2 数据。该数据是由 Fathom 公司与布里斯托大学共同研发的全球数字高程模型数据, 基于哥白尼数字高程模型 (Copernicus DEM, COP DEM) 去除了森林和建筑物的高度偏差, 空间分辨率为 30 m。数据从布里斯托大学网站(<https://doi.org/10.5523/bris.s5hqmjcdj8yo2ibzi9b4ew3sn>) 获取, 本研究利用该数据提取冰湖的高程参数。

(4) GMBA Mountain Inventory V2.0 数据。这是一个具有层级结构的全球山脉矢量数据集, 涵盖多尺度山脉范围, 可通过 EarthEnv 平台 (<https://www.earthenv.org/mountains>) 获取。本研究利用该数据集提取山系尺度的空间分析单元。

(5) Global Dam Watch (GDW)数据集。该数据集是由加拿大麦吉尔大学研究团队主导开发的全球大坝与水库综合数据库 (<https://www.globaldamwatch.org/directory>), 包括全球大坝地理参考数据库 (GOODD)、全球水库和大坝数据库 (GRAND) 以及未来水电站水库和大坝数据库 (FHReD) 3 个核心子集。本研究使用该数据集剔除非冰湖水体。

1.3 数据处理

(1) 冰湖潜在分布区确定

鉴于末次冰期冰盛期冰盖与现代山系边界重合度较高^[9-10], 本研究采用一种改进的冰湖分布范围界定方法, 即将最小单元山系边界与现代冰川边界 10 km 缓冲区空间并集, 划定为冰湖潜在分布范围 (图 2)。

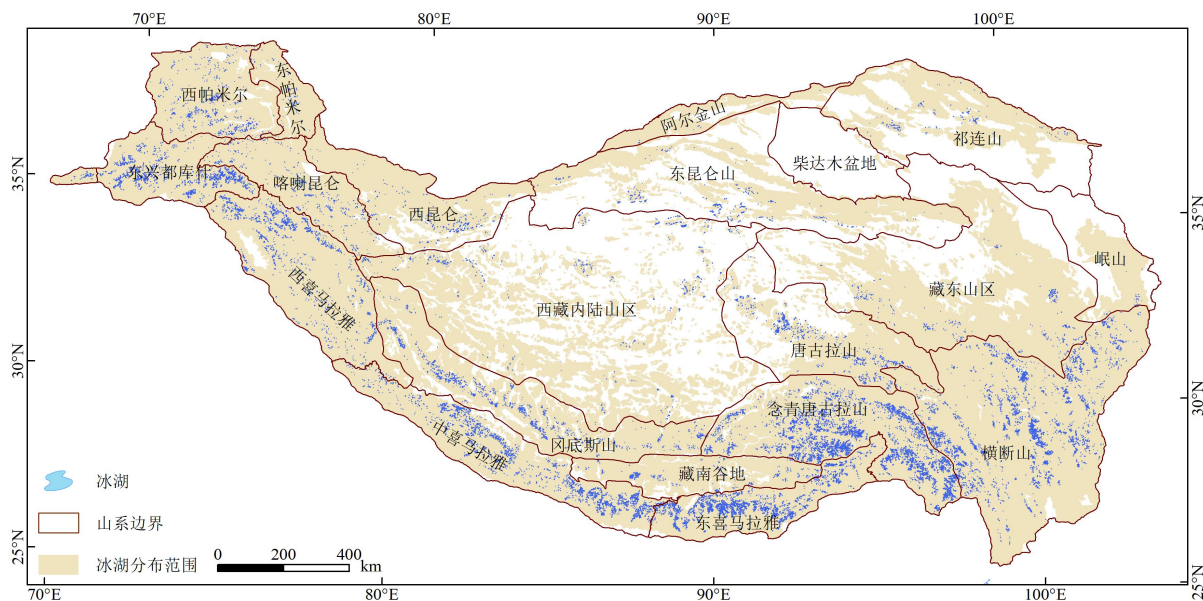


图 2 青藏高原冰湖潜在分布区

Figure 2 Potential distribution of glacial lakes on the Tibetan Plateau

基于冰川发育的地形控制理论, 首先采用山系最小单元边界进行初步划定。冰川发育受气候、地形与海拔协同控制, 其中山地地形具有决定性作用: 一方面通过海拔效应形成低温环境, 当山体超过冰川物质平衡线高度时, 可满足冰川形成的温度条件; 另一方面通过地形抬升增强降水, 使迎风坡成为冰川物质积累的来源区。该方法聚焦于具备冰川发育潜力的高海拔区域, 既体现了地形-降水的耦合特征, 也有助于清晰界定冰川侵蚀与堆积的范围, 从而精准识别冰川潜力区与冰湖分布区,

并有效排除低海拔地区干扰，为冰川分布提供了合理的空间框架。

然而，仅依赖山系边界界定冰湖分布范围存在一定局限性，尤其在识别低海拔平缓区的冰湖时更为明显。这主要源于冰川的动态变化：随着全球变暖导致冰川退缩，新生冰湖常形成于冰川前缘及周边区域。冰川退缩不仅改变了冰川分布格局，还通过融水积聚促进了冰湖的形成与扩张。为此，本研究进一步叠加了冰川边界 10 km 缓冲区。该范围设置基于冰川水文学理论知识，综合考虑以下三个因素：冰碛物堆积通常延伸至冰川末端数公里；冰川融水径流影响范围常超过 10 km；以及冰川边缘微地形对冰湖形成具有显著控制作用。通过将山系边界与冰川缓冲区进行空间叠加，本方法在保留地形单元科学性的同时，兼顾了冰川动态变化的影响，从而全面覆盖古冰湖与新生冰湖。该界定方法通过引入山系边界这一自然地理单元，不仅考虑了现代冰川与冰湖的空间关联，也涵盖了古冰川作用形成的冰湖，有效解决了因冰川消失导致的冰湖提取遗漏问题；同时，该方法突破了固定距离阈值的限制，使划定范围更符合冰湖发育的地形控制规律。

（2）基于 Attention DeepLab V3+模型的冰湖提取

近年来，深度学习技术在遥感影像分析领域取得突破性进展，其中语义分割算法的发展尤为显著。Chen 等^[20]提出的 DeepLab V3+是当前主流的语义分割框架之一，其编码器采用空洞空间金字塔池化（Atrous Spatial Pyramid Pooling, ASPP）模块，通过并行不同扩张率的空间卷积有效捕获多尺度上下文信息；同时，解码器通过融合浅层高分辨率特征与深层语义特征，优化了细节信息的恢复能力。然而，DeepLab V3+模型内置的 ASPP 模块未能完全捕获目标局部特征间的联系，导致分割结果可能出现空洞现象。此外，其骨干网络 Xception 网络层较多、参数量较大，增加了模型的复杂度。为此，Chu 等^[21]提出的 Attention DeepLab V3+模型采用更轻量级的 ResNet-34 作为主干网络（图 3），并在 ASPP 模块中加入卷积注意力模块（Convolutional Block Attention Module, CBAM），自适应地增强特征图中目标地物相关通道和空间位置权重，同时抑制无关背景特征，使模型更聚焦于目标提取。为进一步提升模型鲁棒性，Chu 等^[21]还采用了测试时增强（Test Time Augmentation, TTA）策略。实验结果显示，该模型在关键评价指标上优于 FCN、U-Net 及原始 DeepLab V3+等模型，能够有效区分冰川与终碛湖、薄雪等易混淆地物，提取出的地物边界也更为连续。因此，本研究采用 Attention DeepLab V3+模型进行冰湖提取。

样本库的质量直接影响模型的训练效果与泛化性能。在冰湖样本库制作过程中，选取横断山、东喜马拉雅、中喜马拉雅和西喜马拉雅四个典型冰湖密集区作为样本标注区域。这些区域涵盖了不同光学特性的各类冰湖，包括冰川接触型（冰碛湖和冰面湖）与非接触型冰湖，以及不同营养状态的湖泊样本（图 4），为模型训练提供了高质量的数据基础。通过人工目视解译，最终构建了包含 7680 个样本的标准化数据集（512×512 像素，4 波段）。本研究采用的 4 个波段为 Sentinel-2 MSI 影像的蓝光波段（B2, 490 nm）、绿光波段（B3, 560 nm）、红光波段（B4, 665 nm）和近红外波段（B8, 842 nm），该波段组合在增强水体与背景地物光谱差异方面具有明显优势。在数据划分方面，采用随机抽样方法将样本按 9:1 比例划分为训练集（6912 个样本）和验证集（768 个样本），通过 Scikit-learn 的 `train_test_split` 函数实现，并设置随机种子为 42 以保证结果可重复性。

在冰湖数据集制作过程中，采用 PyTorch 深度学习框架实现 Attention DeepLab V3+模型的训练。PyTorch 作为当前主流的深度学习框架，具备动态计算图（define-by-run）机制、自动微分（Autograd）功能以及与 NumPy 兼容的张量运算体系，确保了模型开发过程中的灵活性与高效性。在硬件加速方面，PyTorch 提供了完善的 GPU 并行计算支持，并集成了丰富的优化工具和预训练模型资源，使其成为深度学习模型研发的理想平台。具体训练参数设置如下：训练周期（epoch）为 120 次，批量大

小 (batch size) 为 64, 初始学习率设为 1×10^{-4} 并采用自适应调整策略。训练过程显示, 训练集损失值随迭代次数增加呈稳定下降趋势, 最终收敛至 0.023; 同时验证集准确率持续提升并趋于稳定, 最高达到 0.888 (图 5), 表明模型具有良好的学习性能和泛化能力。

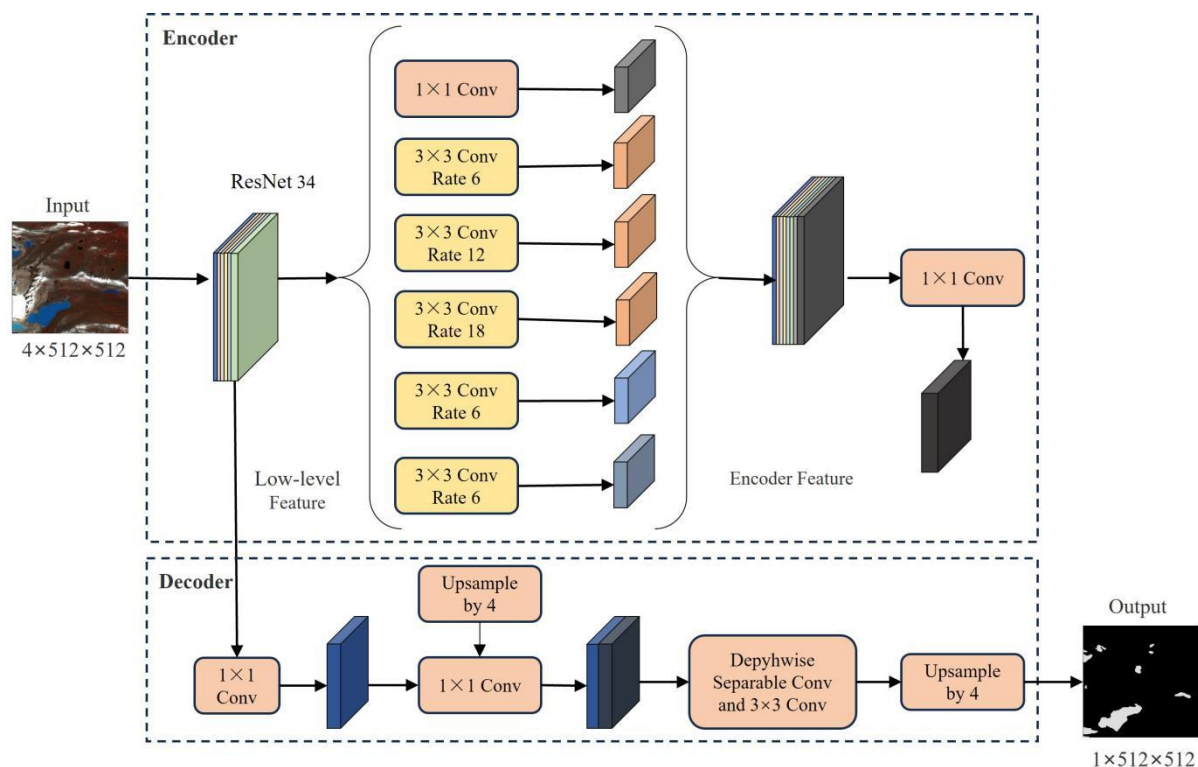


图 3 Attention DeepLab V3+网络架构

Figure 3 Architecture of the Attention DeepLab V3+ network

本研究使用训练好的 Attention DeepLab V3+模型对青藏高原冰湖进行系统提取。提取过程中, 首先以 512×512 像素的滑动窗口对 Sentinel-2 MSI 影像进行分块预测, 随后基于地理坐标信息对各预测窗口进行无缝拼接, 最终生成与原始影像空间分辨率一致的完整预测结果。预测结果后处理流程包括: (1) 将栅格数据转换为矢量数据; (2) 删除面积 $< 0.005 \text{ km}^2$ 的碎斑噪声; (3) 与 GDW 数据集进行空间叠加分析, 剔除误识别的水库等人工水体; (4) 针对结果数据集中存在但现有公开数据集中缺失的细小支流信息和受阴影干扰的冰湖边界, 基于 Google Earth 的亚米级历史影像, 借助 ArcGIS 软件编辑工具进行矢量边界精细化调整。如图 6 所示, 该方法在冰湖边界识别方面表现出良好的性能, 其中真值为人工目视解译结果。

(3) 冰湖属性构建

在提取的青藏高原冰湖矢量数据基础上, 按照冰湖编目规范^[14]完善各冰湖属性信息。为建立清晰且逻辑一致的冰湖分类体系, 本研究综合参考前人研究成果^[11,14], 依据冰湖与冰川的空间位置关系和成因机制, 将冰湖划分为两大类型: 冰川接触湖与非冰川接触湖。其中, 冰川接触湖是指与现代冰川冰体相接触的湖泊, 其形成和存续与冰川作用密切相关。依据地貌特征和发生位置, 冰川接触湖进一步细分为冰面湖和冰碛湖两类, 前者形成于冰川表面的洼地或裂隙中; 后者由冰川退缩后遗留的冰碛物 (如终碛、侧碛) 阻塞形成, 湖盆边界与冰碛垄直接接触。非冰川接触湖则指与现代冰川无直接接触的湖泊, 其形成可能与古冰川作用或其他非冰川地质营力有关。本研究首先基于冰湖与冰川的拓扑关系提取冰面湖, 然后结合 ArcGIS Earth 和 Google Earth 高分辨率影像及三维场景,

确定其他冰湖类型。

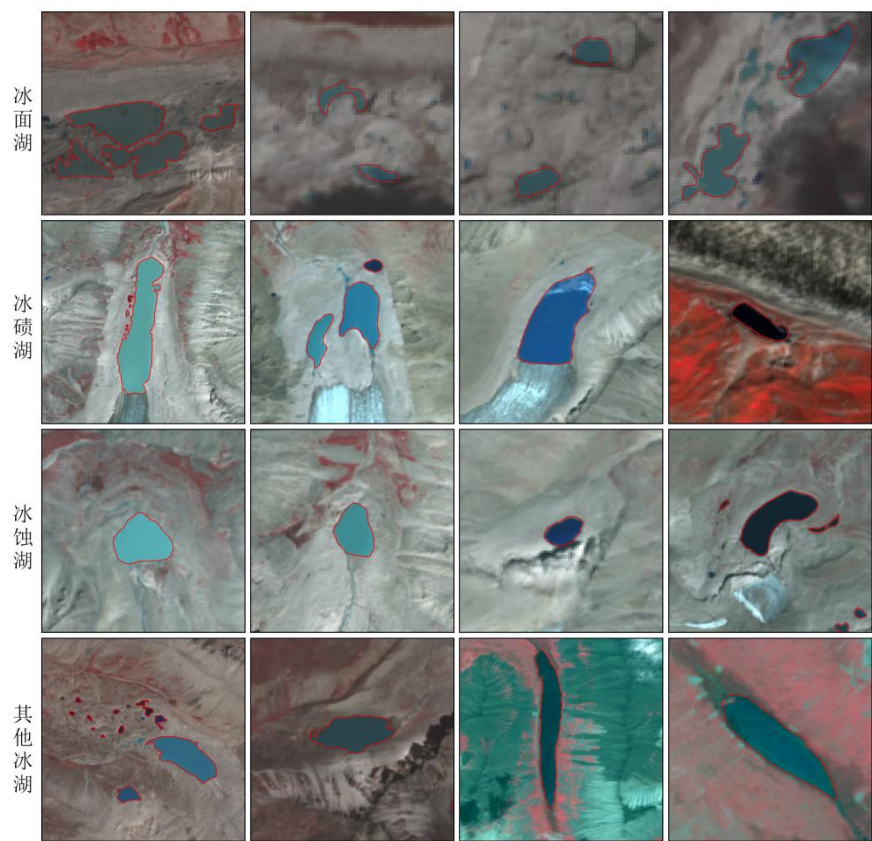


图 4 冰湖样本标注示例

Figure 4 Samples annotation examples of glacial lakes

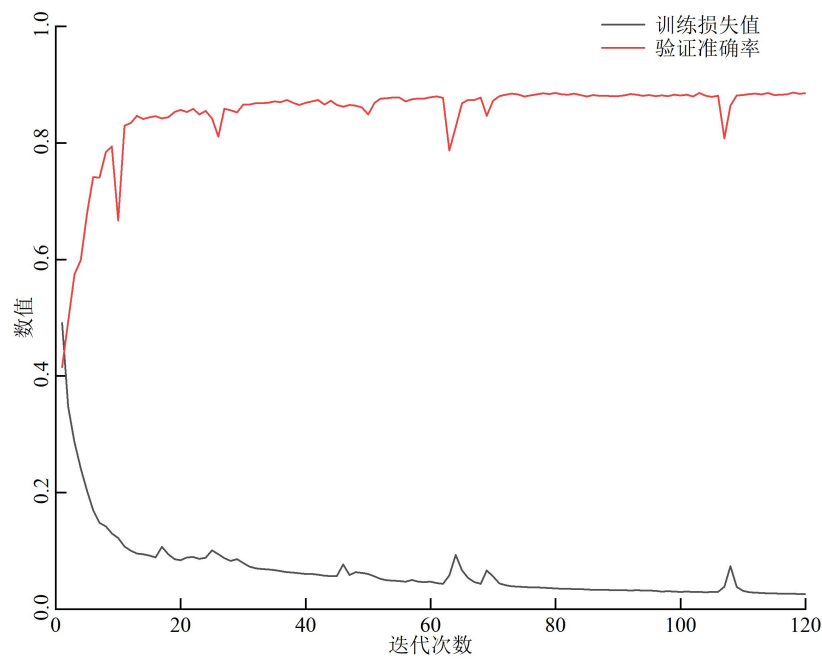


图 5 训练集损失值和验证集准确率

Figure 5 Training set loss values and validation set accuracy

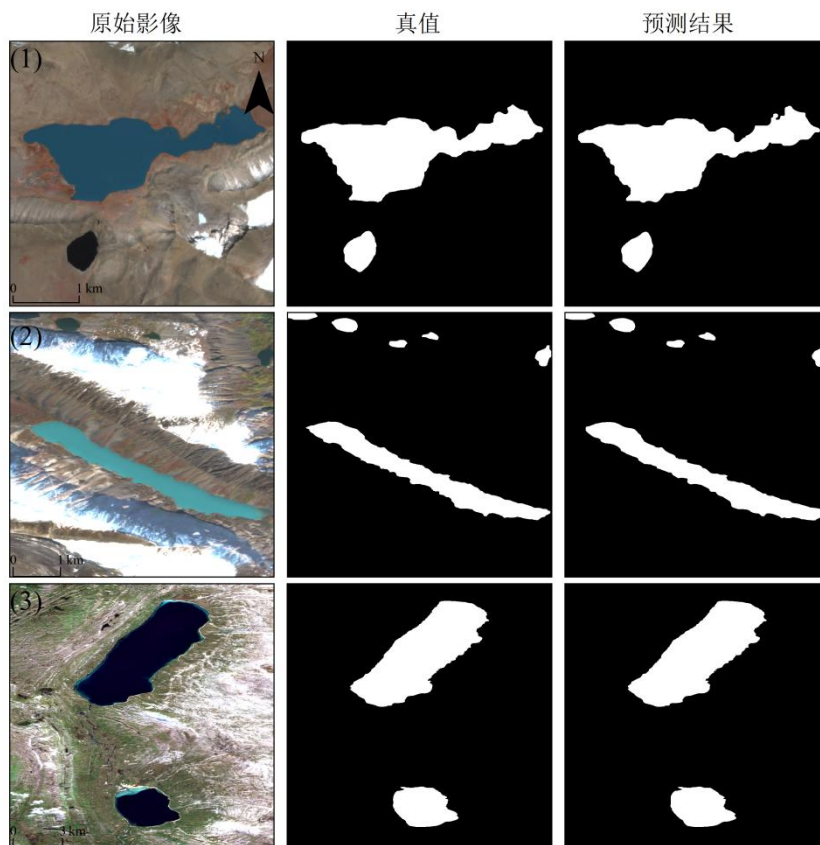


图 6 模型提取冰湖结果对比

Figure 6 Comparison of glacial lake extraction results by the model

冰湖库容采用针对不同类型冰湖提出的库容计算公式^[22]获得，非冰川接触湖、冰面湖和冰碛湖的计算公式如下：

$$V = 17.15V = A^{1.15} \quad (1)$$

$$V = 18.25V = A^{1.31} \quad (2)$$

$$V = 44.47V = A^{1.5} \quad (3)$$

式中， V 代表冰湖库容（ 10^6 m^3 ）， A 代表冰湖面积（ km^2 ）。

2 数据样本描述

2.1 数据属性表

青藏高原冰湖数据集由 29089 个冰湖矢量图斑构成，属性表包含 10 个字段（表 1），记录了冰湖的空间位置、几何参数等信息。其中，GL_ID 为冰湖编码，以确保数据的唯一性和可追溯性；GL_Lon 和 GL_Lat 字段为冰湖质心的经纬度坐标（采用 WGS84 坐标系）；Elevation 表示冰湖水面平均高程，单位为 m；Area 表示冰湖面积，单位为 km^2 ；Error 表示冰湖面积误差，单位为 km^2 ；Perimeter 表示冰湖周长，单位为 km。为科学评估数据质量并深化数据应用，本数据集增加了以下属性：Volume 表示冰湖库容，单位为 10^6 m^3 ；Type 表示冰湖类型，其中 0 代表非冰川接触湖，1 代表冰面湖，2 代表冰碛湖；Mountain 表示冰湖所在山脉名称。

表 1 2023 年青藏高原冰湖编目数据集属性表说明

Table 1 Description of the attribute table for the dataset of glacial lakes inventory on the Tibetan Plateau in 2023

编号	字段名称	数据类型	字段描述
1	GL_ID	Short Integer	冰湖编码
2	GL_Lon	Float	冰湖质心经度，单位为度
3	GL_Lat	Float	冰湖质心纬度，单位为度
4	Elevation	Float	冰湖水面平均高程，单位为 m
5	Area	Float	冰湖面积，单位为 km ²
6	Error	Float	冰湖面积误差，单位为 km ²
7	Perimeter	Float	冰湖周长，单位为 km
8	Volume	Float	冰湖库容，单位为 10 ⁶ m ³
9	Type	Short Integer	冰湖类型，其中 0 表示非冰川接触湖，1 表示冰面湖，2 表示冰碛湖
10	Mountain	Text	冰湖所在的山脉名称

2.2 数据图形样本

截至 2023 年，青藏高原冰湖总数量为 29089 个，总面积约 2020.65 km²。空间分布上，冰湖集中发育于青藏高原南部和东南部的高山区域（图 7）。其中，念青唐古拉山的冰湖发育最为旺盛，冰湖数量和面积分别为 6306 个和 483.65 km²；其次是横断山、东喜马拉雅山和中喜马拉雅山，冰湖规模分别为 5131 个（256.37 km²）、3619 个（265.26 km²）和 2971 个（249.07 km²）。这 4 个区域冰湖数量和面积分别占全区总量的 61.97%和 62.08%，构成青藏高原冰湖分布的核心区域。东兴都库什、西喜马拉雅山、冈底斯山和唐古拉山的冰湖数量均超过 1000 个，总面积为 401.50 km²。相比较而言，青藏高原北部的东帕米尔、昆仑山、阿尔金山和祁连山等区域冰湖发育较为有限，冰湖数量共 1249 个，总面积仅 103.76 km²。

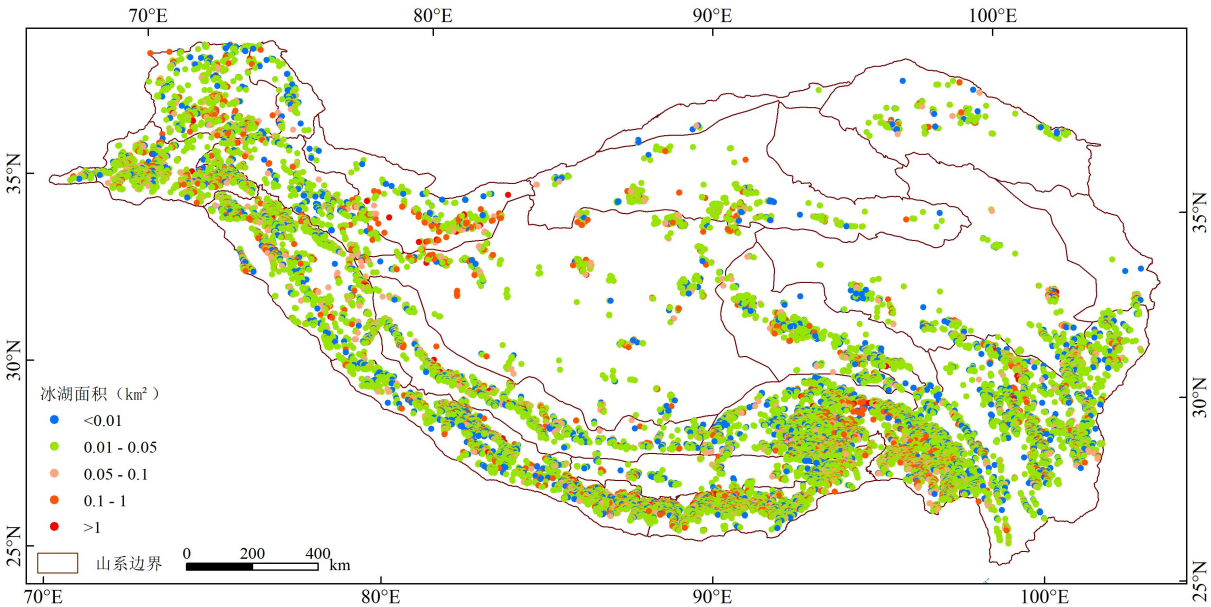


图 7 不同面积等级冰湖空间分布

Figure 7 Spatial distribution of glacial lakes in different area intervals

从数量上而言,青藏高原冰湖以面积 $<0.05\text{ km}^2$ 的小型湖泊为主。其中,面积 $\leq 0.01\text{ km}^2$ 的冰湖共 3334 个(11.46%),总面积仅 29.94 km^2 (1.48%);面积介于 $0.01 - 0.05\text{ km}^2$ 的冰湖数量占绝对优势(18227 个, 62.66%),但其总面积占比仅为 20.39%(图 8),这类冰湖主要密集分布于念青唐古拉山和横断山脉的高海拔区域。虽然面积 $>1\text{ km}^2$ 的大型冰湖数量稀少(197 个, 0.68%),但其总面积高达 489.45 km^2 (24.22%),这类冰湖主要分布在念青唐古拉山和喜马拉雅山中东部等冰川作用强烈的高山区域,对区域水文过程和灾害风险具有重要影响。

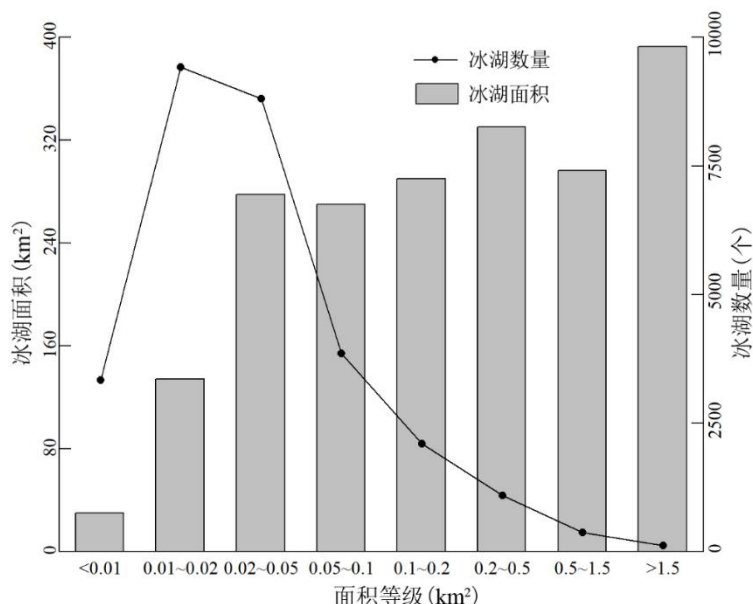


图 8 2023 年青藏高原不同面积等级冰湖数量与面积

Figure 8 Number and area of glacial lakes in different sizes on the Tibetan Plateau in 2023

从冰湖类型来看(表 2),青藏高原冰湖以非冰川接触湖为主,共计 21492 个(73.88%),总面积 1485.23 km^2 (73.50%)。这类冰湖在各山系中均有分布,其中念青唐古拉山(5163 个, 392.37 km^2)、横断山(4670 个, 240.97 km^2)和东喜马拉雅山(2838 个, 175.79 km^2)分布最为广泛。冰碛湖作为第二大类冰湖,共计 7059 个(24.27%),总面积 523.53 km^2 (25.91%)。此类冰湖在中喜马拉雅山(1348 个, 152.78 km^2)、念青唐古拉山(1106 个, 89.61 km^2)和东喜马拉雅山(746 个, 88.72 km^2)发育尤为显著。值得注意的是,中喜马拉雅山的冰碛湖面积占比高达 61.34%,该区域同时也是青藏高原冰湖溃决洪水的多发区^[23]。冰面湖数量最少(538 个, 1.85%),总面积仅 11.89 km^2 (0.59%),主要集中分布在喀喇昆仑山(150 个, 2.91 km^2)和中喜马拉雅山(124 个, 2.58 km^2)。其中,喀喇昆仑山的冰面湖数量占该区域冰湖总数的 27.88%,显著高于其他地区。

表 2 青藏高原各山系不同类型冰湖数量和面积

Table 2 Number and area of different types of glacial lakes in each mountain range on the Tibetan Plateau

山系	非冰川接触湖		冰川接触湖			
			冰碛湖		冰面湖	
	数量	面积/ km^2	数量	面积/ km^2	数量	面积/ km^2
阿尔金山	9	0.14	9	0.42	0	0
祁连山	148	8.44	79	6.13	4	0.07
岷山	96	3.60	6	0.41	0	0

山系	非冰川接触湖		冰川接触湖			
			冰碛湖		冰面湖	
	数量	面积/km ²	数量	面积/km ²	数量	面积/km ²
东喜马拉雅	2838	175.79	746	88.72	35	0.75
中喜马拉雅	1499	93.71	1348	152.78	124	2.58
西喜马拉雅	1128	67.14	748	35.22	37	0.53
东兴都库什	1441	71.83	496	25.17	5	0.06
东帕米尔	48	4.17	46	2.12	13	0.21
西帕米尔	420	75.92	367	22.92	24	0.44
喀喇昆仑山	151	12.38	256	11.82	150	2.91
东昆仑	357	17.48	176	9.77	9	0.24
西昆仑	189	43.57	134	10.04	28	0.96
西藏内陆山区	626	53.50	195	18.39	29	0.72
冈底斯山	884	111.35	576	20.01	16	0.29
藏南谷底	378	21.13	61	2.57	2	0.02
念青唐古拉山	5163	392.37	1106	89.61	37	1.68
唐古拉山	1014	58.63	215	11.05	12	0.22
藏东地区	433	33.11	45	1.20	2	0.03
横断山	4670	240.97	450	15.19	11	0.21
总计	21492	1485.23	7059	523.53	538	11.89

3 数据质量控制和评估

冰湖边界的提取精度主要受遥感影像质量、空间分辨率和提取方法的准确性影响。一方面，本数据集选用了云量<10%、冰湖处于未冻结状态的 Sentinel-2 MSI 影像，这在很大程度上确保了影像的高质量；另一方面，相比人工目视解译和水体指数法，本研究所采用的 Attention DeepLab V3+模型具有更高的计算效率和更强的适用性，能够有效保证数据提取结果一致性。然而，与亚米级等更高空间分辨率卫星遥感影像相比，Sentinel-2 MSI 影像空间分辨率较低，一定程度上制约了冰湖边界细节的刻画。为客观评估数据精度，本研究采用下式^[24]计算各冰湖的面积误差（对应属性表中的 Error 字段）：

$$\varepsilon = N \times A/2 \quad (4)$$

式中， ε 为影像空间分辨率造成的冰湖面积误差； N 为冰湖轮廓的周长； A 为单个像元的边长（10 m）。经计算，由影像空间分辨率造成的冰湖总面积误差为 134.88 km²，占冰湖总面积的 6.68%。

需要说明的是，本数据集仍存在一定的局限性：首先，Sentinel-2 MSI 影像在识别面积<0.005 km²的极小冰湖以及精确刻画其边界方面仍存在固有缺陷，可能导致部分极小冰湖被遗漏；其次，尽管模型训练样本已覆盖横断山、喜马拉雅等四个典型冰湖密集区，但由于青藏高原不同区域冰湖的光学特性存在差异，当前样本库尚未实现区域全覆盖，这可能影响模型在局部地区的泛化能力和提取精度。为应对上述问题，未来可考虑融合更高分辨率的遥感影像（如 Planet 数据）进行交叉验证与

信息补充，并通过扩展训练样本类型和空间覆盖范围，进一步提升模型的鲁棒性与适应性。

4 数据价值

本数据集全面记录了 2023 年青藏高原冰湖的空间位置、几何特征和类型属性等关键参数。在科学层面，本数据集采用深度学习技术实现了对青藏高原全域冰湖的高精度调查，为探究气候—冰川—冰湖相互作用提供了基础数据，尤其对量化冰川退缩与冰湖扩张之间的耦合机制研究具有重要价值。在应用层面，本数据集可支撑区域水资源储量评估、冰湖溃决洪水灾害风险预警模型构建和高原水文系统碳循环等相关研究。

5 数据使用方法和建议

2023 年青藏高原冰湖编目数据集采用 Shapefile 格式存储，地理坐标系为 WGS-1984，投影坐标系为 Albers 等积投影。使用者可以通过 ArcMap、QGIS、SuperMap 等地理信息系统软件进行数据的读取、编辑和更新等操作。

致 谢

感谢 Google Earth Engine (GEE) 平台提供的 Sentinel-2 MSI 数据，同时感谢布里斯托大学提供的 FABDEM V1-2 数字高程数据，以及 Global Land Ice Measurements from Space (GLIMS) 提供的 Randolph Glacier Inventory V7.0 数据和 EarthEnv 平台提供的 GMBA Mountain Inventory v2.0 山脉边界数据。

数据作者分工职责

田志娟，女，甘肃省康乐县人，硕士研究生，研究方向为 GIS 设计与开发。主要承担工作：数据下载，模型训练与冰湖提取，文稿撰写。

姚晓军，男，山西省夏县人，博士，教授，研究方向为地理信息技术与冰冻圈变化。主要承担工作：研究思路设计和文稿修改。

常辉，男，河南省民权县人，硕士研究生，研究方向为 GIS 设计与开发。主要承担工作：模型环境的安装和模型参数的调整。

段红玉，女，甘肃省临洮县人，博士，工程师，研究方向为冰冻圈灾害与地物遥感解译。主要承担工作：文稿修改。

参考文献

- [1] ZEMP M, HUSS M, THIBERT E, et al. Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016[J]. Nature, 2019, 568(7752): 382–386. DOI:10.1038/s41586-019-1071-0.
- [2] ZEMP M, FREY H, GÄRTNER-ROER I, et al. Historically unprecedented global glacier decline in the early 21st century[J]. Journal of Glaciology, 2015, 61(228): 745–762.

- DOI:10.3189/2015jog15j017.
- [3] HUGONNET R, MCNABB R, BERTHIER E, et al. Accelerated global glacier mass loss in the early twenty-first century[J]. *Nature*, 2021, 592(7856): 726–731. DOI:10.1038/s41586-021-03436-z.
- [4] SHUGAR D H, BURR A, HARITASHYA U K, et al. Rapid worldwide growth of glacial lakes since 1990[J]. *Nature Climate Change*, 2020, 10(10): 939–945. DOI:10.1038/s41558-020-0855-4.
- [5] CARRIVICK J L, TWEED F S. A global assessment of the societal impacts of glacier outburst floods[J]. *Global and Planetary Change*, 2016, 144: 1–16. DOI:10.1016/j.gloplacha.2016.07.001.
- [6] RAWAT M, JAIN S K, AHMED R, et al. Glacial lake outburst flood risk assessment using remote sensing and hydrodynamic modeling: a case study of Satluj basin, Western Himalayas, India[J]. *Environmental Science and Pollution Research*, 2023, 30(14): 41591–41608. DOI:10.1007/s11356-023-25134-1.
- [7] FURIAN W, MAUSSION F, SCHNEIDER C. Projected 21st-century glacial lake evolution in high mountain Asia[J]. *Frontiers in Earth Science*, 2022, 10: 821798. DOI:10.3389/feart.2022.821798.
- [8] VEI G, KORUP O, WALZ A. Hazard from Himalayan glacier lake outburst floods[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, 117(2): 907–912. DOI:10.1073/pnas.191489811.
- [9] 施雅风. 中国第四纪冰川与环境变化[M]. 石家庄: 河北科学技术出版社, 2006. [SHI Y F. The Quaternary Glaciations and Environmental Variations in China[M]. Shijiazhuang: Hebei Science & Technology Press, 2006.]
- [10] 施雅风. 中国第四纪冰川新论[M]. 上海: 上海科学普及出版社, 2011. [SHI Y F. New Understanding of Quaternary Glaciations in China[M]. Shanghai: Shanghai Popular Science Press, 2011.]
- [11] 姚晓军, 刘时银, 韩磊, 等. 冰湖的界定与分类体系: 面向冰湖编目和冰湖灾害研究[J]. *地理学报*, 2017, 72(7): 1173–1183. DOI:10.11821/dlxb201707004. [YAO X J, LIU S Y, HAN L, et al. Definition and classification systems of glacial lake for inventory and hazards study[J]. *Acta Geographica Sinica*, 2017, 72(7): 1173–1183. DOI:10.11821/dlxb201707004.]
- [12] 杨成德, 王欣, 魏俊峰, 等. 基于 3S 技术方法的中国冰湖编目[J]. *地理学报*, 2019, 74(3): 544 – 556. DOI:10.11821/dlxb201903011. [YANG C D, WANG X, WEI J F, et al. Chinese glacial lake inventory based on 3S technology method[J]. *Acta Geographica Sinica*, 2019, 74(3): 544–556. DOI:10.11821/dlxb201903011.]
- [13] 杨成德, 王欣, 魏俊峰, 等. 2015 年中国西部冰湖编目数据集[J/OL]. *中国科学数据*, 2018, 3(4): 32–40. DOI:10.11922/csdata.2018.0038.zh. [YANG C D, WANG X, WEI J F, et al. A dataset of glacial lake inventory of West China in 2015[J/OL]. *China Scientific Data*, 2018, 3(4): 32–40. DOI:10.11922/csdata.2018.0038.zh.]
- [14] WANG X, GUO X Y, YANG C D, et al. Glacial lake inventory of high-mountain Asia in 1990 and 2018 derived from Landsat images[J]. *Earth System Science Data*, 2020, 12(3): 2169–2182. DOI:10.5194/essd-12-2169-2020.
- [15] WANG Z B, GAO X, ZHANG Y N, et al. MSLWENet: a novel deep learning network for lake water body extraction of Google remote sensing images[J]. *Remote Sensing*, 2020, 12(24): 4140.

DOI:10.3390/rs12244140.

- [16] THATI J, ARI S. A systematic extraction of glacial lakes for satellite imagery using deep learning based technique[J]. Measurement, 2022, 192: 110858. DOI:10.1016/j.measurement.2022.110858.
- [17] NIE Y, SHENG Y W, LIU Q, et al. A regional-scale assessment of Himalayan glacial lake changes using satellite observations from 1990 to 2015[J]. Remote Sensing of Environment, 2017, 189: 1–13. DOI:10.1016/j.rse.2016.11.008.
- [18] GARDELLE J, ARNAUD Y, BERTHIER E. Contrasted evolution of glacial lakes along the Hindu Kush Himalaya Mountain range between 1990 and 2009[J]. Global and Planetary Change, 2011, 75(1/2): 47–55. DOI:10.1016/j.gloplacha.2010.10.003.
- [19] ZHANG G Q, YAO T D, XIE H J, et al. An inventory of glacial lakes in the Third Pole region and their changes in response to global warming[J]. Global and Planetary Change, 2015, 131: 148–157. DOI:10.1016/j.gloplacha.2015.05.013.
- [20] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Computer Vision–ECCV 2018. Cham: Springer, 2018: 833–851. DOI:10.1007/978-3-030-01234-2_49.
- [21] CHU X D, YAO X J, DUAN H Y, et al. Glacier extraction based on high-spatial-resolution remote-sensing images using a deep-learning approach with attention mechanism[J]. The Cryosphere, 2022, 16(10): 4273–4289. DOI:10.5194/tc-16-4273-2022.
- [22] ZHANG T G, WANG W C, AN B S. A conceptual model for glacial lake bathymetric distribution[J]. The Cryosphere, 2023, 17(12): 5137–5154. DOI:10.5194/tc-17-5137-2023.
- [23] 姚晓军, 刘时银, 孙美平, 等. 20 世纪以来西藏冰湖溃决灾害事件梳理[J]. 自然资源学报, 2014, 29(8): 1377–1390. DOI:10.11849/zrzyxb.2014.08.010. [YAO X J, LIU S Y, SUN M P, et al. Study on the glacial lake outburst flood events in Tibet since the 20th century[J]. Journal of Natural Resources, 2014, 29(8): 1377–1390. DOI:10.11849/zrzyxb.2014.08.010.]
- [24] 刘娟, 姚晓军, 刘时银, 等. 1970–2016 年冈底斯山冰川变化[J]. 地理学报, 2019, 74(7): 1333–1344. DOI:10.11821/dlxb201907005. [LIU J, YAO X J, LIU S Y, et al. Glacier changes in the Gangdisê Mountains from 1970 to 2016[J]. Acta Geographica Sinica, 2019, 74(7): 1333–1344. DOI:10.11821/dlxb201907005.]

论文引用格式

田志娟, 姚晓军, 常辉, 等. 2023 年青藏高原冰湖编目数据集[J/OL]. 中国科学数据, 2026, 11(1). (2026-02-28). DOI: 10.11922/11-6035.ncdc.2025.0112.zh.

数据引用格式

田志娟, 姚晓军, 常辉, 等. 2023 年青藏高原冰湖编目数据集[DS/OL]. V1. Science Data Bank, 2025. (2025-09-12). DOI: 10.57760/sciencedb.j00001.01517.

A dataset of the glacial lake inventory on the Tibetan Plateau in 2023

TIAN Zhijuan¹, YAO Xiaojun^{1,2*}, CHANG Hui¹, DUAN Hongyu^{1,3}

1. College of Geography and Environmental Science, Northwest Normal University, Lanzhou 730070, P.R. China

2. Gansu Provincial Key Laboratory of Oasis Resources, Environment, and Sustainable Development, Lanzhou 730070, P.R. China

3. Gansu Coalfield Geological Bureau No. 149 Team, Lanzhou 730020, P.R. China

*Email: yaoxj_nwnu@163.com

Abstract: Glacial lakes, as characteristic hydrological and geomorphological units formed by glacial ablation, have exhibited rapid expansion trends in alpine and polar regions worldwide in recent years. The systematic inventorying of glacial lakes documenting their spatial locations and geometric parameters, holds significant scientific value for understanding glacier evolution, responses to climate change, freshwater resource assessment, and early warning of glacial lake outburst floods (GLOF) hazards. To address persistent challenges such as ongoing debates over the delineation of potential glacial lake distribution areas and the heavy reliance on manual visual interpretation in remote sensing -based glacial lake extraction, this study developed an innovative approach that integrates mountain units with glacier boundaries to define potential distribution areas of glacial lakes. Based on Sentinel-2 MSI imagery and the Attention DeepLab V3+ deep learning model, we achieved high-precision, fully automated extraction and systematic inventorying of glacial lakes across the entire Tibetan Plateau. The results indicate that there were 29,089 glacial lakes with an area $\geq 0.005 \text{ km}^2$, covering a total area of 2020.65 km^2 on the Qinghai-Tibet Plateau in 2023. These lakes are primarily distributed in the southeastern part of the plateau, including the Nyainqêntanglha Mountains, the Eastern Himalayas, and the Hengduan Mountains. In terms of size structure, small glacial lakes ($0.01\text{-}0.05 \text{ km}^2$) account for the largest proportion numerically (62.66%) yet contribute only 20.39% to the total area. In contrast, large glacial lakes ($>1 \text{ km}^2$), though rare in number (0.68%), but contribute 24.22% of the total lake area. Regarding lake types, non-glacier-contact lakes dominate (73.88%), followed by moraine-dammed lakes (24.27%), while supraglacial lakes represent the smallest proportion (1.85%). The area uncertainty caused by the spatial resolution of the remote sensing imagery is estimated at 134.88 km^2 , accounting for 6.68% of the total glacial lake area. Overall, this dataset not only accurately reflects the current status of glacial lake resources on the Qinghai-Tibet Plateau, but also provides essential data support for key applications such as GLOF risk assessment and regional hydropower development planning.

Keywords: glacial lake inventory; sentinel; deep learning; Qinghai-Tibet Plateau

Dataset Profile

Title	A dataset of the glacial lake inventory on the Tibetan Plateau in 2023
Data authors	TIAN Zhijuan, YAO Xiaojun, CHANG Hui, DUAN Hongyu

Data corresponding author	YAO Xiaojun (yaoxj_nwnu@163.com)
Time range	2023
Geographical scope	Tibetan Plateau (25°N–40°N, 74°E–104°E)
Spatial resolution	10 m
Data volume	81.71 MB
Data format	ESRI Shapefile (compressed in *.zip)
Data service system	https://doi.org/10.57760/sciencedb.j00001.01517 https://www.ncdc.ac.cn/portal/metadata/fe4fb623-3ffe-44a8-a4c9-4f6346f972ec
Sources of funding	Graduate Research Funding Project of Northwest Normal University in 2023 (2023KYZZ-S181); Strategic Action Plan of Oasis Science (NWNULZKX-202301).
Dataset composition	The dataset comprises a single file, archived in compressed format and titled "Tibetan_Plateau_glacial_lakes_inventory_2023", which includes 29089 glacial lakes with an area ≥ 0.005 km ² .