

LIU Fangchao, ZHANG Li, GUO Dijun, LIU Bin, XIE Bin, LÜ Yingbo, CHEN Jian, LING Zongcheng. Impact Crater Database of 10 Landing Regions from Apollo and Chang'E Missions: Construction and Distribution Patterns of Small Impact Crater Databases (in Chinese). *Chinese Journal of Space Science*, 2026, 46(3): 836–849. CSTR:32142.14.cjss.2025–0135. DOI:10.11728/cjss.2026.03.2025–0135

数据论文

月球阿波罗与嫦娥任务 10 个典型着陆区撞击坑数据库：小撞击坑数据库构建与分布规律*

刘方超¹ 张立² 郭弟均³ 刘斌⁴
谢彬⁴ 吕英波¹ 陈剑¹ 凌宗成¹

1(山东大学空间科学与技术学院 威海 264209)

2(山东大学低空科学与工程学院 威海 264209)

3(中国科学院国家空间科学中心 太阳活动与空间天气全国重点实验室 北京 100190)

4(中国科学院空天信息创新研究院 遥感与数字地球全国重点实验室 北京 100101)

摘要 月球是撞击坑保存最完整的内太阳系天体, 拥有就位探测数据和返回样品的月球典型着陆区是认识撞击事件和撞击改造效应的重要研究对象. 其中小撞击坑对月壤的形成和演化具有重要影响, 但是现有的月球撞击坑数据库对直径<100 m 的小尺度撞击坑覆盖不足. 因此, 筛选太阳入射角在 50°~70°之间的月球高分辨率 LROC NAC 影像, 制作了覆盖 10 个着陆区 20 km×20 km 的镶嵌数据, 利用改进的 YOLO11+SAHI 深度学习模型对这 10 个着陆区中直径≥15 m 的撞击坑进行自动提取, 经人工校验后建成了含 357764 条记录的撞击坑数据库. 系统构建了覆盖 6 个阿波罗任务与 4 个嫦娥任务共 10 个典型着陆区、直径≥15 m 的撞击坑数据库, 且数据库中的撞击坑数据完整性优于现有研究. 进一步分析小尺寸撞击坑的密度分布与直径-频率分布特征. 本数据集可为月球地质年代标定、撞击通量演化、月表过程研究及样品分析提供高质量数据支持, 并为未来的撞击坑智能识别模型提供训练与验证基础.

关键词 小撞击坑, 月球, 着陆区, 数据集, 深度学习

中图分类号 P691

数据库(集)基本信息

数据库(集)名称	月球典型着陆区撞击坑数据集
数据通信作者	郭弟均 (guodijun@nssc.ac.cn), 吕英波 (lyb@sdu.edu.cn)
数据作者	郭弟均, 吕英波, 刘方超, 张立, 刘斌, 谢彬, 陈剑, 凌宗成
数据时间范围	2009–2025年
数据空间范围	阿波罗11, 12, 14, 15, 16, 17和嫦娥-3, 4, 5, 6共10次任务着陆点20 km范围内的区域
数据分辨率	0.5~2 m·pixel ⁻¹
数据量	357764条撞击坑记录
数据格式	*.csv, *.shp
数据库(集)访问网址	https://doi.org/10.12176/03.05.00008-V01 , https://cstr.cn/14804.11.03.05.00008-V01
基金项目	国家重点研发计划项目(2022YFF0711400); 国家自然科学基金项目(42472304)
数据库(集)组成	数据集包含以上典型着陆区直径≥15 m的撞击坑. 数据以shapefile文件和 csv文件格式保存, 撞击坑条目命名采用“任务-数字编号”的形式, 例如A11-1, CE6-1.

* 国家重点研发计划项目(2022YFF0711400)和国家自然科学基金项目(42472304)共同资助

2025-07-31 收到原稿, 2026-03-09 收到修定稿

©The Author(s) 2026. This is an open access article under the CC-BY 4.0 License
(<https://creativecommons.org/licenses/by/4.0/>)

Impact Crater Database of 10 Landing Regions from Apollo and Chang'E Missions: Construction and Distribution Patterns of Small Impact Crater Databases

LIU Fangchao¹ ZHANG Li² GUO Dijun³ LIU Bin⁴
XIE Bin⁴ LÜ Yingbo¹ CHEN Jian¹ LING Zongcheng¹

¹(School of Space Science and Technology, Shandong University, Weihai 264209)

²(School of Low-altitude Science and Engineering, Shandong University, Weihai 264209)

³(State Key Laboratory of Solar Activity and Space Weather, National Space Science Center,
Chinese Academy of Sciences, Beijing 100190)

⁴(State Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute,
Chinese Academy of Sciences, Beijing 100101)

Abstract Among the celestial bodies within the solar system, the Moon maintains the most pristine conditions of impact craters. Some landing sites on the lunar surface have *in-situ* exploration data and experimental measurement results from the returned samples, providing distinguished meanings to understand the impact events and their reshaping effects on the lunar surface. Small craters play a significant role in the formation and evolution of lunar regolith. However, existing lunar impact crater databases lack comprehensive coverage of small-scale impact craters with diameters less than 100 m. Therefore, this study produced ten mosaic images of 20 km×20 km in width, using the high-resolution Lunar Reconnaissance Orbiter Camera Narrow Angle Camera images acquired under solar incidence angles between 50° and 70°. These images cover ten lunar landing sites, including six Apollo missions and four Chang'E missions. An improved YOLO11+SAHI deep learning model was then applied to automatically extract craters with diameters ≥ 15 m within these regions. After manual verification, a high-quality crater database containing 357764 records was constructed. Compared with existing datasets, the proposed database demonstrates superior crater completeness. Based on this database, the density distribution and diameter-frequency characteristics of small-sized impact craters were further calculated and analyzed. This dataset can provide robust support for studies of lunar geological chronology, impact flux evolution, surface process, and sample interpretation. In addition, it supplies a valuable training and validation resource for future artificial intelligence models of detecting the impact craters.

Key words Small-scale impact craters, Moon, Landing sites, Datasets, Deep learning

Database/Dataset Profile

Title	A crater catalog for typical landing regions of the Moon
Data Corresponding Author	GUO Dijun (guodijun@nssc.ac.cn), LÜ Yingbo (lyb@sdu.edu.cn)
Data Author (s)	GUO Dijun, LÜ Yingbo, LIU Fangchao, ZHANG Li, LIU Bin, XIE Bin, CHEN Jian, LING Zongcheng
Time Range	2009–2025
Spatial Range	The areas of 20 km×20 km around 10 landing sites including Apollo 11, 12, 14, 15, 16, and 17, along with Chang'E-3, 4, 5, and 6
Data Resolution	0.5~2 m·pixel ⁻¹
Data Volume	357764 impact crater records
Data Format	*.csv, *.shp
Database/Dataset URL	https://doi.org/10.12176/03.05.00008-V01 , https://cstr.cn/14804.11.03.05.00008-V01
Source (s) of Funding	National Key Research and Development Program of China (2022YFF0711400), National Natural Science Foundation of China (42472304)
Dataset/Database Composition	The dataset includes impact craters with diameters ≥ 15 m in the aforementioned typical landing zones. Data is stored in shapefile and csv formats, with crater entries named in the format "Mission-Numeric ID," such as A11-1 and CE6-1

0 引言

撞击坑作为月球表面最显著的地貌特征,其分布规律与形态特征记录了太阳系撞击历史的演变过程,是研究月球表面年代学、地质演化及撞击通量变化的重要依据^[1-4].尤其小尺度范围(直径<100 m)的撞击坑对月壤的形成与演化、表面空间环境及着陆区安全评估均具有至关重要的影响.因此,构建高分辨率、高完备性的小尺度撞击坑数据库已成为当前月球科学研究与探测任务支撑中的基础性课题,具有重要的科学和工程价值.

目前,相关领域的研究已经建立了多个全球性或区域性的月球撞击坑数据库.例如2019年Robbins^[5]基于LROC WAC (Lunar Reconnaissance Orbiter Camera, Wide Angle Camera)数据构建了涵盖上百万个直径 $\geq 1\sim 2$ km的撞击坑数据库;2021年Wang等^[6]结合DEM (Digital Elevation Model)数据与机器学习方法生成了包含约132万个撞击坑的全全球数据集;2025年La Grassa等^[7]发布了LU5 M812 TGT数据库,该数据库借助深度学习技术将撞击坑的尺度下限推进至直径400 m.然而,这些数据库主要聚焦于大、中尺度撞击坑,对直径小于100 m的小尺度撞击坑覆盖严重不足^[8].

为解决上述数据不足的问题,开展了特定着陆区高分辨率撞击坑的识别研究.2020年Jia等^[9]构建了嫦娥五号着陆区中直径大于200 m的撞击坑目录.2024年Wang等^[10]利用LROC NAC (Narrow Angle Camera)数据在嫦娥六号着陆区识别了直径 ≥ 30 m的770731个撞击坑.相对前面仅覆盖单个区域的撞击坑数据库,同年Yang等^[11]利用机器学习方法在10个着陆区识别了包含最小直径约5 m的大量撞击坑.这是当前已有的相对最全面的小撞击坑数据库,但仍存在一定局限性:数据空间覆盖范围仅为 $40\sim 50$ km²,未能完全涵盖巡视探测的实际范围;选择的研究区未能完全覆盖中国嫦娥系列任务的着陆区;对小撞击坑直径区间的完整性检测能力可能不足.

针对现有研究的不足,本文系统地构建了一个覆盖6个阿波罗任务与4个嫦娥任务共10个典型着陆区、直径 ≥ 15 m的高分辨率撞击坑数据库,本文从以下两个方面开展了创新性的工作.一是系统整合了多个任务着陆区,涵盖了月球正面与背面、月海与高地

等多种地质单元,识别的空间范围达 $20\text{ km}\times 20\text{ km}$,有效覆盖了巡视探测与采样返回的实际作业区域,为多地质背景对比研究提供了完备的数据基础.二是基于LROC NAC的高分辨率影像($0.5\sim 2\text{ m}\cdot\text{pixel}^{-1}$),对直径 ≥ 15 m的小撞击坑实现了高精度的识别,并对检测结果进行严格的人工校验,显著提升了小撞击坑编目的完整性和可靠性,弥补了现有数据库在小尺度撞击坑方面的数据不足.本文详细介绍了数据采集与处理方法、数据集样本描述、质量控制与评估流程,并基于撞击坑的密度与尺寸-频率分布揭示了不同着陆区的分布特征.本数据库的建立旨在为月球科学研究提供高质量的基础数据支持,推动中国行星科学数据的共享与应用.

1 数据采集与处理方法

1.1 数据采集方法

所建数据库包括Apollo 11~17与Chang'E-3~6共10个着陆区 $20\text{ km}\times 20\text{ km}$ 范围357764个撞击坑.该数据库中数据的主要采集方式为通过深度学习模型自动识别上述10个着陆区中的直径 ≥ 15 m的撞击坑.其中的核心技术为YOLO11+SAHI深度学习模型的构建,具体包括训练集构建、网络优化与模型训练三部分构成.

1.1.1 训练集构建

为了提升工作效率和深度学习模型的泛化能力,在构建训练集时并没有使用全部10个着陆区 $20\text{ km}\times 20\text{ km}$ 范围的全部影像,而是仅从中选出9个着陆区中的单轨影像数据制作数据集,并用于模型训练.这些数据均来自NASA发布的LROC NAC数据,其最高分辨率约为 $0.5\sim 2\text{ m}\cdot\text{pixel}^{-1}$ ^[12-14].在构建训练集时使用的是NAC RDR级别的数据,并对这些数据进行几何校正和地理投影.训练集中图像的质量是决定目标检测算法的关键因素之一^[15].因此,具有适宜光照条件的NAC图像有助于训练出性能优异、鲁棒性强的撞击坑检测模型^[16],而太高或太低的太阳入射角下拍摄的影像数据均不利于撞击坑自动识别模型的训练.基于已有的研究结果,选择覆盖不同地质背景条件的9个着陆区、入射角在 $50^\circ\sim 70^\circ$ 之间的单轨NAC图像构建了高质量的训练集^[17,18],具体信息列于表1.使用Labelbee软件*完成数据集

* <https://github.com/open-mmlab/labelbee>

的标注, 其中包含丰富的小撞击坑样本, 这对于小尺度撞击坑自动识别模型的训练至关重要^[19].

1.1.2 网络优化

聚焦于提升在月球高分辨率遥感图像中直径 <100 m 的小撞击坑自动检测能力, 特别是在复杂地形和撞击坑密集区域的撞击坑识别能力. 为此, 设计了一种基于 YOLO11 (You Only Look Once v11) 模型并集成切片辅助超推理 (Slicing-Aided Hyper Inference, SAHI) 策略的自动识别方法, 其网络模型结构如图 1 所示.

从图 1 可以看出, YOLO11 在网络结构上的创

新显著提升了其在月球撞击坑检测任务上的精度与效率, 其核心改进体现在 C3 K2 模块、空间金字塔快速池化 (SPFF) 模块和 C2 PSA 模块的协同设计. C3 K2 模块通过引入核尺寸为 2 的跨阶段部分连接, 在降低计算复杂度的同时增强了局部特征的提取能力, 尤其适用于月球表面撞击坑边缘的微小纹理捕捉^[20,21]. SPFF 模块采用多尺度特征融合策略, 有效解决了撞击坑因光照条件差异导致的尺度敏感性问题. 实验表明, 该模块使小尺度撞击坑 (直径 <50 m) 的检测召回率提升了 23.6%^[20,22]. C2 PSA 模块则通过多次叠加 PSA (金字塔切片注意力) 空间注意力机制、

表 1 构建训练集时选择的 9 个典型着陆区域的 LRO NAC 影像信息

Table 1 LRO NAC imagery information for the 9 landing regions where sample collection was conducted

着陆区	数据ID	分辨率/(m·pixel ⁻¹)	太阳入射角/(°)	相位角/(°)
A11	M117338434 R	0.475	74.41	83.35
A12	M162466771 L	0.497	76.27	87.18
A14	M162426054 L	0.477	76.07	77.75
A15	M119829425 L	0.501	58.21	41.45
A16	M177535538 L	0.469	69.64	73.36
A17	M162107606 L	0.478	73.86	77.58
CE3	M1152001999 R	1.357	56.56	55.65
CE4	M1303619844 R	1.400	62.64	67.2
CE5	M1428381550 L+R	0.905	59.91	61.43

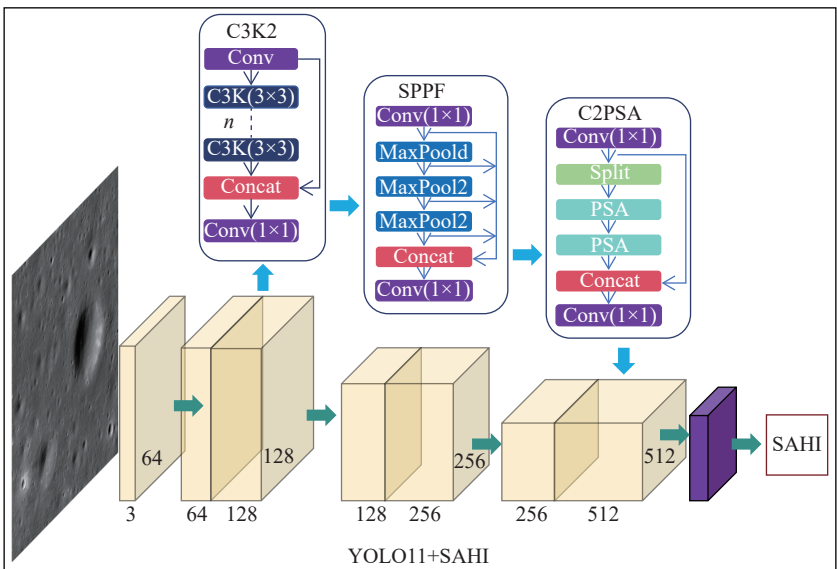


图 1 YOLO11+SAHI 模型的网络结构

Fig. 1 Architecture diagram for YOLO11+SAHI network

并对特征图空间位置进行动态加权以提高目标检测能力. 在嫦娥六号着陆区复杂地形数据的检测案例中, C2 PSA 模块成功地将玄武岩基质区域的误检率降低至 4.8% 以下^[23].

虽然 YOLO11 在网络结构方面做了较大提升, 但其作为一个目标检测模型在处理高分辨率图像中的小目标和密集目标时仍面临挑战, 容易出现漏检^[24]. 为了解决这一问题, 本文在 YOLO11 的检测端引入了 SAHI (Slicing Aided Hyper Inference) 策略^[24]. SAHI 是一种通用的辅助推理框架, 其通过将大尺寸图像分割成更小的、重叠切片, 然后对每个切片独立进行目标检测, 最后将各切片的检测结果进行合并和后处理, 从而有效提升模型对小尺寸目标的检测性能^[25]. 在高分辨率遥感图像中, 由于目标 (例如撞击坑) 在图像中所占的像素比例较小, 采用 SAHI 策略可以将这些小目标在切片中放大, 使其更容易被模型检测到^[25]. 本文将 SAHI 模块集成到 YOLO11 的检测头后, 可以在保持 YOLO11 整体高效性的同时, 显著提高其对高分辨率遥感图像中直径 <100 m 的撞击坑的检测能力. 在复杂地形和撞击坑密集分布的区域, 这种性能的提升表现得尤为明显. 这种方法克服了传统深度学习方法难以有效处理小尺寸和密集分布目标的局限性, 为着陆区撞击坑的自动化识别提供了有效的技术方案.

YOLO 目标检测算法输出的目标坐标格式通常为归一化的中心坐标格式. 每个检测到的目标用 T_{box} 表示, 有

$$T_{\text{box}} = (\text{cls}, x, y, w, h).$$

其中, cls 为目标所属类别的标识符, (x, y) 为边界框中心点的归一化坐标, (w, h) 为边界框的归一化宽度和高度. 归一化坐标 (取值范围为 0~1) 使检测结果与原始图像分辨率无关, 从而便于处理不同尺寸的图像.

这里利用 Python 结合 GDAL 库的方式, 将目标在图像上的相对位置坐标转换为绝对地理坐标. 鉴于 YOLO 输出的目标表示为一个矩形边界框, 在最终的可视化结果中可以将其转换为圆形进行绘制. 转换规则如下.

(1) 圆形的圆心对应矩形边界框的中心点坐标为 (x, y) .

(2) 圆形的直径取矩形边界框的归一化宽度 w 和高度 h 的平均值, 即 $(w + h) / 2$.

1.1.3 模型训练与测试

在模型训练过程中, 将模型的初始学习率设置为 0.01, 将权重衰减设置为 0.0005, 将训练优化器设置为 auto, 以实现基于模型配置的自动选择, 将 batch, imgsiz 和 epochs 分别设置为 16, 640 和 400. 模型训练完成后, 对其识别结果进行了初步的测试与分析. 由于月球表面的地形特征表现为明显的二分性, 即分别为月海和高地. 月海的地势起伏较小, 而高地的地势起伏较大. 月海和高地分别存在反照率不同的区域和不同尺度的撞击坑. 因此, 选取不同地形和不同光照条件的 3 个区域来显示模型的识别效果 (见图 2), 图 2(a)~(c) 为数据集真值 (GT), (d)~(f) 为模型的预测值 (Pred). 从图 2 可以看出, 本模型在不同环境下对各类撞击坑的识别效果均较好, 因此将该模型应用于 10 个着陆区全部撞击坑数据的提取工作.

1.2 数据预处理

本文的主要目标是构建月球 10 个着陆区 20 km×20 km 范围的撞击坑数据库, 但由于 NAC RDR 级别的影像数据在这 10 个着陆区上的覆盖范围不全, 因此, 这里采用了未进行校正和地理投影的 NAC EDR 级别数据用于撞击坑识别以及数据库构建. 为了将原始 EDR 影像转化为可用于撞击坑提取的数据, 研究使用 ISIS3 软件进行了几何校准和地图投影处理. 由于不同轨 NAC 数据存在几何偏差, 使用 SELENE/Kaguya TC 数据作为底图对处理后的 NAC 影像进行了几何配准^[12]. 其配准误差控制在一个像素以内, 即优于所使用的 NAC 影像自身分辨率 ($0.5 \sim 2 \text{ m} \cdot \text{pixel}^{-1}$). 该精度要求是为了保证撞击坑地理定位的准确性, 使其位置不确定度小于影像的地面采样间隔 ($0.5 \sim 2 \text{ m}$). 该精度足以保证后续撞击坑识别与定位的准确性. 为了消减因投影导致的影像几何形变问题, 将影像投影到以各自着陆点为中心的等距离投影下. 最终利用这些影像生成了着陆点周围 20 km×20 km 范围的高分辨率镶嵌影像数据, 其经纬度范围如表 2 所示. 后续本文将使用这些预处理好的影像数据进行撞击坑的信息提取.

2 数据样本描述

2.1 撞击坑数据库制作

为了规范化和便于后续分析, 将识别出的撞击坑信息 (包括位置、直径、检测置信度等) 整理并存储于

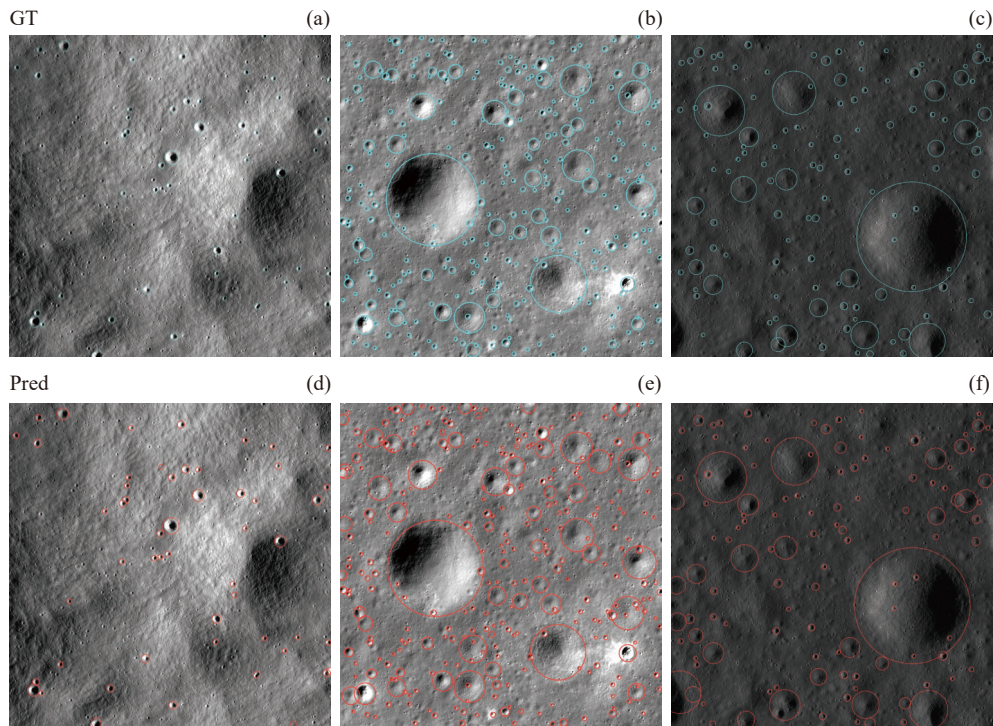


图 2 选取的具有不同地形与不同光照条件的三个区域. GT 表示真值, Pred 表示预测值. (a)(d) 反照率较高的高地区域, 整体撞击坑尺寸较小; (b)(e) 反照率较低的月海区域, 撞击坑尺度为中等大小; (c)(f) 反照率较低的月海区域, 存在一些大尺度的撞击坑

Fig. 2 Schematic images with different illumination and terrain character. (a) (d) High-albedo elevated regions with generally smaller crater sizes. (b) (e) Low-albedo lunar mare regions with medium-sized impact craters. (c) (f) Low-albedo lunar mare regions containing some large-scale impact craters

表 2 10 个着陆区的镶嵌数据集范围
Table 2 Mosaic dataset coverage of ten landing regions

着陆区	纬度范围	经度范围
A11	0°20'40.75"—1°0'15.04"	23°8'35.54"—23°48'10.06"
A12	−3°20'32.49"—−2°40'58.20"	−23°45'7.82"—−23°5'30.18"
A14	−3°58'30.17"—−3°18'55.88"	−17°48'6.48"—−17°8'27.31"
A15	25°48'8.91"—26°27'43.20"	3°15'59.58"—4°0'4.21"
A16	−9°18'11.21"—−8°38'36.92"	15°10'2.43"—15°50'6.22"
A17	19°51'39.79"—20°31'14.08"	30°25'13.16"—31°7'22.93"
CE3	43°47'28.51"—44°27'2.80"	−19°58'15.06"—−19°3'7.74"
CE4	−45°47'14.84"—−45°7'40.55"	177°7'10.91"—178°3'36.00"
CE5	42°43'33.08"—43°23'7.37"	−52°22'8.01"—−51°27'58.71"
CE6	−41°58'4.97"—−41°18'30.68"	−154°25'36.71"—−153°32'39.61"

一个统一的数据库中. 基于本文得到的 LROC NAC 高分辨率影像镶嵌数据, 使用训练好的 YOLO11+ SAHI 模型对每个着陆点周围 20 km×20 km 范围的影像进行撞击坑识别. 由于自动化检测模型可能存在

撞击坑的漏检和误检现象, 为了保证数据集的准确性, 本研究基于以下标准和流程对数据集中的目标进行人工校验.

(1) 校验标准. 目标具有基本闭合、连续的环形

边缘结构; 坑内与周边区域在纹理或阴影上应存在可辨识差异; 目标在图像中的像素尺寸不小于 15 pixel, 以确保数据集覆盖所有物理直径 ≥ 15 m 的撞击坑; 排除由图像噪声、岩石阴影、或相邻坑重叠边界模糊所造成的疑似目标。

(2) 校验流程. 由至少两名经过培训的判读人员独立对自动识别结果进行校验; 初校后, 对存在争议或判断不一致的撞击坑, 由项目组资深研究员进行仲裁, 形成最终裁定; 最终, 对所有裁定后的撞击坑进行统一的编号与属性录入。

使用 YOLO11+SAHI 模型在表 2 所示的 10 个着陆区中共识别出 358762 个候选撞击坑. 经人工校验与修正后, 最终建成包含 357764 条记录的撞击坑数据库. 由于模型本身准确率高 (平均 F1-score 0.962), 人工干预比例约为 0.3%, 最后得到的撞击坑标记如图 3 所示. 图中绿色线表示研究区包含多个地质单元时着陆点所在地质单元的边界^[26].

由于 Apollo 15, Apollo 17, Chang'E-3 和 Chang'E-4 等部分着陆区所包括的区域比较大, 覆盖了不同

年代的地质单元, 这有可能使后续的分析变得更复杂. 为减少不同地质单元带来的干扰, 本文在进行后续分析的时候, 仅考虑着陆点所在的地质单元, 如图 3 中绿色线所标注的范围。

2.2 撞击坑数量统计

设定撞击坑直径下限为 15 m. 此阈值主要基于所用 LROC NAC 影像的分辨率 ($0.5 \sim 2 \text{ m} \cdot \text{pixel}^{-1}$), 旨在确保有足够的像元数量呈现撞击坑的环形结构, 并测量直径. 为更好显示不同区间内撞击坑数量占比, 使用自然间断点分级法 (Jenks Natural Breaks, Jenks) 对所有编目的着陆区撞击坑直径进行分段, 并统计了不同区间的撞击坑数量. Jenks 算法通过迭代优化自动寻找数据中固有的断点, 因此能识别出月球撞击坑的小坑密集区、中等坑过渡区、大坑稀疏区之间的自然分界, 避免出现因人为设定固定间隔 (例如等间隔法) 而出现的小坑区间隔过小失去区分度、大坑区间隔过大掩盖细节等问题. 方法中无严格的数学理论依据设定器分类数 k 值, 只是依据经验和数据对 k 值进行人工确定. 由于所研究的月球撞击坑直径跨

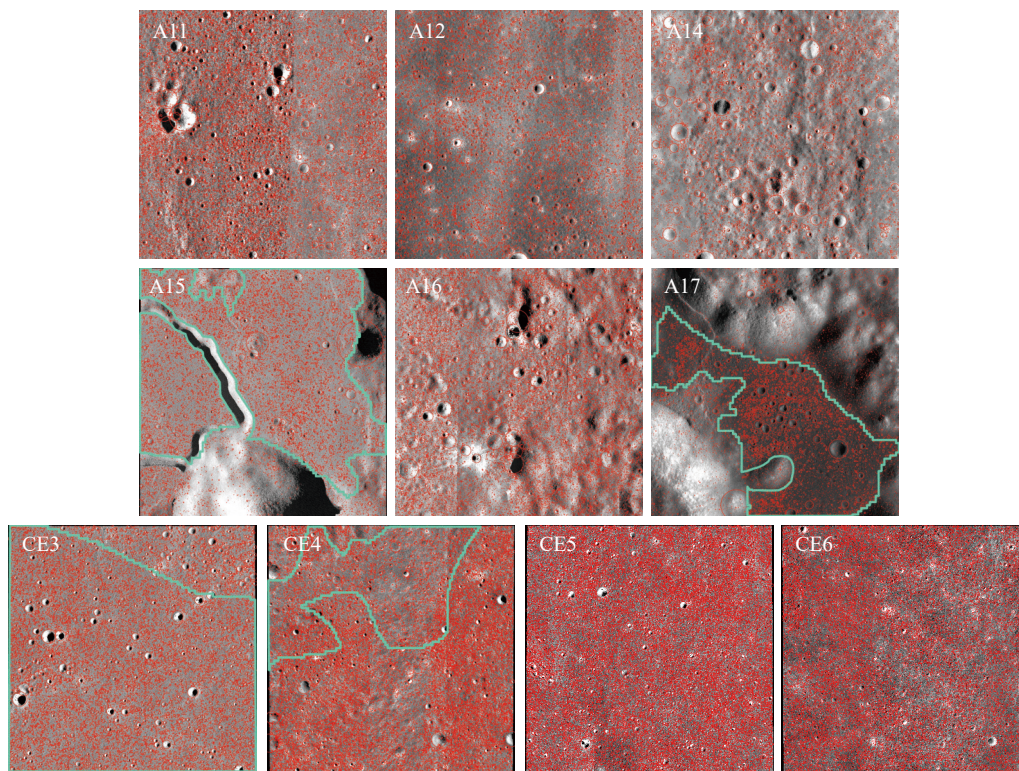


图 3 10 个着陆区的撞击坑标记 (绿色线表示研究区包含多个地质单元时着陆点所在地质单元的边界)

Fig. 3 Impact crater maps density maps for ten landing regions (The green line indicates the boundary of the geological unit where the landing site is located when the study area contains multiple geological units)

度较大, 且小坑数量较多、大坑数量较少 (符合幂律分布), 因此需要较多的类别来捕捉小坑范围内的精细差异. 经过系列测试后, 选择分类数为 8. 使用 Jenks 方法对 10 个着陆区的撞击坑进行分段划分的流程如下.

- (1) 将撞击坑直径数据按从小到大排序.
 - (2) 设定分类数 $k = 8$.
 - (3) 基于 Jenks 算法, 使用 ArcGIS 软件寻找 7 个断点值 (因为 8 个区间需要 7 个分界点).
 - (4) 算法会遍历多种可能的 7 个断点组合, 计算每种组合下的总组内平方和 (Total Sum of Squares Within, TSSW).
 - (5) 最终选择使 TSSW 最小的 7 个断点值, 这些值即为自然间断点.
- 对于上述 10 个着陆区内编目的撞击坑, 每个着

陆区识别的撞击坑直径区间和区间内的撞击坑数量分布如表 3 所示.

Jenks 分类结果显示, 在 10 个着陆区内, 撞击坑数量随直径增大呈显著的幂律衰减趋势. 以 Apollo 11 着陆区为例, 直径最小的区间 (15~26 m) 包含 29397 个撞击坑, 而直径最大的区间 (560~1534 m) 仅包含 34 个, 数量相差超过 3 个数量级. 其他着陆区也表现出类似的规律, 具体数据见表 3.

3 数据质量控制与评估

3.1 撞击坑检测性能

将标注好的数据集划分为训练集、验证集和测试集, 其包含图像的比例为 7:1:2. 本文使用测试集对 YOLO11+SAHI 模型在每个着陆区上的识别性能,

表 3 使用自然间断点分级法获得的 10 个着陆区各区间撞击坑直径区间划分以及区间内的桩径数量统计
Table 3 Statistics on the number of impact craters in each section of ten landing regions using Jenks methods

着陆区	区间划分及撞击坑分区数量								总数量
A11	15~26	26~46	46~77	77~130	130~205	205~319	318~562	562~2240	39024
	29438	6116	1764	828	414	271	158	35	
A12	15~21	21~30	30~44	44~63	63~93	93~147	147~387	387~1610	19355
	9258	4267	1921	1421	1227	805	421	35	
A14	15~21	21~32	32~49	49~76	76~141	141~272	272~472	472~1641	17862
	10025	4409	1918	840	418	133	55	64	
A15	15~24	24~36	36~54	54~83	83~128	128~206	206~437	437~3741	27309
	17678	5486	2219	1034	499	277	104	12	
A16	15~27	27~44	44~72	72~121	121~197	197~320	320~615	615~1818	46889
	34608	8971	2196	712	226	83	48	45	
A17	15~22	22~32	32~50	50~79	79~124	124~201	201~355	355~2489	19066
	11628	4355	1808	525	367	237	102	44	
CE3	15~31	31~57	57~93	93~143	143~208	208~305	305~504	504~1216	31497
	20297	5537	3310	1278	659	269	113	34	
CE4	15~26	26~43	43~68	68~100	100~142	142~251	251~515	515~1484	57259
	33801	12529	6478	2689	1203	361	180	18	
CE5	15~36	36~60	60~96	96~140	140~201	201~297	297~447	447~966	56545
	40650	8036	4980	1930	640	245	54	10	
CE6	15~30	30~47	47~71	71~98	98~133	133~198	198~338	338~853	42958
	25470	11247	3018	1607	924	464	194	34	

注 第一列为着陆区名称, 每个着陆区对应数据的第一行为Jenks划分直径区间 (单位: m), 第二行为各区间内统计的撞击坑数量, 最后一竖列为该着陆区内撞击坑总数量.

使用的评估指标为精确率 (σ_{pre})、召回率 (σ_{rec}) 和 F1 分数 (F_1).

精确率用来衡量模型识别出的撞击坑中有多少是真实的撞击坑, 其计算公式为

$$\sigma_{\text{pre}} = \frac{T_p}{T_p + F_p}. \quad (1)$$

其中, T_p 为模型识别结果中, 经专家判定为正确识别为真实撞击坑的目标数量; F_p 为模型识别结果中, 被专家判定为误检的非撞击坑的目标数量, 即为误检的数量; 精确率 Precision 表征了模型识别结果的准确性.

召回率衡量实际撞击坑中被模型成功识别的比例, 其计算公式为

$$\sigma_{\text{rec}} = \frac{T_p}{T_p + F_n}. \quad (2)$$

其中, T_p 与式 (1) 中的 T_p 相同, F_n 为由专家标注的数据集中未被模型识别的撞击坑的数量, 即漏检的数量. 召回率表示模型识别目标的完备性.

在目标检测领域, 通常情况下, 精确率和召回率难以兼得. 因此引入一个新指标 F_1 , 即精确率和召回率的调和平均数, 综合反映模型的检测性能. 其计算公式为

$$F_1 = \frac{2T_p}{2T_p + F_p + F_n}. \quad (3)$$

如表 4 所示, 分别统计模型在 9 个着陆点的测试集上的识别结果. 这些统计结果显示, 使用的

YOLO11+SAHI 模型在 9 个着陆区测试集上的各项指标均在 90% 以上. 因为这 9 个着陆区的地质背景差异较大, 因此这些结果表明, 模型在各种复杂场景下均具有较好的鲁棒性和泛化能力.

研究针对月球高分辨率影像中密集分布的小尺度撞击坑检测难题, 通过引入融合 C3 K2、SPFF 与 C2 PSA 模块的改进型 YOLO11 网络, 显著增强了对坑体边缘纹理与多尺度特征的提取能力; 同时, 结合 SAHI 切片推理策略, 有效缓解了小目标在高分辨率影像中的漏检问题. 该方法在多个地质背景各异的着陆区均表现出高精度与强泛化性, 实现了对 20 km×20 km 大范围区域的高效、自动化处理, 为构建大规模、高可靠性撞击坑数据库提供了创新的技术解决方案.

3.2 完整性分析

完整性直径是评估撞击坑数据库统计完备性的关键指标. 对于某一特定研究区域, 在该区域内撞击坑数量最多的直径区间之后, 沿直径区间增大的方向, 下一个撞击坑数量最多的直径区间的中间值, 即为该区域撞击坑的完整性直径^[5]. 图 4 给出了 10 个着陆区的撞击坑大小频率的增量分布, 图中的竖线代表了每个着陆区对应的完整性直径.

表 5 汇总了本文在 10 个着陆区编码的直径 ≥ 15 m 的撞击坑的完整性直径数值. 从表 5 的数据可以看出, 构建的数据库所包含的各着陆区撞击坑的完整性直径集中在 20~30 m 之间. 随后通过比较大于

表 4 参与训练的 9 个着陆点数据集详细信息及模型在其测试集上的准确率
Table 4 Detailed information on the nine landing site datasets used in training
and the model's accuracy on their test sets

着陆区	数据ID	标签数量	精确率	召回率	F1分数
A11	M117338434 R	1178	0.964	0.933	0.948
A12	M162466771 L	1131	0.938	0.902	0.919
A14	M162426054 L	1321	1	0.919	0.958
A15	M119829425 L	1282	0.993	0.939	0.965
A16	M177535538 L	3073	0.985	0.996	0.990
A17	M162107606 L	1823	1	0.964	0.982
CE3	M1152001999 R	1279	1	0.963	0.981
CE4	M1303619844 R	20785	0.993	0.915	0.952
CE5	M1428381550 L+R	5598	0.988	0.929	0.959
均值			0.985	0.94	0.962

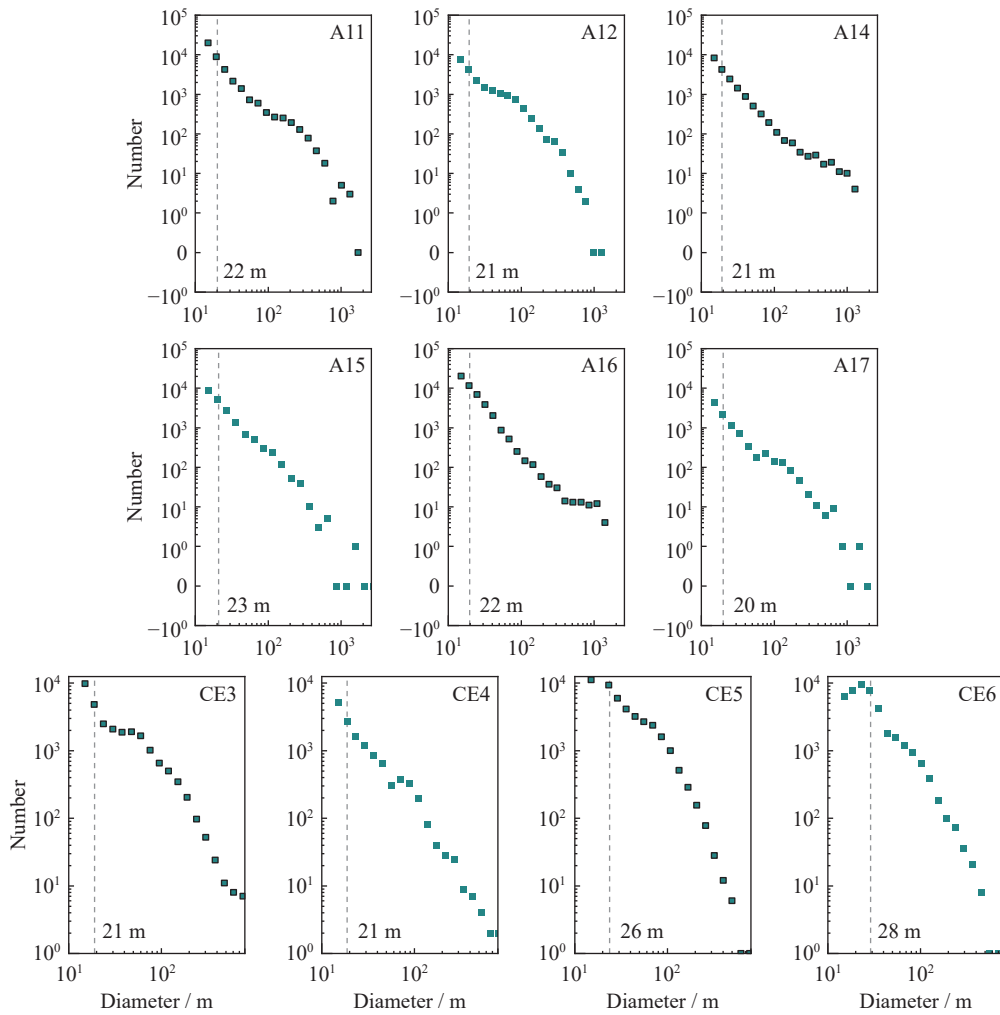


图 4 10 个着陆区中的撞击坑大小频率累计分布 (垂直虚线表示完整性直径的位置)

Fig. 4 Cumulative frequency distribution of crater sizes across ten landing regions (The vertical dashed lines indicate the positions of the integrity completeness diameter)

完整性直径的撞击坑数量, 发现本文的数据库在完备性上优于之前文献中的月表撞击坑数据库。

4 数据价值

数据集提供了涵盖月球正面与背面、月海与高地等多种地质背景的 10 个典型着陆区的高完备性撞击坑编目数据。为显示其在月球科学研究中的直接应用潜力与价值, 基于本数据集的统计结果, 通过撞击坑密度与尺寸-频率分布分析, 可以揭示不同地质背景下撞击坑的分布规律, 并探讨其背后的地质演化意义。

4.1 撞击坑密度分布

撞击坑面密度 (单位面积内的撞击坑数量) 是估算月球表面绝对模式年龄的关键参数^[8]。利用本数

据集, 以 50 m×50 m 为网格、100 m 为搜索半径, 可以计算各着陆区 (着陆点所在主要地质单元) 的撞击坑核密度分布 (见图 5), 统计其平均面密度 (见表 6)。

通过对密度差异与已知地质背景 (见表 6) 的综合分析, 本数据分布特征表明, 撞击坑密度与地质单元年龄可能存在复杂的规律。例如, 位于古老笛卡尔高地的 Apollo 16 着陆区 (年龄 > 3.7 Ga), 其撞击坑密度最高 (117.22 km^{-2}); 而位于相对年轻的风暴洋月海平原的 Apollo 12 着陆区 (年龄约 3.2 Ga), 密度则较低 (51.49 km^{-2})。这一对比清晰印证了越古老的表面, 撞击坑密度通常越高的基本规律。然而嫦娥五号返回样品表明该着陆区拥有月球已知最年轻的玄武

表 5 10 个着陆区检测到的撞击坑直径统计信息

Table 5 Statistical information on crater diameters detected across ten landing regions

着陆区	最小直径/m	最大直径/m	完整性直径/m	大于完整性直径的数量
A11	15	2239	22	14389
A12	15	1609	21	10023
A14	15	1641	21	7971
A15	15	3740	23	10808
A16	15	1817	22	19990
A17	15	2488	20	9184
CE3	15	1216	21	16975
CE4	15	1483	21	31203
CE5	15	965	26	26242
CE6	15	852	28	19618

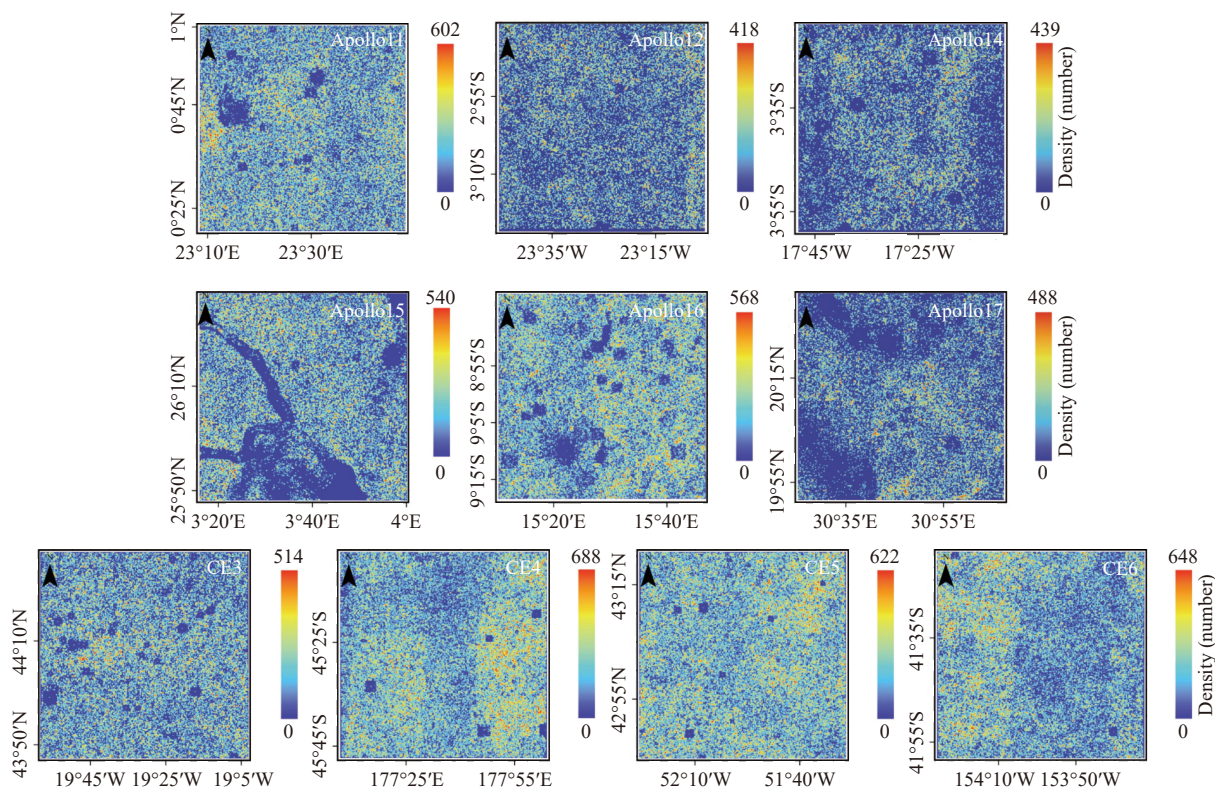


图 5 10 个着陆区撞击坑核密度分布

Fig. 5 Kernel density distribution of impact craters in ten landing regions

岩之一 (约 2.0 Ga), 但本数据集计算出的撞击坑密度 (141.36 km^{-2}) 却相对较高. 这一现象揭示了年轻月海表面的复杂性: 密集的小尺度撞击坑可能反映了其形成以来持续的撞击累积、较高的二次撞击坑贡献^[27], 或表明该表面尚未达到撞击平衡状态. 本数据集为量化这类复杂区域的撞击记录、进而优化年龄解

释模型提供了关键数据基础.

4.2 撞击坑尺寸-频率分布

撞击坑的尺寸-频率分布 (Crater Size-Frequency Distribution, CSFD) 包含了丰富的表面演化信息. 基于本数据集, 绘制各着陆区累积尺寸-频率分布 (见图 6) 和相对尺寸-频率分布 (R 图, 见图 7). 累积分布

的斜率反映表面年龄和撞击历史, 而 R 图的峰值则能指示特定尺寸撞击坑的富集, 常与二次撞击坑群或局部改造事件相关.

经过分析, 数据集能有效区分不同成因的撞击坑种群. 例如 Apollo 12 着陆区的 R 图在约 300 m 处出现明显峰值 (见图 7a), 结合其地质背景, 这很可能指示了来自哥白尼等撞击坑的二次撞击坑群富集^[13]. 相反, 古老的 Apollo 16 高地着陆区, 其 R 图峰值出现在大直径 (>1 km) 的范围内, 且小直径区间无显著峰

值 (见图 7a), 因此二次坑存在证据较弱. 对于嫦娥系列着陆区, 嫦娥三号着陆区 R 图在 300~500 m 范围内的峰值可能与其位于雨海溅射覆盖区有关^[33]; 而嫦娥五号着陆区 R 图在约 100 m 处的峰值, 可能揭示该年轻月表上活跃的小尺度撞击改造过程或局部二次坑群的存在^[34]. 这些分析凸显了本数据集在解析局部撞击历史、识别二次撞击效应及约束表面改造过程方面的独特价值.

月球撞击坑的识别和计数正越来越多地采用机

表 6 10 个着陆点所在地质单元的撞击坑密度和地质背景
Table 6 Impact crater density and geological background of the address units
where the ten landing sites are located

着陆区	单个地质单元 面积/km ²	单个地质单元 撞击坑数量	撞击坑 密度/km ⁻²	年龄/Ga	地质单元	地质单元起伏情况
A11	400	39176	97.94	3.58~3.85 ^[28]	静海	平坦
A12	400	20595	51.49	3.15~3.22 ^[28]	风暴洋月海平原	平坦
A14	400	18564	46.41	3.77~3.85 ^[28]	弗拉·摩罗高地	崎岖
A15	247	20331	82.31	3.28~3.33 ^[28]	靠近盆缘高地的月海平原	平坦
A16	400	46889	117.22	3.77~3.85 ^[28]	笛卡尔高地	崎岖
A17	139	9560	68.78	3.5~3.85 ^[28]	靠近高地的月海平原	相对平坦
CE3	342	27685	80.95	2~3 ^[29]	虹湾月海平原	相对平坦
CE4	117	13653	116.69	2.95~3.67 ^[30]	被溅射物覆盖的月海平原	相对崎岖
CE5	400	56545	141.36	2.03 ^[31]	月海平原	相对平坦
CE6	400	42966	107.42	2.83 ^[32]	月海平原	平坦

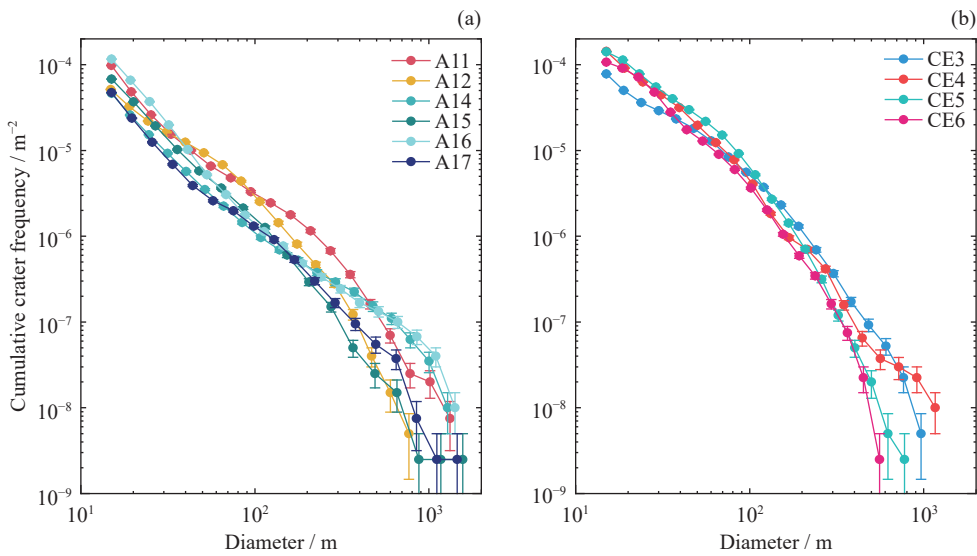


图 6 10 个着陆区的撞击坑直径-频率的累积分布. (a) Apollo 11~17, (b) Chang'E-3~6
Fig. 6 Cumulative distribution plots of crater diameter versus frequency for ten landing regions.
(a) Apollo 11~17, (b) Chang'E-3~6

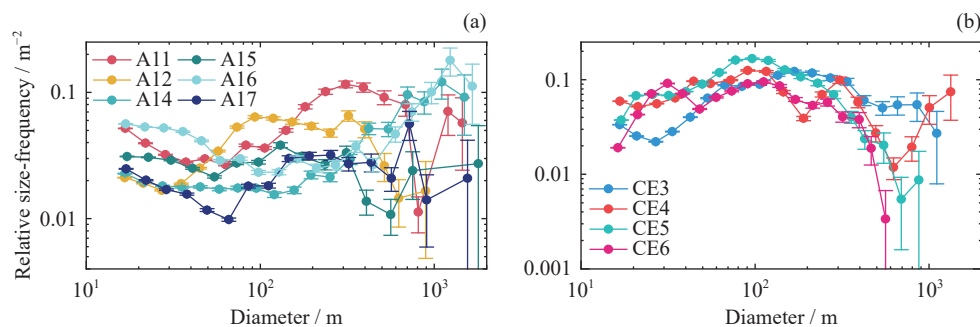


图7 10个着陆区的撞击坑直径-频率的R分布. (a) Apollo 11~17, (b) Chang'E-3~6

Fig. 7 R-plots of crater diameter-frequency distributions for ten landing regions.

(a) Apollo 11~17, (b) Chang'E-3~6

器学习和深度学习算法. 本文所提供的覆盖不同地质背景的阿波罗和嫦娥着陆区的数据集经过准确的机器识别和质量控制, 因此行星探测领域可以用这些参考数据集作为训练数据来训练人工智能模型, 然后再将其应用到月球全球或其他行星的撞击坑识别工作中, 能够极大地提高研究效率, 促进科学产出.

本数据集可以用于进一步优化月球的撞击坑统计定年函数. 在目前领域内常用的撞击坑定年法中, 其产率函数和年代函数的建立主要来自阿波罗任务所获得的地质年龄数据和撞击坑人工解译结果. 本研究建立的撞击坑数据集来自高分辨率的遥感数据, 在撞击坑识别和直径测量上更为准确, 可以为定年相关函数的优化提供更优质的数据.

5 数据使用方法与建议

本文推荐使用 ArcGIS 等地理信息系统软件进行数据的浏览、分析和深加工.

6 数据可用性声明

本数据库已于线上完全公开、共享, 用户可通过国家科技资源共享服务平台 (国家空间科学数据中心) 下载, 具体信息见本文的数据库 (集) 基本信息简介.

利益冲突 所有作者均声明不存在利益冲突.

参考文献

- [1] XIAO Z Y, DI K C, XIE M G, *et al.* Impact flux on the moon[J]. *Space: Science & Technology*, 2024, **4**(4): 0148
- [2] SHYLAJA B S. Determination of lunar surface ages from crater frequency—size distribution[J]. *Journal of Earth*

System Science, 2005, **114**(6): 609-612

- [3] CHEN D, HU F, ZHANG L Q, *et al.* Impact crater recognition methods: a review[J]. *Science China Earth Sciences*, 2024, **67**(6): 1719-1742
- [4] MICHAEL G G, NEUKUM G. Planetary surface dating from crater size-frequency distribution measurements: partial resurfacing events and statistical age uncertainty[J]. *Earth and Planetary Science Letters*, 2010, **294**(3-4): 223-229
- [5] ROBBINS S J. A new global database of lunar impact craters >1-2 km: 1. Crater locations and sizes, comparisons with published databases, and global analysis[J]. *Journal of Geophysical Research: Planets*, 2019, **124**(4): 871-892
- [6] WANG Y R, WU B, XUE H O, *et al.* An improved global catalog of lunar impact craters (≥ 1 km) with 3D morphometric information and updates on global crater analysis[J]. *Journal of Geophysical Research: Planets*, 2021, **126**(9): e2020JE006728
- [7] LA GRASSA R, MARTELLATO E, CREMONESE G, *et al.* LU5M812TGT: an AI-powered global database of impact craters ≥ 0.4 km on the Moon[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2025, **220**: 75-84
- [8] ZHAO D D. Intelligent Identification and Spatial Distribution Analysis of Small Craters in Lunar Landing Area[D]. Changchun: Jilin University, 2022 (赵丹冬. 月球着陆区小型撞击坑智能识别与空间分布特征分析 [D]. 长春: 吉林大学, 2022)
- [9] JIA M N, YUE Z Y, DI K C, *et al.* A catalogue of impact craters larger than 200 m and surface age analysis in the Chang'E-5 landing area[J]. *Earth and Planetary Science Letters*, 2020, **541**: 116272
- [10] WANG Y X, NAN J, ZHAO C X, *et al.* A catalogue of impact craters and surface age analysis in the Chang'E-6 landing area[J]. *Remote Sensing*, 2024, **16**(11): 2014
- [11] YANG C, WANG X L, ZHAO D D, *et al.* Accurate mapping and evaluation of small impact craters within the lunar landing area[J]. *Remote Sensing*, 2024, **16**(12): 2165
- [12] IQBAL W, HIESINGER H, BORISOV D, *et al.* Geological mapping and chronology of lunar landing sites: Apollo 14[J]. *Icarus*, 2023, **406**: 115732

- [13] IQBAL W, HIESINGER H, VAN DER BOGERT C H. Geological mapping and chronology of lunar landing sites: Apollo 12[J]. *Icarus*, 2020, **352**: 113991
- [14] CADOGAN P H. Automated precision counting of small lunar craters - a broader view[J]. *Icarus*, 2024, **408**: 115796
- [15] DELATTE D M, CRITES S T, GUTTENBERG N, *et al*. Automated crater detection algorithms from a machine learning perspective in the convolutional neural network era[J]. *Advances in Space Research*, 2019, **64**(8): 1615-1628
- [16] FAIRWEATHER J H, LAGAIN A, SERVIS K, *et al*. Automatic mapping of small lunar impact craters using LRO-NAC images[J]. *Earth and Space Science*, 2022, **9**(7): e2021EA002177
- [17] HEAD III J W, FASSETT C I, KADISH S J, *et al*. Global distribution of large lunar craters: implications for resurfacing and impactor populations[J]. *Science*, 2010, **329**(5998): 1504-1507
- [18] CADOGAN P H. Automated precision counting of very small craters at lunar landing sites[J]. *Icarus*, 2020, **348**: 113822
- [19] YANG J, ZHANG S, LI L, *et al*. Topographic knowledge-aware network for automatic small-scale impact crater detection from lunar digital elevation models[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2024, **129**: 103831
- [20] KHANAM R, HUSSAIN M. YOLOv11: An overview of the Key Architectural Enhancements[OL]. arXiv preprint arXiv: 2410.17725, 2024
- [21] ZHUO S L, BAI H, JIANG L F, *et al*. SCL-YOLOv11: A lightweight object detection network for low-illumination environments[J]. *IEEE Access*, 2025, **13**: 47653-47662
- [22] WANG G R, CHEN S Y, HU G, *et al*. Detection algorithm of abnormal flow state fluid on closed vibrating screen based on improved YOLOv5[J]. *Engineering Applications of Artificial Intelligence*, 2023, **123**: 106272
- [23] NAN J, WANG Y X, DI K C, *et al*. YOLOv8-LCNET: an improved YOLOv8 automatic crater detection algorithm and application in the Chang'E-6 landing area[J]. *Sensors*, 2025, **25**(1): 243
- [24] AKYON F C, ALTINUC S O, TEMIZEL A. Slicing aided hyper inference and fine-tuning for small object detection[C]//Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP). Bordeaux: IEEE, 2022: 966-970
- [25] ZHANG H, HAO C Y, SONG W R, *et al*. Adaptive slicing-aided hyper inference for small object detection in high-resolution remote sensing images[J]. *Remote Sensing*, 2023, **15**(5): 1249
- [26] JI J Z, GUO D J, LIU J Z, *et al*. The 1: 2, 500, 000-scale geologic map of the global moon[J]. *Science Bulletin*, 2022, **67**(15): 1544-1548
- [27] QIAN Y Q, XIAO L, ZHAO J W, *et al*. First magnetic and spectroscopic constraints on attenuated space weathering at the Chang'E-5 landing site[J]. *Icarus*, 2024, **410**: 115892
- [28] Meyer C. Lunar Sample Compendium[R]. Washington: NASA, 2005
- [29] LI Bo, LING Zongcheng, ZHANG Jiang, *et al*. Geochronology, petrogenesis and geological significance of the lunar basalts around CE-3 landing site[J]. *Acta Petrologica Sinica*, 2016, **32**(1): 19-28 (李勃, 凌宗成, 张江, 等. 嫦娥三号着陆区月海玄武岩的年龄、成因及地质意义 [J]. 岩石学报, 2016, **32**(1): 19-28)
- [30] YANG Meiping, YUE Zongyu, DI Kaichang, *et al*. Statistical analysis of secondary craters in the Chang'E-4 landing area based on panoramic camera data[J]. *Bulletin of Mineralogy, Petrology and Geochemistry*, 2021, **40**(3): 720-729 (杨梅萍, 岳宗玉, 邸凯昌, 等. 基于全景相机数据的嫦娥四号着陆区次级坑统计分析 [J]. 矿物岩石地球化学通报, 2021, **40**(3): 720-729)
- [31] LI Q L, ZHOU Q, LIU Y, *et al*. Two-billion-year-old volcanism on the moon from Chang'E -5 basalts[J]. *Nature*, 2021, **600**(7887): 54-58
- [32] CUI Z X, YANG Q, ZHANG Y Q, *et al*. A sample of the Moon's far side retrieved by Chang'E -6 contains 2.83-billion-year-old basalt[J]. *Science*, 2024, **386**(6728): 1395-1399
- [33] LIU S C, DU K C, TONG X H, *et al*. In-situ mapping of iron and titanium with the Visible and Near-infrared Image Spectrometer (VNIS) along the Yutu-2 rover traverse on the far side of the moon[J]. *Icarus*, 2024, **412**: 116003
- [34] QIAO L, XU L Y, YANG Y Z, *et al*. Cratering records in the Chang'E -5 mare unit: filling the "age gap" of the lunar crater chronology and preparation for its recalibration[J]. *Geophysical Research Letters*, 2021, **48**(22): e2021GL095132

作者简介



刘方超 男, 1992 年 10 月出生于山东省潍坊市, 现为山东大学空间科学与技术学院博士研究生, 主要研究方向为行星科学与大数据。
E-mail: you_mu_sdu@163.com



郭弟均 (通信作者) 男, 1988 年 9 月出生于四川省巴中市, 现为中国科学院国家空间科学中心副研究员、硕士生导师, 主要研究方向为月球与行星地质、行星遥感。
E-mail: guodijun@nssc.ac.cn



吕英波 (通信作者) 男, 1980 年 3 月出生于山东省莱芜市, 现为山东大学空间科学与技术学院教授、博士生导师, 主要研究方向为空间科学数据分析、深度学习与数据挖掘、深空资源勘探与利用、空间太阳能利用、材料模拟与工程仿真等相关的科学研究与技术攻关工作。
E-mail: lyb@sdu.edu.cn