ISPRS Journal of Photogrammetry and Remote Sensing 243 (2027) 446–463

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Globe-CraterSAM: a training-free generalizable framework for mapping sub-kilometer martian craters

Yiling Lin [a,b], Xie Hu [a,b,*]

[a] College of Urban and Environmental Sciences, State Key laboratory of Vegetation Structure, Function and Construction (VegLab), and Key Laboratory for Earth Surface Processes of the Ministry of Education, Peking University, Beijing, China
[b] Beijing Key Laboratory of Spatio-temporal Perception and Urban Resilience, China

ARTICLE INFO

Keywords:
Sub-kilometer craters
Adaptation
Generalized mapping
Training-free

ABSTRACT

Sub-kilometer craters are essential for resolving fine-scale Martian surface chronology, yet their global-scale mapping remains limited by the poor generalizability of existing methods and lack of systematic global-scale validation. Here, we propose Globe-CraterSAM, a training-free and generalized framework for sub-kilometer Martian crater mapping. Central to our framework is an adaptation strategy that identifies craters from Segment Anything Model (SAM)-segmented landform units by exploiting common circular depression characteristics of craters. Before crater identification, we introduce a Buffer-Split Loop Strategy (BSLS) to separate overlapping landform units, preventing adjacent landform units from distorting the circular-depression morphology of craters. A circularity index is then applied to exclude non-circular landform units. We further reconstruct 6-m-resolution quasi-3D terrain from monocular optical images using a Shape-from-Shading-based approach and apply Double-Gaussian Fitting to define quantitative criteria for crater-like depressions. This design enables topographic information to be incorporated into sub-kilometer crater mapping despite the limited resolution of existing global Digital Elevation Models (DEMs). To address the lack of systematic validation for sub-kilometer crater mapping, we construct the first globally comprehensive benchmark dataset, comprising more than 10,000 craters from 50 sub-areas spanning 26 distinct landform types across Mars. Validation results demonstrate the strong generalizability of Globe-CraterSAM, with an average F1 score of 90.09%, outperforming nine state-of-the-art (SOTA) models by 12.38–65.88 percentage points. The framework also demonstrates strong feasibility for large-extent applications, achieving an F1 score of 91.58% across 12 $1^\circ \times 1^\circ$ longitude-latitude tiles. Together, Globe-CraterSAM and the benchmark dataset provide a scalable foundation for sub-kilometer crater mapping and open new opportunities for high-resolution planetary investigation.

1. Introduction

Cratering is a widespread process shaping celestial bodies and provides an important record of their surface formation and modification histories (Hartmann, 1965; Melosh, 1989; Neukum et al., 2001;Blanco-Rojas et al., 2024). Geological ages inferred from crater size-frequency distributions (SFDs) act as a foundation for numerous planetary studies, and inform the landing site selection and rover navigation (Liu et al., 2020; Ye et al., 2025). Accordingly, developing catalogs that document crater sizes and spatial distributions has long been a fundamental task in the planetary science community.

While current global catalogs have documented Martian craters larger than 1 km in diameter (e.g., Barlow, 1988; Stepinski et al., 2009; Salamuniccar et al., 2010; Robbins & Hynek, 2012), the proliferation of meter-resolution imagery now provides an opportunity to catalog sub-kilometer craters. These smaller craters can reveal surface processes operating at finer scales that existing catalogs cannot resolve: They are essential for dating young or spatially limited regions, where large impacts are rare and age estimates must rely on small craters (Robbins et al., 2011; Hartmann et al., 2010; Quantin et al., 2004). Their relatively frequent formation enables detection of newly formed craters within limited orbital observation periods, constraining present-day

* Corresponding author at: College of Urban and Environmental Sciences, State Key laboratory of Vegetation Structure, Function and Construction (VegLab), Key Laboratory for Earth Surface Processes of the Ministry of Education, and Beijing Key Laboratory of Spatio-temporal Perception and Urban Resilience, Peking University, Beijing, China.
E-mail address: hu.xie@pku.edu.cn (X. Hu).

https://doi.org/10.1016/j.isprsjprs.2026.10.005
Received 22 May 2026; Received in revised form 11 September 2026; Accepted 6 October 2026
0924-2716/© 2026 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

impact rates (Daubar et al., 2013). Meanwhile, their susceptibility to degradation makes them sensitive indicators of erosion, resurfacing, and environmental change on Mars (Michael & Neukum, 2010).

Despite this importance, existing sub-kilometer crater catalogs still fail to achieve both spatial coverage and mapping accuracy. High-accuracy catalogs are restricted to regional areas (e.g., Huang et al., 2023; Crosta et al., 2025), whereas global-scale datasets remain limited. To the best of our knowledge, the catalog compiled by Lagain et al. (2021) is currently the only global sub-kilometer crater dataset. However, its recall rate is ~ 75%, indicating that many craters remain undetected. The absence of a mature sub-kilometer Martian crater catalog increases the effort required for crater cataloging in related studies and constrains fine-scale exploration of the Martian surface.

Substantial efforts have been made to address the gap in sub-kilometer crater cataloging. In particular, numerous deep learning-based mapping methods have been developed and have demonstrated high accuracy in certain regions (e.g. Lin et al., 2024; Wang et al., 2022). However, applying these models at the global scale remains challenging. First, few of these models have been validated at large or global scales, leaving considerable uncertainty about their performance across diverse geological settings. Moreover, models that have been validated over larger spatial extents exhibit significant performance degradation. For instance, Fairweather et al. (2022) evaluated their model on six lunar areas with varying lighting and observed a marked drop in accuracy under poor illumination. Similarly, to assess the reliability of their global sub-kilometer crater catalog, Lagain et al. (2021) conducted an analysis using ~ 2,000 craters across six terrain types: volcanic plains, highlands, lowlands, aprons, impact zones, and high-latitude areas. Their results reported a notable performance decline in mid- and high-latitude areas.

The limited generalization stems from the heterogeneity of crater representation in observational data. Firstly, given the low resolution of global Digital Elevation Models (DEM) (~O(100 m)), crater detection models rely primarily on images, (Smith et al., 2001). However, image quality is strongly affected by multiple factors: the brightness of craters varies considerably with changes in illumination, sensor viewing geometry, and atmospheric effects (Cadogan, 2020; Wang et al., 2022). Secondly, crater morphology and distribution patterns are diverse on Mars. Craters often occur as overlapping and multi-ring structures (Wyrick et al., 2004), and are susceptible to wind-, water-, and ice-related processes (Day & Kocurek, 2016; Bamber et al., 2022; Baker & Carter, 2018). At the sub-kilometer scale, the complex spatial relationships between craters and non-crater landforms — occurring alongside, within, or as background to craters — complicate crater isolation. Moreover, landforms with similar circular or depressed morphologies (e.g., rootless cones, mud volcanoes and caves) are easily misclassified as craters. These factors can cause models trained on spatially limited datasets to fail when generalized across diverse geological settings, creating a critical barrier to global-scale crater mapping. These limitations underscore the need for scalable, accurate methods to map sub-kilometer craters globally.

The recently developed foundation model, Segment Anything Model (SAM), demonstrates unprecedented generalizability and strong perceptual capabilities for generic object boundaries by training on the largest real-world dataset to date (Kirillov et al., 2023). These capabilities enable SAM to delineate typical landforms, including diverse small craters (Osco et al., 2023), offering a new avenue to address the challenge in global-scale crater mapping. An essential step in applying SAM to small crater detection is adapting it from general landform delineation to crater-specific mapping. Currently, the most widely adopted adaptation methods are fine-tuning. However, generating crater datasets on a scale comparable to SAM's pretraining data is unrealistic. As a result, fine-tuning on small downstream datasets often causes "catastrophic forgetting", where the model overfits the new data and loses the original generalizability (Wang et al., 2026). Alternatively, some feature-based approaches avoid training by extracting craters from segmented landforms using circular or elliptical indices (e.g. Giannakis et al., 2024; Zhao & Ye, 2024). However, these methods often fail to separate craters from similar circular landforms such as mounds, and cannot effectively retain overlapping craters without typical circular shapes. Therefore, there is an urgent need for more effective adaptation strategies that can better leverage SAM's generalization ability to accurately detect diverse sub-kilometer craters.

Here, to address the gap in global sub-kilometer crater mapping, we propose a SAM-based framework for sub-kilometer crater mapping and construct a global-scale validation benchmark. The main contributions of this study are as follows: First, we propose an integrated, training-free framework for identifying sub-kilometer Martian craters from SAM segmentation results. The framework combines landform-unit separation, geometric screening, qusi-terrain generation based on Shape from shading, and terrain-based depression constraints within a unified workflow, exploiting the fundamental morphological characteristic that impact craters are generally circular depressions distinct from surrounding landforms. This integrated design enables robust crater identification across diverse crater morphologies and heterogeneous terrain conditions. Second, we construct the first globally comprehensive benchmark dataset for sub-kilometer crater mapping, comprising more than 10,000 craters from 50 sub-areas across 26 distinct landform types distributed throughout Mars. Beyond validating our framework, the benchmark dataset holds broad potential to support methodological development in sub-kilometer crater mapping.

2. Datasets

2.1. CTX imagery

To map craters as small as 100 m in diameter over large areas, we use Global Context Camera (CTX) Mosaic of Mars dataset in 500–800 nm visible band as input of our model. The dataset is generated from 86,571 registered CTX images with ~ 6 m resolution, and covers 99.5% of the Martian surface between 88°S and 88°N (Dickson et al., 2018). It includes detailed metadata (e.g., subsolar longitudes) of the original CTX images which are used to comprise the mosaic. The high transparency of imaging geometry allows us to conduct tile-by-tile SfS processing. To reduce computational cost, we resample the 5-m resolution mosaic dataset back to 6 m.

2.2. Global dataset

To validate the model, we establish the first global-scale validation benchmark dataset for sub-kilometer craters. We manually delineate > 10,000 craters from 50 sub-areas spanning Mars from the equator to the poles (Fig. 1). Each sub-area comprises ~ 2,000 × 2,000 pixels and covers an area of ~ 144 km^2. These sub-areas can be categorized into four groups: (1) Craters-only sub-areas, which contain diverse craters with varying erosion states, formation mechanisms, distribution patterns, illumination conditions exclusively. This group also includes craters in polar regions affected by projection-induced distortion, as well as craters located in the landing-site areas of the Tianwen-1 and InSight missions (Fig. 2a). (2) Typical landform sub-areas, where craters coexist with common Martian landforms such as mesas, landslides, and sand dunes (Fig. 2b); (3) crater-like landform sub-areas, featuring landforms with a circular appearance, such as rootless cones, mounds, and mud volcanoes that can be easily misidentified as craters (Fig. 2c); and (4) varying surface sub-areas, where craters are embedded within complex terrain backgrounds (e.g., chaos terrain and brain terrain) (Fig. 2d).

The dataset is compiled by two annotators: the first creates the initial labels, and the second reviews all annotations to identify and correct missed craters, false positives, and inaccurate boundaries. The annotation protocol is as follows: (1) a crater must have a circular shape and depression; (2) only craters with diameters between 100 and 1,000 m

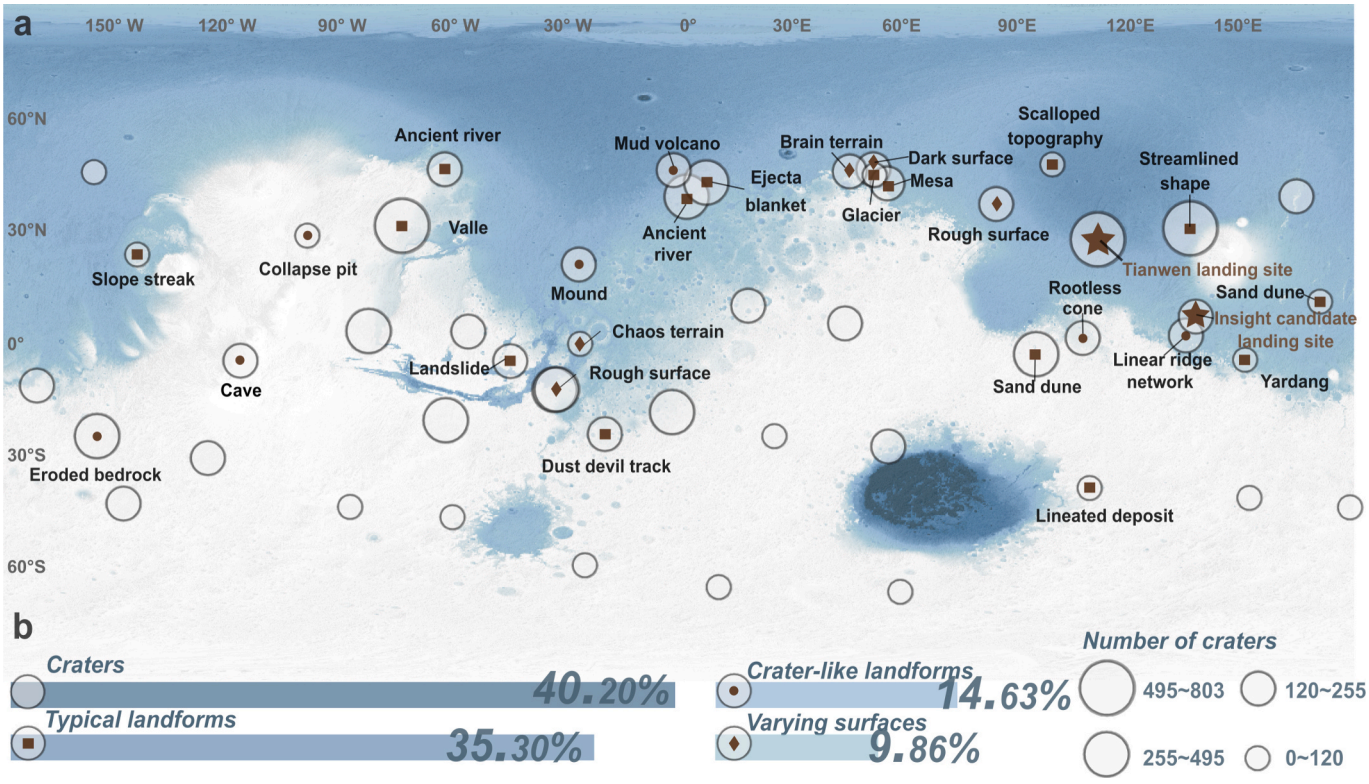


Fig. 1. Global testing dataset overview on Mars. (a) Locations of 50 sub-areas. (b) Proportion of crater counts within the four categories of sub-areas.

are included; (3) degraded craters are included if the rim or depression is still visible; (4) primary and secondary craters are both included; (5) uncertain craters are kept only when both annotators agree that the feature is a crater; and (6) the annotation follows the crater rim. If the rim is incomplete, the boundary is estimated from the visible rim segments and the overall circular or elliptical shape of the crater.

The dataset is divided into a validation dataset and a testing dataset. The validation dataset consists of 15 randomly selected sub-areas and is used for parameter sensitivity experiments of Globe-CraterSAM and for training the supervised SOTA comparison models. The remaining sub-areas constitute the testing dataset, which is used to evaluate model performance of Globe-CraterSAM and the SOTA methods.

3. Globe-CraterSAM

3.1. Methodological overview

We present Globe-CraterSAM, a SAM-based framework for addressing the limited accuracy and generalizability of global-scale sub-kilometer crater mapping. Central to the framework is a zero-shot adaptation approach that identifies craters directly from SAM's segmentation results on meter-scale imagery. The adaptation process exploits a fundamental characteristic: despite their diversity, craters are circular depressions, distinguishing them from other landforms. To exploit this characteristic, the framework integrates four components. It begins with a Buffer-Split Loop Strategy (BSLS) to disentangle overlapping landform units, followed by a circularity filter to filter out irregular ones. Quasi-3D terrain is then reconstructed from monocular optical images using a Shape-from-Shading (SfS) technique. Based on the reconstructed terrain, elevation profiles are analyzed using Double-Gaussian Fitting (DGF) to establish quantitative morphological criteria, excluding circular non-crater landforms (Fig. 3).

3.2. Segment Anything model (SAM)

By training a vast number of parameters on an unprecedentedly large real-world dataset, SAM achieves remarkable generalizability, demonstrated by its human-level understanding of object concepts. When fed CTX images, SAM can delineate the boundaries of landforms beyond craters, such as sand dunes, mounds, rootless cones, and dark streaks (Fig. 4). The mixture of craters with other landform units happens as SAM lacks task-specific information about craters and requires further adaptation for crater detection tasks.

When using SAM to map landform units, due to limited GPU memory, CTX images are cropped into sub-images of size 1024×1024 pixels. To avoid missing landform units close to the edges, a 30% overlap is applied between adjacent sub-images. If an overlapping pixel is labeled as a landform unit in any sub-image, it is considered a landform unit in the final prediction. The prediction results from all sub-images are then mosaicked into a single raster image. Finally, the raster image is converted into the vector format using GDAL FPolygonize with the default 4-connected pixel connectivity for subsequent adaptation procedures. The detailed configuration of SAM model is provided in Table S1.

3.3. Buffer-Split Loop strategy (BSLS) and circularity filter

Due to the binary nature of SAM's direct outputs, adjacent landform units in the vectorized results often exhibit undesired adhesion. This occurs not only between crater (Fig. 5a-b) but also between craters and non-crater landform units (Fig. 5c-b). The consequent irregular geometries hinder the exclusion of non-crater landforms and cause the erroneous removal of overlapping craters. Moreover, adhered craters introduce counting errors in SFD and contaminate the chronological results. The erosion-dilation approach in the field of mathematical morphology has been widely used to disconnect thin bridges between adjacent objects (Balado et al., 2020). However, a simple opening operation (erosion-dilation) is insufficient to separate densely clustered craters from other landform units.

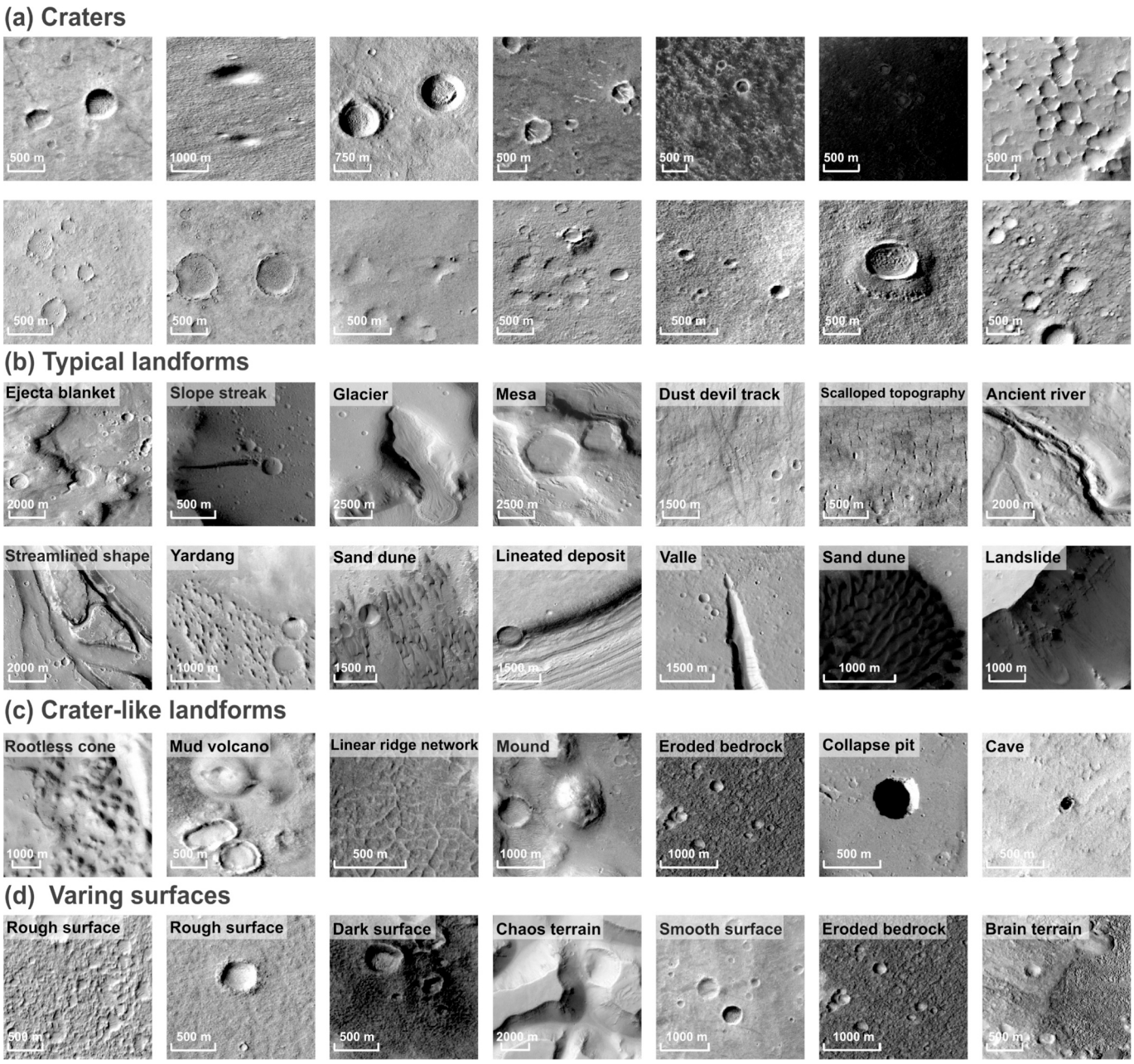


Fig. 2. Diverse craters and representative landforms included in the validation dataset.

Inspired by the opening operation in raster image processing, we develop the Buffer-Split Loop Strategy (BSLS) to separate adhered landform units in the vector domain. The basic idea is to progressively shrink a polygon until narrow connections between adjacent unit break, and then expand the separated parts to restore their shapes. For example, two touching crater-like objects may initially form a single polygon, but erosion can break their narrow connection and separate them into two individual units.

BSLS implements erosion and dilation using negative and positive buffer operations, respectively. All input landform units are first placed in a processing queue, which stores polygons that still need to be examined. For each polygon, the erosion magnitude is increased by a fixed step at each iteration. If erosion separates a single polygon into multiple distinct polygons, the resulting sub-polygons are restored using the corresponding positive buffer distance and returned to the processing queue as independent units for further examination. If the polygon is completely eroded before any split occurs, it is considered an indivisible landform unit. To retain its original shapes, we directly move the unprocessed vectors to the processed queue. If neither condition is met, the erosion magnitude is increased and the same polygon is examined again. This process continues until the processing queue is empty, at which point all indivisible landform units stored in the processed queue are exported as the final BSLS result (Fig. 6a). A representative example illustrating the iterative BSLS separation process for an adhered crater cluster is shown in Fig. 6b.

After BSLS, a circularity index, calculated using Equation (1), is applied to all individual landform units. The threshold of circularity is determined based on the sensitivity experiment (See Section 4.2 and 5.1). Notably, landform unit vectors located in the polar regions are subject to geometric distortion under the equirectangular projection employed by the CTX mosaic dataset, leading to reduced circularity values. To mitigate this risk, all vectors situated at latitudes greater than 60° are reprojected into a polar stereographic coordinate system (POLAR_STEREOGRAPHIC_MARS) prior to the circularity-based

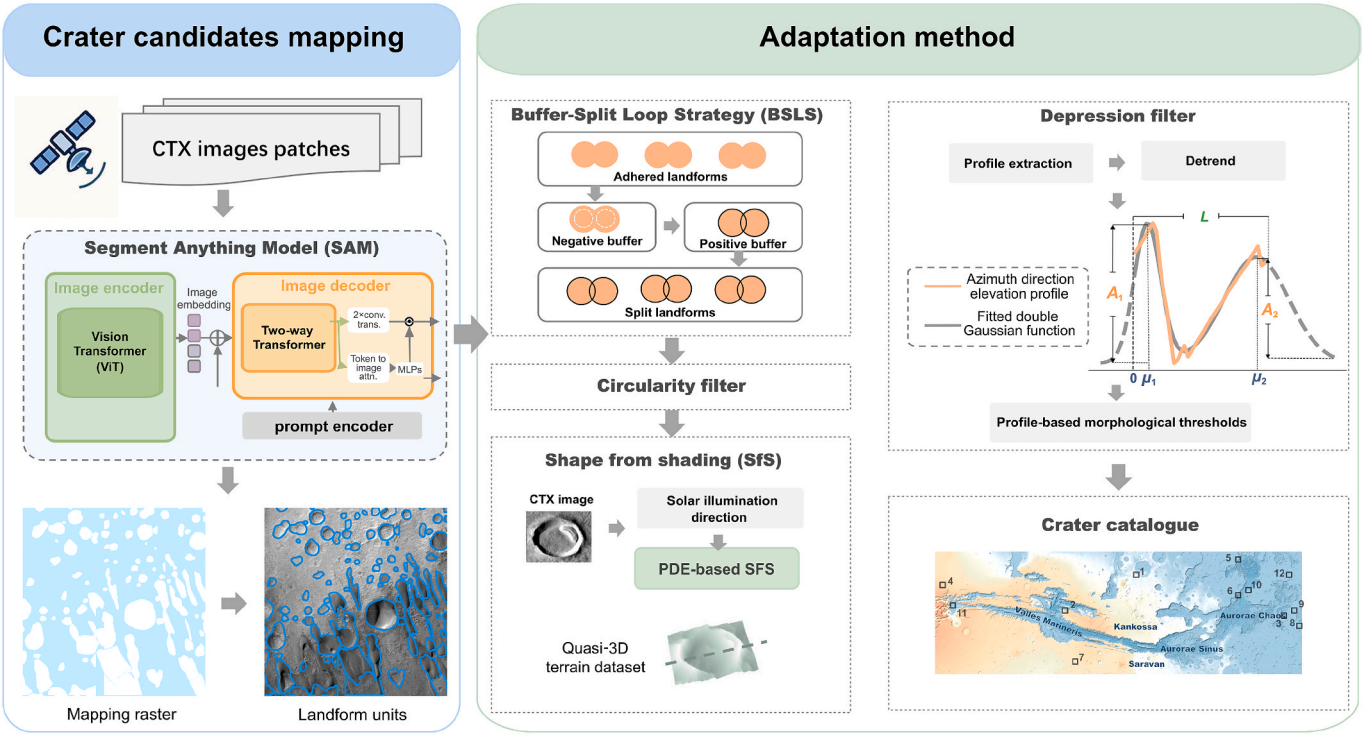


Fig. 3. The flowchart of Globe-CraterSAM, including CTX image input, SAM-based landform-unit mapping, adhered-feature separation using BSLS, circularity filtering, SfS-based quasi-3D terrain reconstruction, depression filtering, and final crater-catalogue generation.

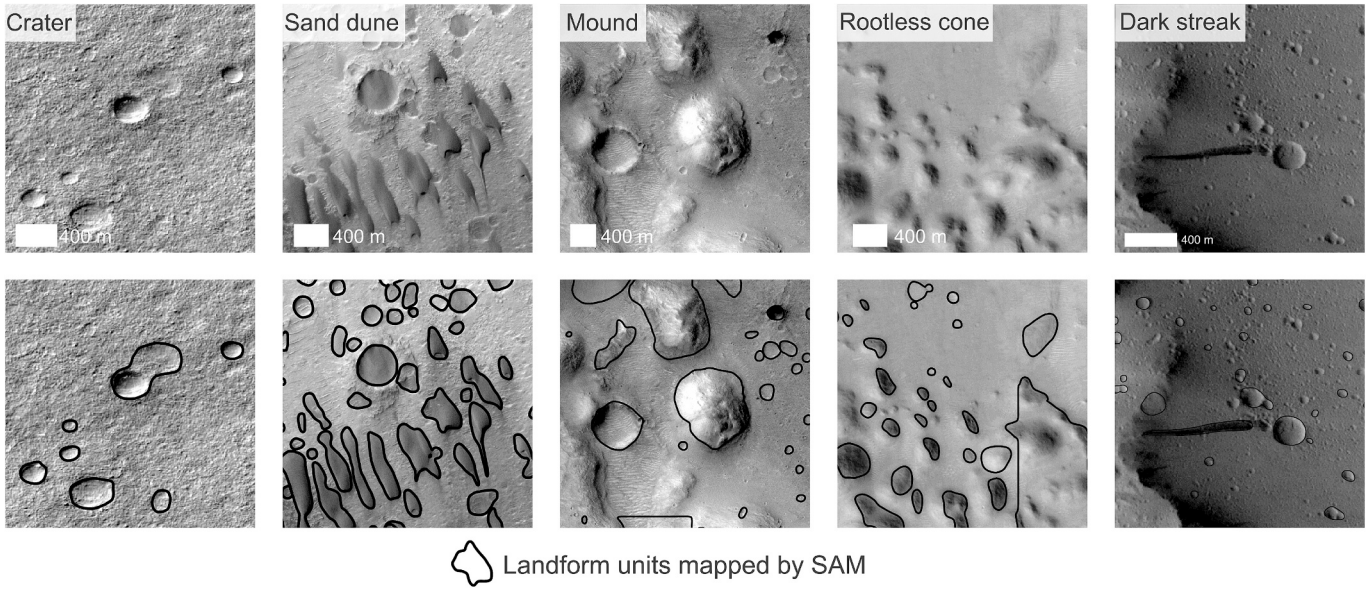


Fig. 4. Landform units mapped by SAM.

filtering. This projection is conformal and therefore preserves local angles and shapes relatively well, making circular and elliptical geometries more reliable in polar regions.

$$Circularity = \frac{4\pi \bullet Area}{Perimeter^2} \quad (1)$$

3.4. Shape from Shading (SfS)

Shape-from-shading, SfS, estimates three-dimensional (3D) terrain solely from the brightness variations in a single image. The method is effective in capturing subtle surface details and serves as a valuable complement to the DEM, particularly in scenarios where high-resolution DEM is unavailable (Wu et al., 2018). It has been widely used for landing site terrain reconstruction in previous planetary missions (Beyer, 2017; Liu & Wu, 2020).

The estimation relies on the relationship between photometric image intensity and surface geometry, which can be expressed as Equation (2):

$$I(x,y) = \rho(x,y) \times G(x,y) \quad (2)$$

where I presents intensity of a grayscale image, ρ denotes the surface reflectance properties (i.e., albedo), and G is the reflectance model

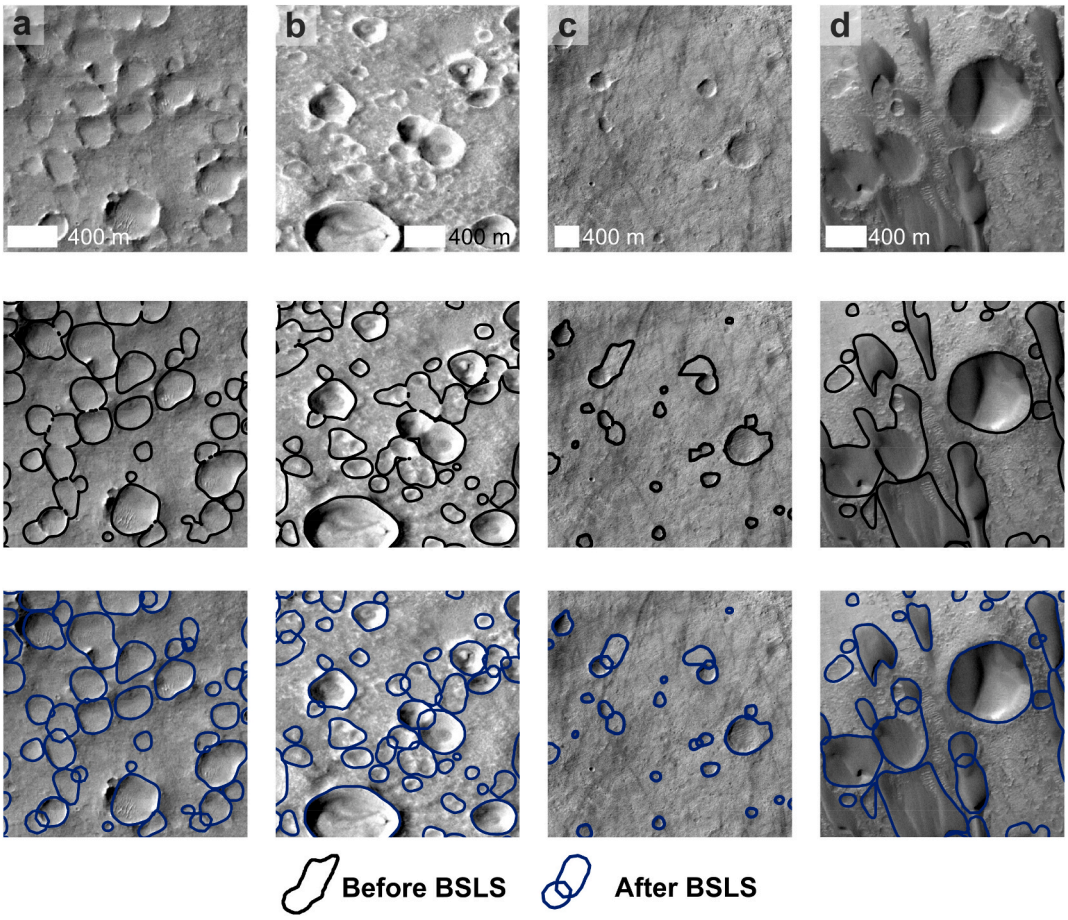


Fig. 5. Landform units before and after the BSLS operation. (a-b) Separation of overlapping craters and crater clusters. (c-d) Separation of craters and non-crater landform units.

describing the relationship between the image intensity I and the surface normal $n\,(x, y)$. The surface normal solved from Equation (2) can be used to infer the surface terrain. We adopt a well-known reflectance model that incorporates the complete illumination information through a Spherical Harmonics (SH) model (Basri & Jacobs, 2003). In the model, the SH coefficient l serves as the link between the surface normal $n\,(x, y)$ and the image intensity I:

$$I(x,y) = \rho(x,y) \times [\boldsymbol{l}\cdot\boldsymbol{n}(x,y)] \tag{3}$$

The SH model approximates complex illumination $L(\theta,\phi)$ with orthogonal basis functions $Y_l^m(\theta,\phi)$ defined on the unit sphere. And the SH coefficients l_{lm} represent the weights of corresponding basis function:

$$L(\theta,\phi) \approx \sum_{l=0}^{N}\sum_{m=-l}^{l} l_{lm}\cdot Y_l^m(\theta,\phi) \tag{4}$$

Each basis function captures a particular pattern of how light intensity varies with direction on the sphere. First-order basis functions ($l = 1$) represent linear directional variations and correspond to the three principal axes (x, y, z) in the Cartesian coordinate system, enabling the modeling of directional lighting. Higher-order basis functions can capture more complex lighting scenarios, such as ring-shaped sources, diffuse skylight, or multi-directional illumination. On Mars, the Sun is the sole source of illumination, and its light can be approximated as a directional light from a distant point source. First-order spherical harmonics are sufficient to represent such directional light, and Equation (3) can be rewritten as:

$$I(x,y) = \rho(x,y)[l_1, l_2, l_3]\begin{bmatrix} n_1(x,y) \\ n_2(x,y) \\ n_3(x,y) \end{bmatrix} \tag{5}$$

where $[n_1\,(x, y), n_2\,(x, y), n_3\,(x, y)]^{\mathrm{T}}$ denotes the field of unit-length outward surface normal in the Cartesian coordinate system along the x , y, and z axes, and $[l_1, l_2, l_3]$ are corresponding to the first-order SH coefficients.

With a single image, it is underdetermined to simultaneously solve for the three unknowns (i.e. albedo, surface normal, and SH coefficients). Therefore, additional constraints or prior assumptions are required. Here, we assume a perfectly Lambertian white surface and set the albedo uniformly to 1. A common approach to estimate SH coefficients is to combine image intensity with a coarse-resolution DEM (Gardner et al., 2019). However, atmospheric scattering significantly affects the image intensity distribution, leading to inaccuracies in the estimated coefficients (Liu & Wu, 2023). Therefore, we alternatively infer the coefficients based on the illumination geometry. Since the first-order SH basis functions correspond to the x , y, and z axes, the SH coefficients are proportional to the solar illumination direction projected onto these axes. This projected direction can be described in terms of the subsolar incidence angle θ and azimuth angle ϕ, which are available in metadata of CTX images:

$$\begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} \propto \begin{bmatrix} sin\theta cos\phi \\ sin\theta sin\phi \\ cos\theta \end{bmatrix} \tag{6}$$

Since l is required to be a normalized vector, and the expression derived from the incidence and azimuth angles is already unit-length, the two are proportional and therefore equivalent. As such, the l in Equation (5) can be directly replaced by the right side of Equation (6) and rewritten as:

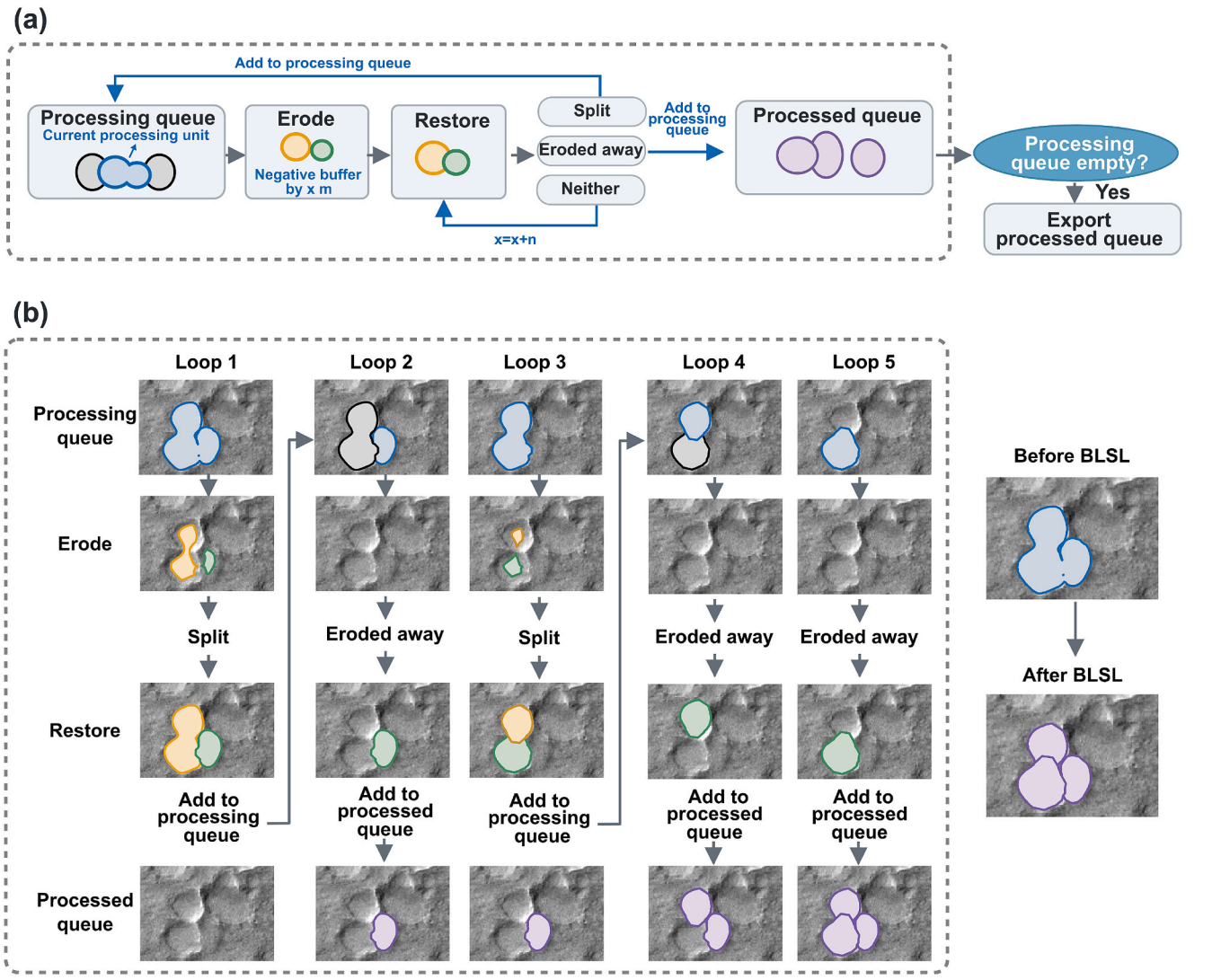


Fig. 6. Workflow and example application of BSLS. (a) Workflow of BSLS. (b) Example of the iterative BSLS separation process for an adhered crater cluster. The blue polygons indicate the polygons currently being processed in the loop.

$$I(x,y) = \rho(x,y)[sin\theta cos\phi, sin\theta sin\phi, cos\theta] \begin{bmatrix} n_1(x,y) \\ n_2(x,y) \\ n_3(x,y) \end{bmatrix} \tag{7}$$

Now the surface normal n is the only remaining unknown, and the equation becomes solvable. A widely adopted approach in previous studies is a two-stage shape recovery method, which first estimates the surface-normal n from Equation (7), and then integrates it to obtain the final depth map (Wu et al., 2018). However, this two-stage process is prone to bias propagation, as errors introduced during normal estimation can significantly affect the subsequent integration step. Therefore, we adopt the Partial Differential Equation (PDE)-based method proposed by Quéau et al. (2018), which simultaneously estimates the surface normal and the depth map through a PDE constraint.

Based on this method, we generate a dimensionless quasi-3D terrain dataset that represents relative surface relief rather than absolute elevation. The relief is reconstructed from local grayscale variations in the image, such as brightness differences between crater walls and crater floors, with larger local numerical differences indicating more pronounced surface relief. This relative relief information is suitable for crater identification, because the task primarily depends on detecting crater-specific local depressions rather than measuring precise absolute elevation. Therefore, the quasi-3D terrain dataset provides the local topographic signatures needed to distinguish craters from other landforms (Fig. 7).

3.5. Depression filter

Based on the quasi-3D dataset, we develop a depression filter to exclude circular non-crater landform units. For each landform unit, a terrain profile is initially extracted along the illumination azimuth direction, where the quasi-3D data demonstrates the highest sensitivity to topographic variation (See section 6.1). The profile length corresponds to the diagonal of the unit's bounding box. To eliminate the influence of large-extent background terrain, a detrending operation is applied to remove linear trend from the profile. Subsequently, the minimum value of the detrended profile is subtracted to normalize the baseline, facilitating consistent comparisons of relative relief across different landform units. After these pre-processing steps, a double-Gaussian Function (DGF) (Equation (8), which has been shown to be effective for extracting the morphometric parameters of craters, is applied to fit the terrain profiles (Wang et al., 2021).

$$f(x) = A_1 e^{\left(-\frac{(x-\mu_1)^2}{c_1}\right)} + A_2 e^{\left(-\frac{(x-\mu_2)^2}{c_2}\right)} \tag{8}$$

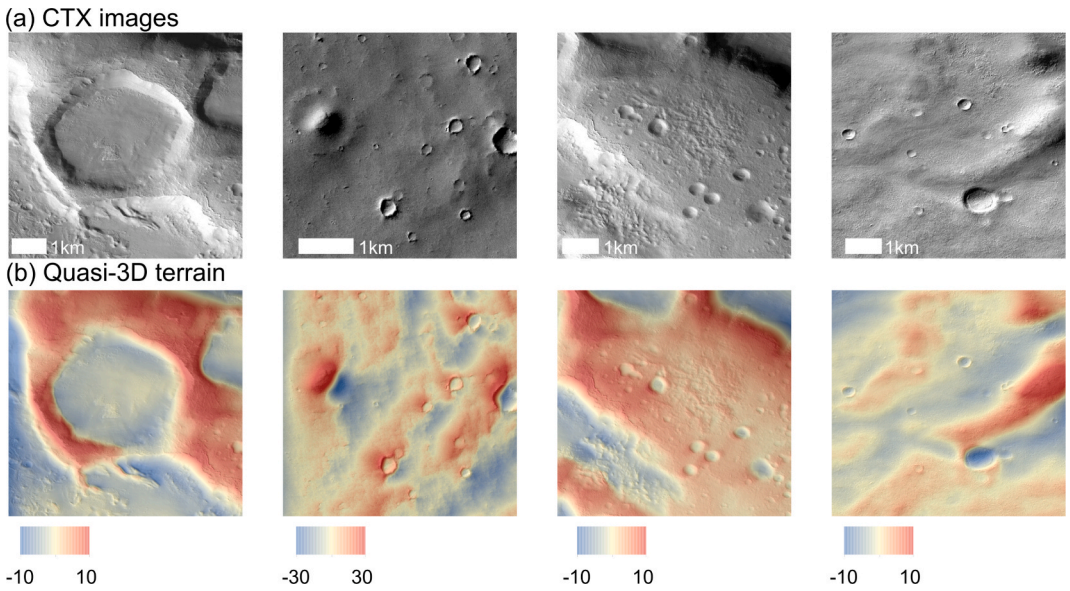


Fig. 7. The recovered quasi-3D terrain dataset. (a) CTX images used for terrain reconstruction. (b) Recovered quasi-3D terrain dataset.

The DGF yields a bowl-shaped curve with two peaks, where A_1 and A_2 denote the amplitudes of the peaks, μ_1 and μ_2 represent their central positions, and C_1 and C_2 control the width of each peak. Based on these parameters, we design two quantitative morphological criteria to identify crater-like concavities. The first criterion requires a sufficiently deep central depression. We define the midpoint between the two fitted peaks as $x_m = (\mu_1 + \mu_2)/2$, and quantify the depression depth as the vertical difference between the fitted value at this midpoint and the higher of the two fitted peaks: Where H represents the depth of the landform unit, and landform units with H smaller than $H_{threshold}$ are filtered out. This criterion helps remove circular but flat non-crater landform units that lack substantial topographic relief (e.g., Fig. 8 k).

$$\begin{cases} x_m = \dfrac{\mu_1 + \mu_2}{2} \\ H = \max[\mathrm{f}(\mu_1), \mathrm{f}(\mu_2)] - \mathrm{f}(x_m) \\ H > H_{\mathrm{threshold}} \end{cases} \tag{9}$$

In addition to flat landform units, many circular landform units exhibit topographic relief and cannot be filtered out using $H_{threshold}$ alone. However, their profile patterns do not match the bowl-shaped depressions characteristic of craters, which are typically defined by two peaks located on opposite sides of the profile. Representative examples include mud volcanoes, rootless cones, mounds, and linear ridge networks, whose peaks are often located near the center of the profile (e.g., Fig.8 g-j). Therefore, we introduce a second criterion that requires two peaks to occur on opposite sides of the profile, ensuring these non-crater landform units with topographic fluctuations are excluded. This criterion can be quantified as follows:

$$\begin{cases} R_1 = min(\mu_1, \mu_2) \\ R_2 = max(\mu_1, \mu_2) \\ R_1 < R_{threshold} \cdot L \\ R_2 > (1 - R_{threshold}) \cdot L \end{cases} \tag{10}$$

where L represent the length of the profile, R_1 and R_2 are positions of left and right peaks, respectively. $R_{\mathrm{threshold}}$ is a proportional threshold used to determine whether the two peaks are located near the two sides of the profile. The first condition requires the left peak to occur within the left $R_{\mathrm{threshold}}$ portion of the profile, while the second condition requires the right peak to occur within the right $R_{\mathrm{threshold}}$ portion. Together, these constraints ensure that the two peaks are positioned on opposite sides of the profile, consistent with the bowl-shaped depression characteristic of craters. Only landform units that meet both standards are identified as craters and retained in the final crater catalogue. Through sensitivity experiments, we set the $H_{threshold}$ as 0.5 and $R_{\mathrm{threshold}}$ as 0.33 (See Section 4.2 and 5.1).

4. Performance evaluation

4.1. Metrics

We evaluate Globe-CraterSAM using three groups of metrics. First, mapping accuracy is assessed using crater-level Precision, Recall, and F1 at IoU thresholds of 0.30, 0.40, 0.50, and 0.75, where IoU is defined as the intersection area divided by the union area between predicted and ground-truth craters. Precision measures the fraction of predicted craters that are correctly matched, Recall measures the fraction of ground-truth craters that are successfully detected, and F1 is their harmonic mean. Pixel-level IoU (pIoU) is calculated to quantify the overall spatial agreement. Moreover, IoU values of all ground truth craters are calculated and their overall distribution is used to assess the alignment between the model's predictions and ground truth. The IoU distribution offers a continuous assessment of the alignment between predicted and ground-truth boundaries, complementing the discrete, fixed-threshold nature of the F1 score. Second, to evaluate the reliability of the crater catalogue for downstream scientific applications, we calculate crater diameter error, center-location error, and rim-boundary error. Diameter error is the absolute difference between the fitted mean diameters of predicted and ground-truth craters, center-location error is the Euclidean distance between their fitted centers, and rim-boundary error measures the distance between the predicted and ground-truth crater rims. The detailed calculation procedures for these metrics are provided in the Supplementary Methods. We further report Precision, Recall, and F1 within different crater-diameter bins to characterize size-dependent detection performance. Third, computational efficiency is evaluated using the end-to-end processing time normalized by mapped area (s km^{-2}). The complete processing workflows from input imagery to the final crater catalogue are timed on the same hardware mentioned in 4.3 to provide a consistent comparison of computational efficiency.

4.2. Sensitivity and ablation experiments

We conduct sensitivity experiments using the validation dataset with 15 randomly selected sub-areas to determine the optimal thresholds and

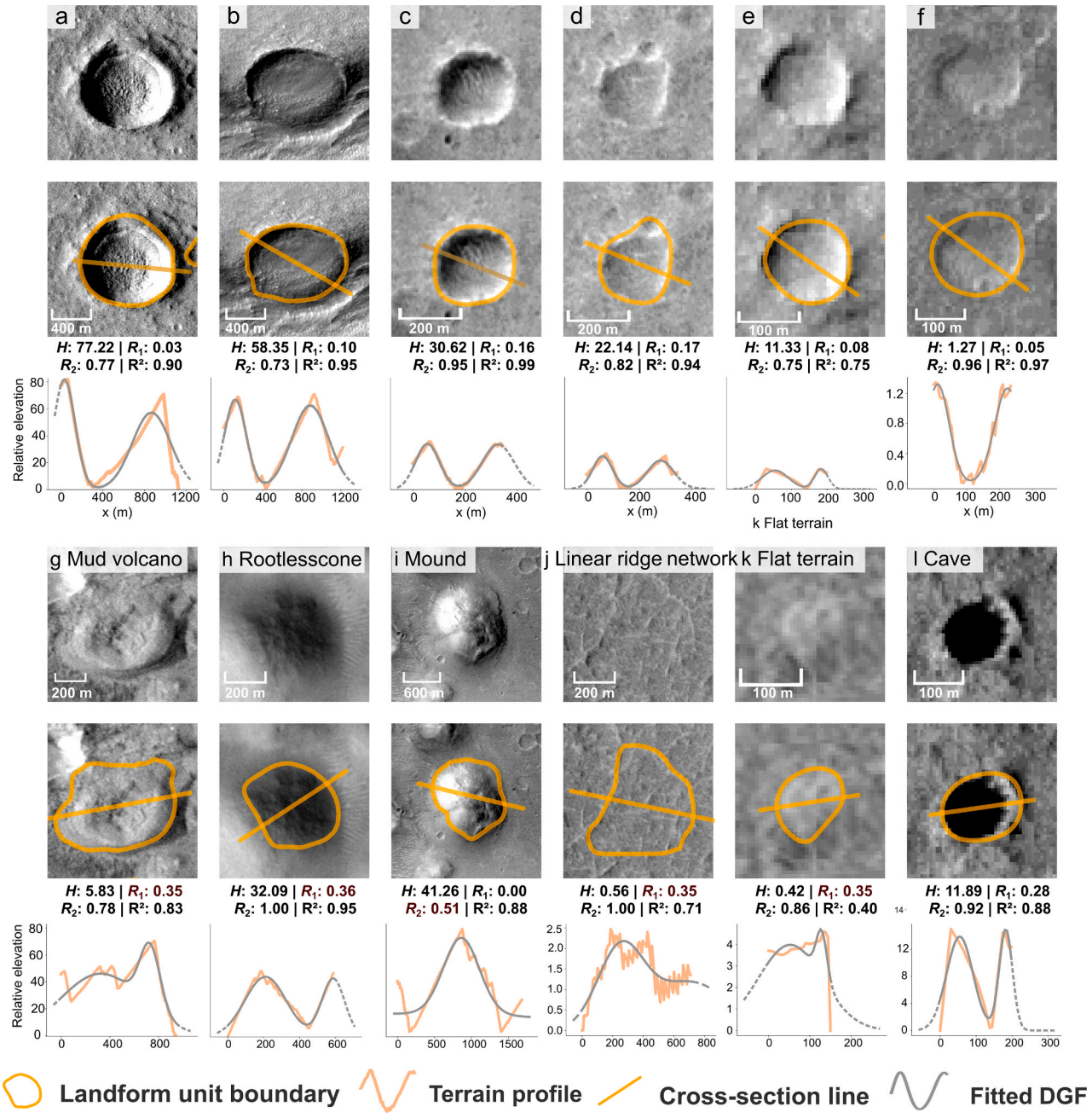


Fig. 8. Terrain profiles of craters and circular non-crater landforms. The bold dark-red values indicate parameters that do not meet the required criteria. (a-f) Deep and shallow craters with diameters of ~ 1000 m, ~250 m, and ~ 100 m. (g-l) Circular non-crater landforms.

assess the robustness of Globe-CraterSAM with respect to parameter variation. Four parameters are evaluated independently while keeping the others fixed at their adopted values. Circularity thresholds are tested from 0.75 to 0.95 in increments of 0.02, H thresholds from 0.1 to 0.9 in increments of 0.1, and R thresholds from 0.21 to 0.45 in increments of 0.03. The buffer step used in BSLS is additionally tested from 10 to 90 m in increments of 10 m.

Component-wise ablation experiments are also conducted to evaluate the contribution of each major module, including the BSLS, circularity filter, and depression filter. Each component is independently removed from the pipeline to quantify its individual effect on overall performance. In addition, the two thresholds used in the depression filter, h and R, are separately ablated to assess their respective contributions. To further examine whether the contribution of each module remains consistent across different terrain conditions, the same ablation experiments are performed separately for the four terrain classes represented in the dataset.

4.3. Comparison with SOTA models

We compare the performance of Globe-CraterSAM with six SOTA segmentation models, including DeepLabv3+ (Chen et al., 2018), DANet (Fu et al., 2019), DMNet (He et al., 2019), SegFormer (Xie et al., 2021), Mask2Former (Cheng et al., 2022), and Segmenter (Strudel et al., 2021). The first three models rely on convolutional neural networks, noted for their multiscale feature extraction capabilities, while the remaining three are Transformer-based and known for their large model capacity and superior generalization performance. For a fair comparison, all six models are fine-tuned using the crater ground truths from the same 15 randomly selected sub-areas used for parameter selection of Globe-

CraterSAM, while the remaining 35 non-overlapping sub-areas are used for evaluation. The input image resolution and the pre-processing and post-processing procedures are kept identical to those used for Globe-CraterSAM.

Because Globe-CraterSAM is training-free, we further compare it with three training-free crater detection methods. The first is a traditional unsupervised crater-detection method based on the Hough transform (Baronia et al., 2024). The other two, Zhao and Ye (2024) and Giannakis et al. (2024), are SAM-based methods that apply geometry-based post-processing for crater identification. For a fair comparison, their SAM implementation and parameter settings are kept identical to those of Globe-CraterSAM. The detailed configurations of these SOTA models are presented in Table S2.

All experiments are conducted on a workstation equipped with an Intel Xeon Platinum 8352 V CPU at 2.10 GHz, and four NVIDIA GeForce RTX 4090 GPUs. The SAM-based framework is implemented in Python using PyTorch 2.0.0 with CUDA 11.7. All the supervised comparison models are implemented using the OpenMMLab framework, including MMSegmentation 1.2.2.

4.4. Validation over large spatial extents

Since Globe-CraterSAM is designed for large-scale applications, we further select 12 additional $1^\circ \times 1^\circ$ latitude-longitude grids to evaluate its performance over larger, spatially continuous extents. These areas provide a validation setting that resembles practical regional-scale crater mapping scenarios. These grids are located within Valles Marineris (VM), the largest canyon system in the Solar System. VM is characterized by highly heterogeneous geomorphic settings, with kilometer-scale relief overprinted by aeolian, fluvial, and mass-wasting processes. This geomorphic diversity makes VM well suited for evaluating the performance of Globe-CraterSAM in large-scale crater mapping (Fig. 12a).

Within each grid, we manually refine all craters mapped by Globe-CraterSAM by removing non-crater landform units and supplementing missed craters. Removed non-crater units are counted as false positives (FPs), added missed craters are counted as false negatives (FNs), and retained correctly identified craters are counted as true positives (TPs). Precision, recall, and F1 score are then calculated for each grid based on the TP, FP, and FN counts.

5. Results

5.1. Sensitivity experiments

Globe-CraterSAM maintains relatively stable performance over broad ranges of the four parameters, indicating that the model's effectiveness stems from its adaptation design rather than parameter settings (Fig. 9). For the circularity threshold, the mean F1 remains above 90% between 0.75 and 0.85 and reaches its maximum of 92.28% at 0.83. As an ideal circular crater has a circularity of 1, the threshold of ~ 0.8 retains the moderately irregular craters while removing elongated or irregular non-crater features. Further increasing the threshold imposes an overly strict condition and is more easily to remove true craters. The model is relatively insensitive to the H threshold, with F1 ranging only from 91.19% to 92.28% across the entire tested range of 0.1–0.9. The highest F1 occurs at $H = 0.5$. For the R threshold, performance remains relatively stable between 0.30 and 0.45, with the maximum F1 of 92.28% obtained at$R = 0.33$. The BSLS buffer step shows limited sensitivity over 10–90 m, with F1 varying from 90.65% to 92.28%. The best performance occurs at a 10 m buffer step, which is close to the CTX pixel size (~6 m). Further reducing the step below the image pixel scale is unlikely to provide meaningful additional splitting detail, because such fine geometric variations cannot be reliably resolved from the source imagery. Conversely, if the step is too large, relatively small adhered units may be entirely eroded before a separable stage is reached, causing the splitting to fail. A 10 m step therefore provides a reasonable balance, being fine to capture the disappearance of narrow connections while avoiding unnecessarily small increments below the effective image resolution. Based on these results, the circularity threshold, H threshold, R threshold, and BSLS buffer step are set to 0.83, 0.5, 0.33, and 10 m, respectively.

5.2. Ablation experiments

The three adaptation components all contribute to the overall performance of Globe-CraterSAM (Table 1). BSLS effectively separates crater clusters, preventing their erroneous removal caused by irregular geometries or non-crater-like depression profiles. It also prevents craters attached to non-crater landform units from being incorrectly removed. Removing BSLS decreases Recall from 91.40% to 77.26% and F1 from 90.09% to 84.71% when IoU = 0.4, indicating that BSLS mainly contributes to improved completeness. The circularity filter primarily suppresses irregular non-crater landforms; removing it decreases Precision from 89.04% to 79.09% and F1 from 90.09% to 85.21%. The depression filter provides the strongest false-positive suppression: removing it reduces Precision to 73.67% and F1 to 82.06%, despite a slight increase in Recall to 93.80%. Further ablation of its two criteria shows that removing the h threshold decreases F1 to 88.50%, whereas removing the

Table 1
Quantitative evaluation of the ablation experiments. The metrics are averaged across 35 testing sub-areas.

Strategy	Precision (IoU = 0.4)	Recall (IoU = 0.4)	F1 (IoU = 0.4)	pIoU
Globe-CraterSAM	89.04	91.40	90.09	0.71
without BSLS	94.70	77.26	84.71	0.63
without circularity filter	79.09	93.12	85.21	0.60
without h threshold	85.45	92.08	88.50	0.70
without R threshold	80.97	92.38	85.99	0.65
without depression filter	73.67	93.80	82.06	0.62

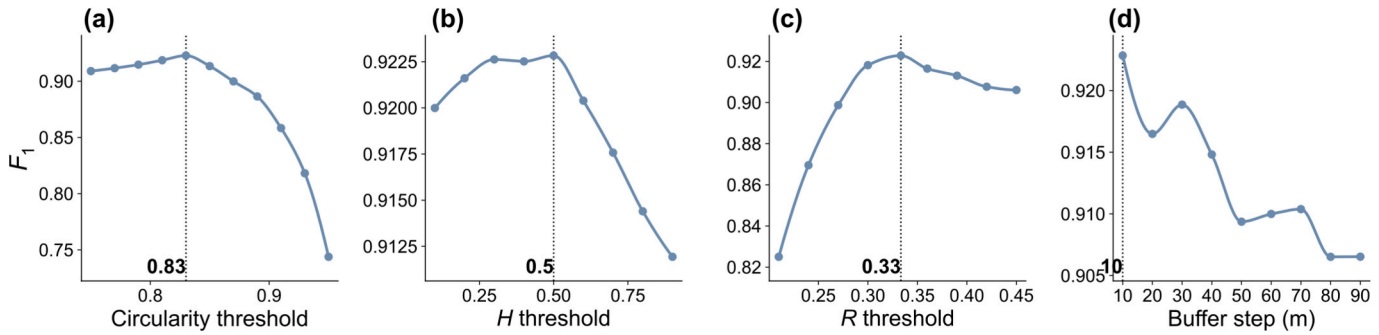


Fig. 9. Parameter sensitivity experiments.

R threshold decreases F1 to 85.99%, indicating a larger contribution of the *R* criterion to false-positive suppression. Consistent trends are also observed for pIoU, which decreases from 0.71 for the full model to 0.63, 0.60, and 0.62 when BSLS, the circularity filter, and the depression filter are removed, respectively. The ablation results show consistent performance degradation after removing each component across all four terrain types, demonstrating that the contributions of BSLS, the circularity filter, and the depression filter are robust to variations in terrain conditions (Table S3).

5.3. Comparison with SOTA models

Globe-CraterSAM demonstrates outstanding accuracy, significantly outperforming all SOTA models. Its average F1 across 35 testing sub-areas reaches 90.09% at IoU threshold of 0.4, while SOTA models only achieve average F1 scores ranging from 24.21 to 77.71%, corresponding to an improvement of 12.38–65.88 percentage points (Fig. 10a). Globe-CraterSAM also achieves a pIoU of 0.71, markedly higher than the 0.17–0.53 obtained by the SOTA models, indicating substantially better overall spatial agreement between the mapped crater regions and ground truth. Among the training-free comparison methods, the two SAM-based training-free methods, Zhao and Ye (2024) and Giannakis et al. (2024), achieve higher F1 scores of 74.26% and 77.71% and pIoU values of 0.44 and 0.53, respectively. Globe-CraterSAM maintains a similarly high Recall compared with the other SAM-based methods, while achieving a substantially higher Precision. This improvement in Precision is mainly attributed to the multiple filtering modules in Globe-CraterSAM, which effectively remove a large number of non-crater landform units (Table 2).

Moreover, Globe-CraterSAM achieves consistently high performance across all sub-areas, underscoring its strong generalizability with F1 scores at IoU threshold of 0.4 span from 78.89 to 100%, whereas those of the SOTA models vary widely across different sub-areas. This highlights the ability of Globe-CraterSAM in detecting craters formed through diverse geological processes and located within varied landforms and backgrounds. In contrast, the comparison models tend to perform well only under relatively simple transfer scenarios, but suffer substantial performance degradation in more complex settings (Fig. 10c and d). For instance, Globe-CraterSAM successfully recalls most craters situated among sand dunes and accurately filters out rootless cones and mud volcanoes that are morphologically similar to craters. In contrast, the competitor models miss the majority of these craters and frequently misidentify rootless cones and mud volcanoes as craters (Fig. S1).

Beyond crater-detection accuracy, the crater mapping results generated by Globe-CraterSAM demonstrate strong potential for downstream scientific applications. Globe-CraterSAM achieves a mean diameter error of 8.93 m, a center-location error of 8.94 m, and a rim-boundary error of 6.16 m, all lower than those of the comparison methods. Even relative to the best-performing competitor for each metric, Globe-CraterSAM reduces the diameter error by 3.24 m, the center-location error by 1.48 m, and the rim-boundary error by 2.37 m. In addition, Globe-CraterSAM maintains high performance across a wide range of crater sizes, whereas the competing methods show greater sensitivity to crater size, with their F1 scores decreasing markedly for

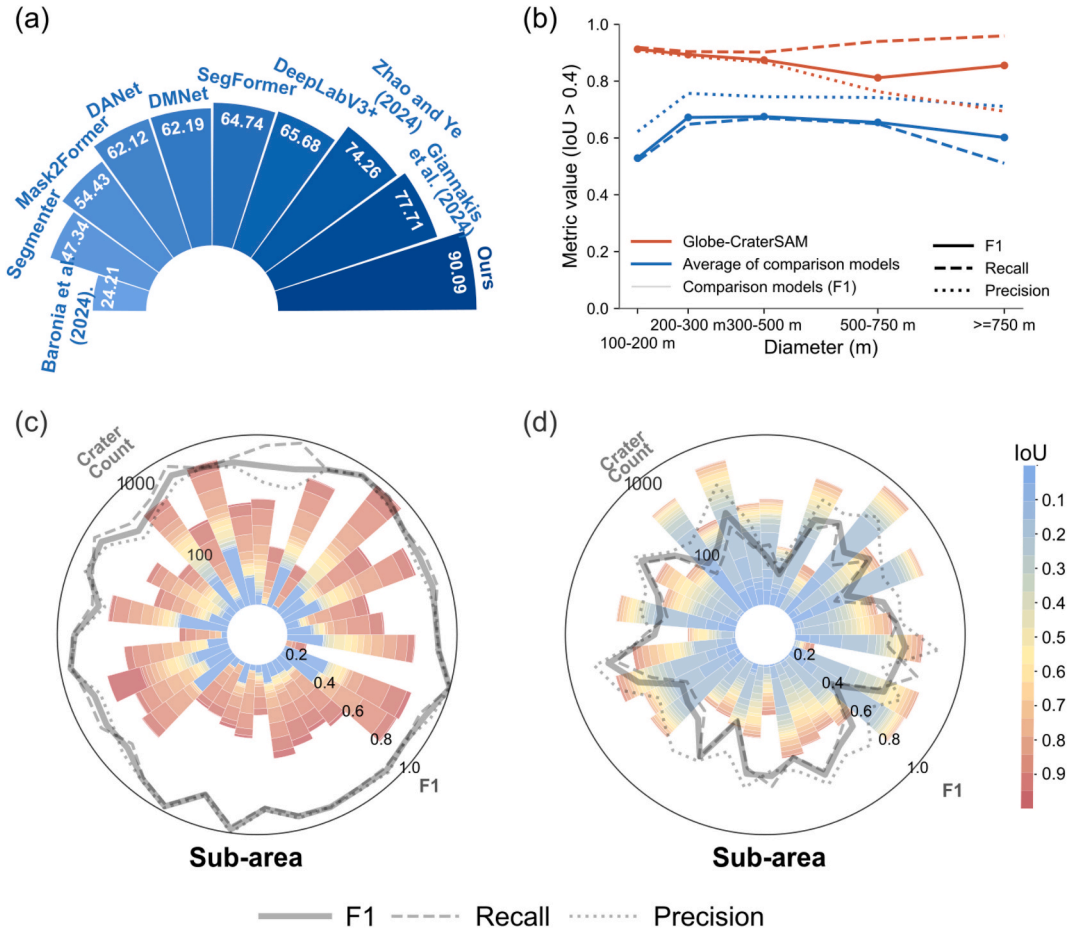


Fig. 10. Performance of Globe-CraterSAM and the comparison models for sub-kilometer crater detection. (a) F1 of Globe-CraterSAM and SOTA models. (b) The F1 scores at IoU threshold of 0.4 across various crater diameters for Globe-CraterSAM and SOTA models. (c) Performance of Globe-CraterSAM across sub-areas. (d) Performance of SOTA models across sub-areas. The bars in c-d illustrate crater boundary delineation accuracy through the IoU distribution, where colored segments in the sectors represent the number of craters within a specific IoU range. Higher IoU indicates more accurate boundary delineation.

Table 2
Quantitative comparison of Globe-CraterSAM with comparison methods.

Model type	Model	IoU threshold	Precision	Recall	F1	pIoU	Diameter error (m)	Center-location error (m)	Rim-boundary error (m)	End-to-end processing time (s km^{-2})
Ours	Globe-CraterSAM	0.30	90.41	92.11	91.14	0.71	8.93	8.94	6.16	~0.60
		0.40	89.04	91.40	90.09					
		0.50	87.22	89.76	88.36					
		0.75	78.99	79.19	79.00					
Supervised models	DeepLabV3+	0.30	75.26	63.01	67.22	0.50	20.30	15.44	9.12	~0.0037
		0.40	73.55	61.56	65.68					
		0.50	71.28	59.52	63.58					
		0.75	53.95	44.63	47.86					
	DANet	0.30	79.26	55.30	63.16	0.46	17.99	14.08	8.53	~0.0034
		0.40	78.01	54.40	62.12					
		0.50	75.46	52.53	60.01					
		0.75	58.18	40.10	46.00					
	DMNet	0.30	78.10	55.68	63.48	0.45	19.64	15.16	9.28	~0.0037
		0.40	76.66	54.50	62.19					
		0.50	73.08	52.23	59.60					
		0.75	53.82	38.31	43.80					
	Mask2Former	0.30	82.34	43.28	55.45	0.45	20.35	16.24	9.34	~0.0060
		0.40	81.00	42.43	54.43					
		0.50	78.60	40.94	52.63					
		0.75	62.97	32.49	41.95					
	SegFormer	0.30	75.53	60.70	66.09	0.52	22.13	15.82	9.64	~0.0060
		0.40	74.08	59.41	64.74					
		0.50	71.84	57.24	62.55					
		0.75	54.74	43.15	47.32					
	Segmenter	0.30	63.44	45.10	51.37	0.36	32.35	23.49	17.33	~0.0029
		0.40	58.33	41.66	47.34					
		0.50	50.39	36.14	40.97					
		0.75	15.88	12.17	13.34					
Unsupervised models	Zhao & Ye (2024)	0.30	66.72	87.66	75.05	0.44	12.17	10.42	11.13	~0.045
		0.40	65.99	86.79	74.26					
		0.50	64.91	84.99	72.89					
		0.75	46.57	58.39	51.22					
	Giannakis et al. (2024)	0.30	69.37	91.61	78.36	0.53	14.28	11.65	8.61	~0.042
		0.40	68.79	90.83	77.71					
		0.50	67.76	89.29	76.47					
		0.75	59.27	76.04	66.03					
	Baronia et al. (2024)	0.30	46.33	24.91	29.64	0.17	55.08	40.71	37.19	~0.032
		0.40	37.90	20.38	24.21					
		0.50	24.90	12.56	15.42					
		0.75	1.39	0.72	0.86					

craters smaller than 300 m in diameter (Fig. 10b). These results indicate that Globe-CraterSAM not only achieves high crater-detection accuracy but also provides more reliable estimates of crater size, location, and rim geometry, which are important for downstream applications such as crater size-frequency analysis and geomorphological characterization.

However, Globe-CraterSAM requires more computation time than the comparison models, with an end-to-end processing time of approximately 0.60 s km^{-2}, compared with 0.0029–0.0060 s km^{-2} for the supervised models and 0.032–0.045 s km^{-2} for the other training-free methods. For large-scale mapping, this corresponds to ~ 35, 30, and 17 min per 1° × 1° tile at 0-1°N, 30-31°N, and 60-61°N, respectively. The main computational bottleneck is the SfS processing, which accounts for ~ 0.56 s km^{-2}, ~94% of the total runtime (Table S4).

5.4. Performance over large spatial extents

Globe-CraterSAM achieves an average F1 score of 91.58% across these regions, confirming the generalizability of our method across diverse geological settings. It also demonstrates balanced performance, with a recall of 90.01% and a precision of 93.22%, indicating that our method improves crater detection completeness while minimizing the inclusion of non-crater features. These results further demonstrate the feasibility of applying the proposed approach across large, continuous spatial extents (Fig. 11b and c).

6. Discussion

6.1. Reliability of quasi-3D terrain data

We select three representative validation sites within the interior overlapping coverage of the SfS reconstructions and corresponding HiRISE DTMs, spanning contrasting illumination geometries, surface roughness, and albedo backgrounds (i.e., dark sand dunes and relatively bright bare land) (Fig. 12a). Since the quasi-3D data is dimensionless relative topography and lacks a common absolute elevation datum, their raw values cannot be directly compared with the absolute elevations of the HiRISE DTMs. We therefore evaluate the consistency of the relative terrain morphology. Specifically, for each candidate landform unit entering the depression filter, we compare normalized terrain profiles from our reconstructed quasi-3D data with those from the corresponding HiRISE DTMs along multiple directions to evaluate the overall fidelity of the reconstructed terrain morphology (See Supplementary method).

The quasi-3D profiles agree well with the HiRISE data, with mean normalized RMSE values of 0.24, 0.21, and 0.16 along the illumination-azimuth direction at Sites 1–3, respectively (Fig. 12b). The agreement varies little with profile orientation (Fig. 12d). The illumination-azimuth direction consistently yields nearly the lowest RMSE values and performs better than the corresponding multi-direction averages (Fig. 12c), supporting its use in the depression filter. Representative examples also show that the quasi-3D profiles align well with the topographic variations observed in the HiRISE DTMs (Fig. 12e-f). These results indicate that the quasi-3D data capture the relative terrain morphology required by the depression filter, rather than merely reflecting local image brightness or shadow contrast.

6.2. Effectiveness of BSLS

To comprehensively evaluate the practical effectiveness and potential risks of BSLS, as well as its benefits for downstream applications, we further quantify its effects on crater detection, over-splitting, and diameter estimation. The results show that the BSLS improves crater detection: Among the original SAM masks split by BSLS in the testing dataset, 66.8% recover at least one true crater missed by the corresponding workflow without BSLS. A representative crater-recovery scenario involves adhered craters, where irregular merged craters are split into more regular sub-regions that can pass subsequent filtering. Another 20.9% of the splits produce no change in final detection relative to the without-BSLS workflow: For example, an object that is filtered in the workflow without BSLS may be split by BSLS, but none of the resulting sub-regions pass the subsequent filtering, leaving the final detection unchanged. Only 12.4% of the splits increase false positives without recovering any crater relative to the without-BSLS workflow. For example, a non-crater feature may be split into multiple sub-regions, and if more than one of these sub-regions is retained in the final

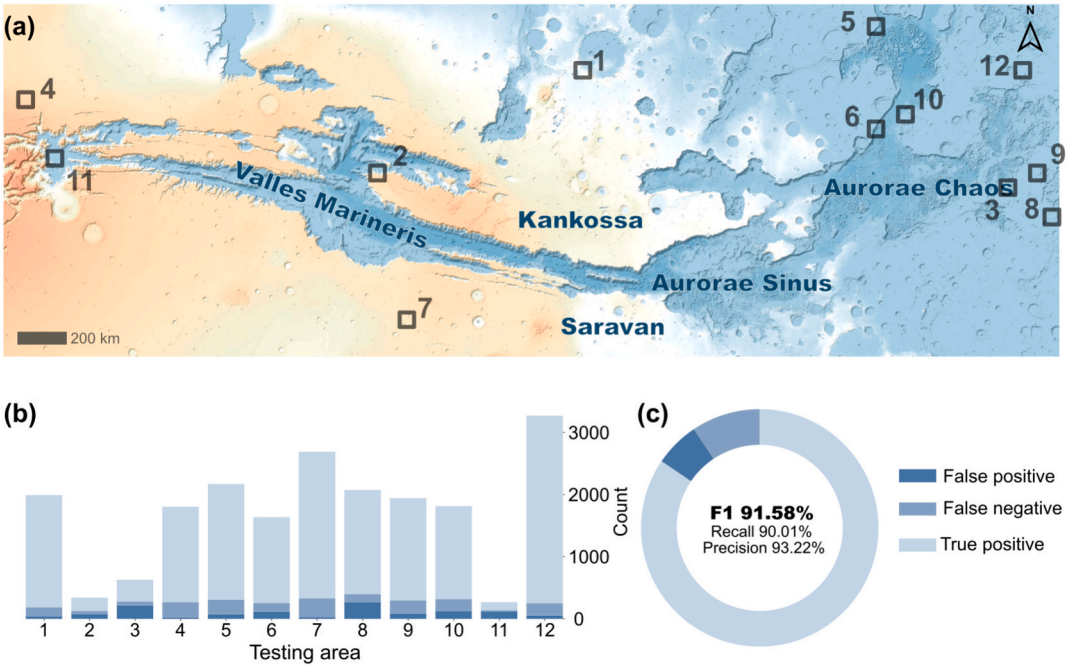


Fig. 11. Performance of Globe-CraterSAM over large spatial extents. (a) Geological locations of the 12 testing areas, each measuring 1° × 1° (longitude × latitude). (b) Distributions of false positives, false negatives, and true positives across the 12 testing areas. (c) Aggregated distribution of false positives, false negatives, and true positives, along with evaluation metrics across all testing areas.

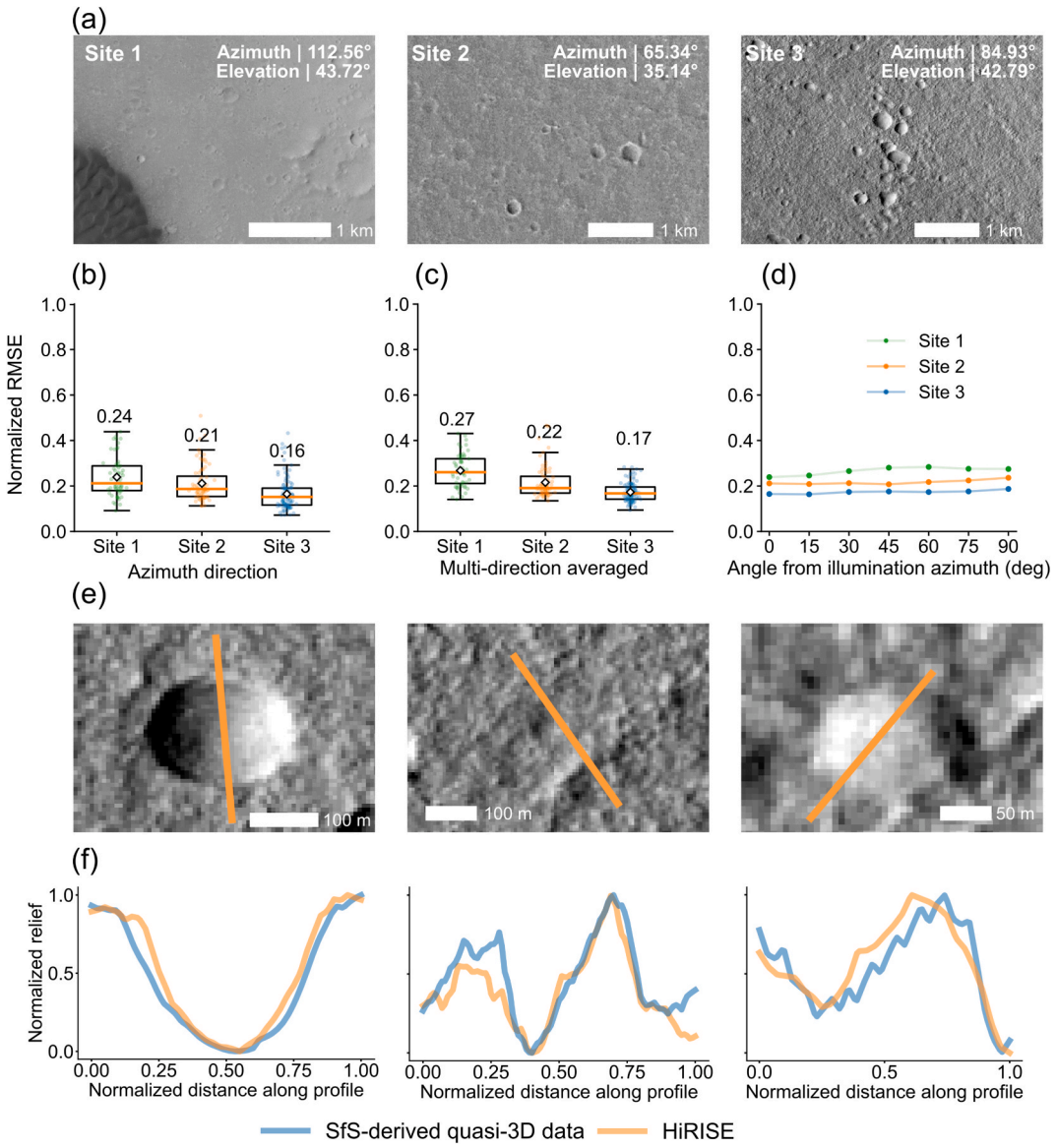


Fig. 12. Validation of quasi-3D terrain against HiRISE DTMs. (a) CTX images of the three validation sites, with illumination azimuth and elevation angles indicated for each site. (b-c) Distributions of normalized RMSE along the illumination azimuth (b) and average normalized RMSE across all profile directions (c). Values above the boxplots indicate the site-level mean RMSE. (d) Mean normalized RMSE of profiles at different directions relative to the illumination azimuth. (e) Representative candidate landform units, with orange lines indicating the directions along which terrain profiles are extracted. (f) Corresponding detrended and normalized terrain profiles from the SfS-derived quasi-3D data and HiRISE DTMs.

predictions, additional false positives are introduced.

The risk of over-splitting is low: only 0.05% of single-crater are over-split into multiple features. This risk is slightly higher for irregular craters but remains limited overall. Using circularity as a measure of crater regularity, the over-split rate is 0.19% for craters in the lowest circularity quartile (the most irregular) and 0% for those in the highest circularity quartile (the most regular). We further observe little evident over-splitting in sub-areas containing large numbers of degraded craters, possibly because SAM can map degraded crater boundaries as relatively complete and regular masks, reducing their susceptibility to fragmentation by BSLS.

BSLS also improves diameter estimation for adhered craters. For those adhered craters successfully separated by BSLS, the median absolute diameter error decreases from 300.3 m for the original SAM masks before BSLS to 11.0 m after splitting. In contrast, for single craters that are not split by BSLS, the original SAM boundary is restored, leaving their diameter estimates unchanged. Overall, BSLS frequently improves detection performance and substantially improves diameter estimation for adhered craters, while posing relatively low risks of over-splitting or altering the diameter estimates of unsplit single craters.

6.3. *Comparison among profile strategies*

Because craters are two-dimensional depressions, their terrain can be characterized using profiles extracted along different directions. We compare three depression-filtering strategies: (1) a single profile along the predefined azimuth direction, (2) the average of 12 directional profiles, and (3) voting among the same 12 profiles. The 12 profiles are extracted along evenly distributed directions using the predefined azimuth as the starting orientation. For consistency, all other settings, including the major filtering thresholds, SFS processing, and SAM segmentation, are held constant across strategies. The voting threshold is optimized on the validation dataset, with the best performance obtained when at least 7 of 12 profiles pass the depression filter.

The single-profile and multi-profile voting approaches show broadly comparable performance, with their F1-scores differing by only 1.52 percentage points. Their relative performance varies across sub-areas, with the single-profile strategy performing better in 68.6% of sub-areas and voting performing better in 31.4%. The broadly comparable performance of the single-profile and multi-profile voting strategies likely arises because many craters retain recognizable depression along several directions, allowing them to be identified by both approaches (Fig. 13). Differences between the two approaches are therefore likely driven by a subset of craters with specific characteristics, revealing a trade-off between the two strategies. For instance, multi-profile voting can recover craters whose depression is weak along the azimuth direction because of degradation or irregular morphology but remains detectable in other directions (e.g., Fig. S2). Conversely, strongly degraded craters may retain a recognizable depression in only a few directions and therefore have too few depression profiles to pass the multi-profile voting criterion. In such cases, the azimuth profile strategy can be advantageous because illumination-related brightness contrasts can produce stronger apparent relief along this direction in the quasi-3D data (e.g., Fig. S3). Because of this trade-off, each strategy has advantages under different conditions. The slightly better overall performance of the single-profile strategy on the testing dataset does not indicate that it is universally superior across all crater conditions, but rather suggests that the testing dataset contains a larger proportion of cases that favor this strategy.

Multi-profile averaging shows a relatively lower F1-score of 86.62%, compared with 90.09% for the single-profile strategy. Moreover, the single-profile strategy performs better in 82.9% of the 35 testing sub-areas, indicating that it is more effective than multi-profile averaging in most testing scenarios (Fig. 13). This advantage may arise because weak depression signals in most directions can dilute the few remaining clear signals for strongly degraded craters. For irregular craters, asymmetric profile shapes from different directions may also partially cancel one another, further weakening the averaged depression signature.

We also evaluate an additional R^2 criterion for double-Gaussian profile fitting. Validation results show that $R^2 > 0.4$ gives the best performance. The overall F1 difference between using and not using the R^2 filter is small (0.1 percentage points). Without R^2 filtering, performance is better in 42.9% of sub-areas, identical in 31.4%, and worse in 25.7%. This small difference arises because the double-Gaussian model fits most terrain profiles reasonably well, so the majority of candidate objects already satisfy the R^2 threshold and are unaffected by the additional filtering. Consequently, differences between the two strategies occur only in a small number of special cases, reflecting a trade-off: R^2 filtering can remove false-positive candidates with poorly fitted profiles, but it can also reject true craters with non-standard depression morphologies, such as craters with highly irregular profiles (Fig. S4a) or with central peaks (Fig. S4b).

6.4. Model's performance across diverse conditions

6.4.1. Robustness of global thresholds

Because a single global threshold configuration is applied throughout the study, evaluating its robustness across diverse conditions is essential for identifying its applicable scenarios and potential risks. We therefore stratify the testing areas by latitude, illumination geometry, crater diameter, and landform type and calculate the mean performance within each group. For each condition group, we compare the adopted global threshold values with the best-performing alternative thresholds tested in the sensitivity experiments. Overall, F1 (IoU = 0.4) remains around 90% across most geographic, illumination, and geomorphological conditions. The adopted global threshold configuration yields the highest mean F1 in most groups, while the gains achieved by alternative thresholds in the few non-optimal groups remain marginal, at only 0.003–0.66 percentage points. Notably, among the four landform categories, crater-dominated areas show the highest performance, while the crater-like landform, varying-surface, and typical-landform groups also maintain F1 values of approximately 90% (Fig. 14).

In addition to these directly quantifiable environmental stratifications, we examine representative sub-areas spanning different surface brightness levels and degrees of crater degradation. The global threshold configuration maintains F1 values of 88.93%-94.80% across these representative conditions (Fig. S5). Because circular non-crater landforms represent another major source of potential confusion with craters, we further establish independent ground-truth samples for these features and construct a confusion matrix to quantify their misclassification as craters. Among the 267 ground-truth circular non-crater features, only 14 are incorrectly identified as craters, corresponding to an overall false-positive rate of ~ 5.2%. These results indicate that the adopted model and global threshold configuration are robust across a broad range of geomorphological settings (Table 3).

6.4.2. Challenging scenarios

Although Globe-CraterSAM with the global threshold configuration performs well across the diverse conditions, several challenging scenarios remain. A notable case involves large craters, for which performance decreases with increasing diameter and is lowest for craters ≥ 500 m (81.8%). Re-optimizing the thresholds improves the group-mean F1 by only 0.66 percentage points, indicating that this performance decline is not primarily caused by inappropriate threshold values. Instead, it is more closely associated with complex terrain backgrounds. Large elongated landforms, such as cliffs and steep slopes, can generate predictions extending along terrain edges. After BSLS, these elongated predictions may be divided into multiple circular candidates. The pronounced local relief and depressions associated with steep terrain can then allow these candidates to pass the depression filter, resulting in false-positives.

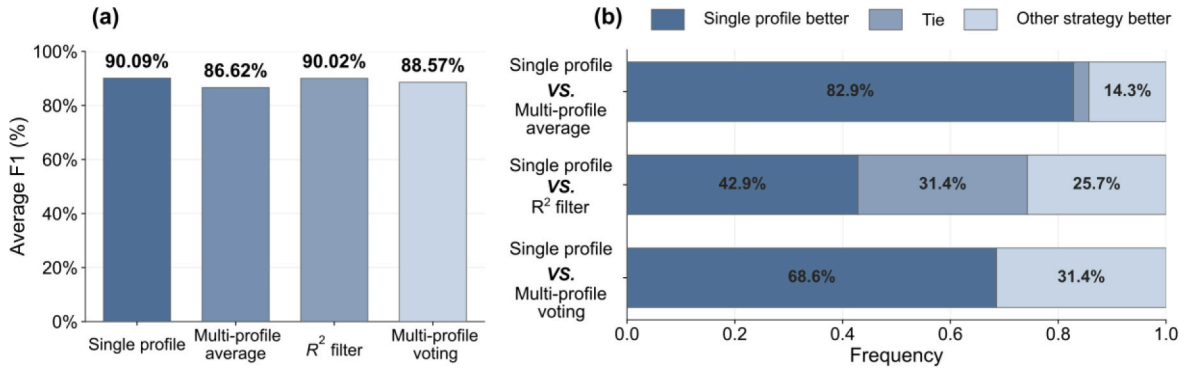


Fig. 13. Performance comparison of different profile strategies. (a) Average F1 of the different profile strategies across the testing sub-areas. (b) Sub-area-wise frequency of better, tied, and worse performance of the single-profile strategy relative to each alternative strategy.

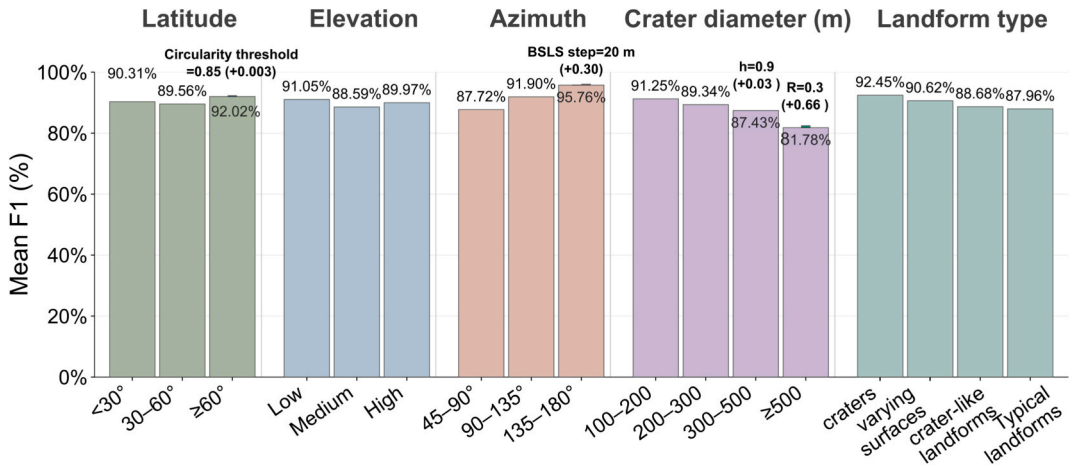


Fig. 14. Performance of Globe-CraterSAM across diverse geographic, illumination, crater-size, and geomorphological conditions. For groups where the adopted global thresholds are not optimal, the optimal parameter setting and corresponding F1 gain are indicated above the bars.

Table 3
Confusion matrix of circular non-crater landforms.

Circular landform	Predicted non-crater	Predicted crater	Total	Confusion rate
Rootless cones	141	4	145	2.8%
Mud volcanoes	88	9	97	9.3%
Mounds	22	0	22	0.0%
Caves	2	0	2	0.0%
Collapse pits	0	1	1	100%

Collapse pits represent another potential source of confusion. This failure is consistent with its crater-like morphology, as collapse pits can exhibit both approximately circular outlines and pronounced internal depressions. Severe crater adhesion presents a different limitation. In a sub-area containing numerous strongly adjoining craters, F1 decreases to 78.9%. BSLS may fail to fully separate these severely adhered craters. The resulting irregular candidate boundaries may subsequently fail the circularity constraint and be incorrectly removed (Table 3).

6.5. Limitation and future work

Despite the high generalizability of Globe-CraterSAM, certain limitations still persist. First, the performance of our framework may decrease under particularly challenging morphological and terrain conditions. Large craters embedded in complex terrain can generate false positives along cliffs or steep slopes, while collapse pits may be confused with craters because of their similar circular outlines and internal depressions. The impact of false positives associated with collapse pits may be limited because such features are rare relative to the global crater population. Compared to over tens of millions of craters on Mars, only slightly over one thousand pit chains and caves, respectively, have been identified across the planet (Wyrick et al., 2004; Cushing, 2015). Moreover, strongly adhered craters may also remain incompletely separated by BSLS and result in false negatives. In addition, craters that are severely degraded or exhibit irregular morphology along the azimuth direction may also be missed. Second, the computational efficiency of Globe-CraterSAM still requires improvement for large-area and planetary-scale applications. Since SfS reconstruction represents the major computational bottleneck, further optimization of the SfS workflow and parallel processing strategy could substantially improve processing efficiency and facilitate broader-scale applications. Third, the current SfS reconstruction assumes spatially uniform albedo and ideal Lambertian reflectance, whereas Martian surfaces can exhibit substantial albedo heterogeneity and non-Lambertian behavior, particularly over dark dunes and dust-covered regions. These effects may introduce uncertainties into the reconstructed terrain. Future work could incorporate multi-angle observations or multispectral constraints to improve reconstruction robustness.

7. Conclusions

With the proliferation of meter-resolution Martian imagery, compiling sub-kilometer crater catalogues has become increasingly important for revealing fine-scale surface processes that remain unresolved with existing kilometer-scale catalogues. However, global-scale sub-kilometer crater mapping remains challenging because existing methods often show limited spatial generalizability across diverse geological settings and lack comprehensive large-scale validation.

To address this limitation, we propose Globe-CraterSAM, a generalized framework for sub-kilometer crater mapping, and construct the first global validation benchmark for sub-kilometer craters. Results show that Globe-CraterSAM achieves strong generalizability across diverse craters and geological settings, with an average F1 score of 90.09%, outperforming 9 SOTA models by 12.38–65.88 percentage points and demonstrates strong feasibility for large, continuous crater mapping, achieving an average F1 score of 91.58%, recall of 90.01%, and precision of 93.22% across 12 randomly selected $1^\circ \times 1^\circ$ testing regions.

Overall, this study represents a significant step toward high-resolution global crater mapping by establishing Globe-CraterSAM as a scalable automated framework, introducing the first comprehensive benchmark for sub-kilometer crater evaluation, and extending Martian crater detection from the kilometer scale down to the 100 m scale.

Data Availability

We provided the global dataset and codes in https://github.com/YilingLin0610/Globe-CraterSAM-A-Training-Free-Generalizable-Framework-for-Mapping-Sub-Kilometer-Martian-Craters.

CRediT authorship contribution statement

Yiling Lin: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Investigation, Data curation, Conceptualization. **Xie Hu:** Writing – review & editing, Validation, Supervision, Resources, Investigation, Funding acquisition, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (42241132), and the Youth Open Grant of the NSSC The National Space Science Center of the Chinese Academy of Sciences (NSSDC2302002). We are grateful to W. Hamish Mitchell and Gaia Stucky de Quay for their valuable discussions and insightful comments on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2026.10.005.

References

Baker, D.M., Carter, L.M., 2018. Probing supraglacial debris on Mars 2: Crater morphology. Icarus 319, 264–280. https://doi.org/10.1016/j.icarus.2018.09.009.

Balado, J., Van Oosterom, P., Díaz-Vilariño, L., Meijers, M., 2020. Mathematical morphology directly applied to point cloud data. ISPRS J. Photogramm. Remote Sens. 168, 208–220. https://doi.org/10.1016/j.isprsjprs.2020.08.011.

Bamber, E.R., Goudge, T.A., Fassett, C.I., Osinski, G.R., De Quay, G.S., 2022. Paleolake inlet valley formation: factors controlling which craters breached on early mars. Geophys. Res. Lett. 49 (24). https://doi.org/10.1029/2022gl101097.

Barlow, N.G., 1988. Crater size-frequency distributions and a revised Martian relative chronology. Icarus 75 (2), 285–305. https://doi.org/10.1016/0019-1035(88)90006-1.

Baronia, A., Sarup, J., Gupta, S., Shanker, R., Chourasia, K., Soni, D., 2024. Multi-scale based approach for crater detection on lunar surface using clustering algorithm. Procedia Comput. Sci. 235, 1145–1156.

Basri, R., Jacobs, D., 2003. Lambertian reflectance and linear subspaces. IEEE Trans. Pattern Anal. Mach. Intell. 25 (2), 218–233. https://doi.org/10.1109/tpami.2003.1177153.

Beyer, R.A., 2017. Meter-scale slopes of Candidate InSight landing sites from point photoclinometry. Space Sci. Rev. 211 (1–4), 97–107. https://doi.org/10.1007/s11214-016-0287-7.

Blanco-Rojas, M., Carroll, M.L., Spradlin, C.S., Caraballo-Vega, J.A., Williams, Z.W., 2024. A novel approach to impact crater mapping and analysis on enceladus, using machine learning. J. Geophys. Res. Planets 129 (2). https://doi.org/10.1029/2023je008010.

Cadogan, P.H., 2020. Automated precision counting of very small craters at lunar landing sites. Icarus 348, 113822. https://doi.org/10.1016/j.icarus.2020.113822.

Chen, L., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation.* Paper presented at Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany. Doi: 10.1007/978-3-030-01234-2_49.

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). *Masked-attention mask transformer for universal image segmentation*. Paper presented at Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA. Doi: 10.1109/cvpr52688.2022.00135.

Crosta, G.B., De Blasio, F.V., Valbuzzi, E., Martinazzi, A., 2025. Landslide chronology on mars: a window into the past of Valles Marineris and martian slopes. Icarus 116653. https://doi.org/10.1016/j.icarus.2025.116653.

Cushing, G. E. (2015). Mars Global Cave Candidate Catalog PDS4 Archive Bundle. [Dataset]. PDS Cartography and Imaging Sciences Node (IMG). Doi: 10.17189/1519222.

Daubar, I., McEwen, A., Byrne, S., Kennedy, & Ivanov, B. (2013). The current martian cratering rate. *Icarus*, *225*(1), 506-516. Doi: 10.1016/j.icarus.2013.04.009.

Day, M., Kocurek, G., 2016. Observations of an aeolian landscape: from surface to orbit in Gale Crater. Icarus 280, 37–71. https://doi.org/10.1016/j.icarus.2015.09.042.

Dickson, J. L., Kerber, L. A., Fassett, C. I., & Ehlmann, B. L. (2018). A Global, blended CTX Mosaic of Mars with vectorized seam mapping: A new mosaicking pipeline using principles of non-destructive image editing. Paper presented at Lunar and Planetary Science Conference, The Woodlands, Texas.

Fairweather, J.H., Lagain, A., Servis, K., Benedix, G.K., Kumar, S.S., Bland, P.A., 2022. Automatic mapping of small lunar impact craters using LRO-NAC images. Earth Space Sci. 9 (7). https://doi.org/10.1029/2021ea002177.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., & Lu, H. (2019). *Dual attention network for scene segmentation.* Paper presented at Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA. Doi: 10.1109/cvpr.2019.00326.

Gardner, M. A., Hold-Geoffroy, Y., Sunkavalli, K., Gagne, C., & Lalonde, J.-F. (2019). *Deep Parametric indoor lighting estimation.* Paper presented at IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea. Doi: 10.1109/iccv.2019.00727.

Giannakis, I., Bhardwaj, A., Sam, L., Leontidis, G., 2024. A flexible deep learning crater detection scheme using Segment Anything Model (SAM). Icarus 408, 115797.

Hartmann, W.K., 1965. Terrestrial and lunar flux of large meteorites in the last two billion years. Icarus 4 (2), 157–165. https://doi.org/10.1016/0019-1035(65)90057-6.

Hartmann, W.K., Quantin, C., Werner, S.C., Popova, O., 2010. Do young martian ray craters have ages consistent with the crater count system? Icarus 208 (2), 621–635. https://doi.org/10.1016/j.icarus.2010.03.030.

He, J., Deng, Z., & Qiao, Y. (2019). *Dynamic Multi-Scale filters for semantic segmentation.* Paper presented at Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea. Doi: 10.1109/iccv.2019.00366.

Huang, H., Wang, X., Chen, Y., Zhang, Q., Zhao, F., Ren, X., et al., 2023. Observations and interpretations of geomorphologic features in the Tianwen-1 landing area on Mars by using orbital imagery data. Earth Planet. Phys. 7 (3), 331–346. https://doi.org/10.26464/epp2023005.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al. (2023). *Segment anything.* Paper presented at 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France. Doi: 10.1109/iccv51070.2023.00371.

Lagain, A., Benedix, G.K., Servis, K., Baratoux, D., Doucet, L.S., Rajšic, A., et al., 2021. The Tharsis mantle source of depleted shergottites revealed by 90 million impact craters. Nat. Commun. 12 (1). https://doi.org/10.1038/s41467-021-26648-3.

Lin, F., Hu, X., Lin, Y., Li, Y., Liu, Y., Li, D., 2024. Dual-branch multi-modal convergence network for crater detection using Chang'e image. Int. J. Appl. Earth Obs. Geoinf. 134 (104215). https://doi.org/10.1016/j.jag.2024.104215.

Liu, J., Zeng, X., Li, C., Ren, X., Yan, W., Tan, X., et al., 2020. Landing site selection and overview of China's lunar landing missions. Space Sci. Rev. 217 (1). https://doi.org/10.1007/s11214-020-00781-9.

Liu, W.C., Wu, B., 2020. An integrated photogrammetric and photoclinometric approach for illumination-invariant pixel-resolution 3D mapping of the lunar surface. ISPRS J. Photogramm. Remote Sens. 159, 153–168. https://doi.org/10.1016/j.isprsjprs.2019.11.017.

Liu, W.C., Wu, B., 2023. Atmosphere-aware photoclinometry for pixel-wise 3D topographic mapping of Mars. ISPRS J. Photogramm. Remote Sens. 204, 237–256. https://doi.org/10.1016/j.isprsjprs.2023.09.017.

Melosh, H. J. (1989). *Impact Cratering: a geologic process*. Cambridge: Cambridge University Press.

Michael, G., Neukum, G., 2010. Planetary surface dating from crater size-frequency distribution measurements: Partial resurfacing events and statistical age uncertainty. Earth Planet. Sci. Lett. 294 (3–4), 223–229. https://doi.org/10.1016/j.epsl.2009.12.041.

Neukum, G., Ivanov, B.A., Hartmann, W.K., 2001. Cratering records in the inner solar system in relation to the lunar reference system. Space Sci. Rev. 96 (1), 55–86. https://doi.org/10.1023/A:1011989004263.

Osco, L.P., Wu, Q., De Lemos, E.L., Gonçalves, W.N., Ramos, A.P.M., Li, J., Marcato, J., 2023. The Segment Anything Model (SAM) for remote sensing applications: from zero to one shot. Int. J. Appl. Earth Obs. Geoinf. 124, 103540. https://doi.org/10.1016/j.jag.2023.103540.

Quantin, C., Allemand, P., Mangold, N., Delacourt, C., 2004. Ages of Valles Marineris (Mars) landslides and implications for canyon history. Icarus 172 (2), 555–572. https://doi.org/10.1016/j.icarus.2004.06.013.

Quéau, Y., Mélou, J., Castan, F., Cremers, D., & Durou, J.-D. (2018). A variational approach to Shape-from-Shading under natural illumination. In *Lecture notes in computer science* (pp. 342-357). Cham: Springer International Publishing. Doi: 10.1007/978-3-319-78199-0_23.

Robbins, S.J., Di Achille, G., Hynek, B.M., 2011. The volcanic history of Mars: High-resolution crater-based studies of the calderas of 20 volcanoes. Icarus 211 (2), 1179–1203. https://doi.org/10.1016/j.icarus.2010.11.012.

Robbins, S.J., Hynek, B.M., 2012. A new global database of Mars impact craters ≥1 km: 1. database creation, properties, and parameters. J. Geophys. Res. Planets 117 (E5). https://doi.org/10.1029/2011je003966.

Salamunićcar, G., Lončarić, S., Pina, P., Bandeira, L., Saraiva, J., 2010. MA130301GT catalogue of Martian impact craters and advanced evaluation of crater detection algorithms using diverse topography and image datasets. Planet. Space Sci. 59 (1), 111–131. https://doi.org/10.1016/j.pss.2010.11.003.

Smith, D.E., Zuber, M.T., Frey, H.V., Garvin, J.B., Head, J.W., Muhleman, D.O., et al., 2001. Mars orbiter laser altimeter: experiment summary after the first year of global mapping of mars. J. Geophys. Res. Planets 106 (E10), 23689–23722. https://doi.org/10.1029/2000je001364.

Stepinski, T.F., Mendenhall, M.P., Bue, B.D., 2009. Machine cataloging of impact craters on Mars. Icarus 203 (1), 77–87. https://doi.org/10.1016/j.icarus.2009.04.026.

Strudel, R., Garcia, R., Laptev, I., & Schmid, C. (2021). *Segmenter: transformer for semantic segmentation.* Paper presented at Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada. Doi: 10.1109/iccv48922.2021.00717.

Wang, Y., Wu, B., Xue, H., Li, X., Ma, J., 2021. An improved global catalog of Lunar impact craters (≥1 km) with 3D morphometric information and updates on global crater analysis. J. Geophys. Res. Planets 126 (9). https://doi.org/10.1029/2020je006728.

Wang, Y., Xie, H., Huang, Q., Feng, Y., Yan, X., Tong, X., et al., 2022. A novel approach for multiscale lunar crater detection by the use of path-profile and isolation forest based on high-resolution planetary images. IEEE Trans. Geosci. Remote Sens. 60, 1–24. https://doi.org/10.1109/tgrs.2022.3153721.

Wang, Z., Ji, K., Wang, D., Zhang, H., & Cheng, F. (2026), SAMCL: Empowering SAM to continually learn from dynamic domains with extreme storage efficiency, *Proceedings of the AAAI Conference on Artificial Intelligence*, *40*(31), 26741-26750, doi:10.1609/aaai.v40i31.39884.

Wu, B., Liu, W.C., Grumpe, A., Wöhler, C., 2018. Construction of pixel-level resolution DEMs from monocular images by shape and albedo from shading constrained with

low-resolution DEM. ISPRS J. Photogramm. Remote Sens. 140, 3–19. https://doi.org/10.1016/j.isprsjprs.2017.03.007.

Wyrick, D., Ferrill, D.A., Morris, A.P., Colton, S.L., Sims, D.W., 2004. Distribution, morphology, and origins of Martian pit crater chains. J. Geophys. Res. Planets 109 (E6). https://doi.org/10.1029/2004je002240.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). *SegFormer: simple and efficient design for semantic segmentation with transformers.* Paper presented at Advances in Neural Information Processing Systems.

Ye, C., Kang, Z., Wan, X., Shao, S., Yang, J., Cao, Z., 2025. Map-Guided Coarse-to-Fine lunar rover localization via multiview crater matching. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 18, 19884–19896. https://doi.org/10.1109/jstars.2025.3594111.

Zhao, Y., Ye, H., 2024. Crater detection and population statistics in Tianwen-1 landing area based on Segment Anything Model (SAM). Remote Sens. (Basel) 16 (10), 1743. https://doi.org/10.3390/rs16101743.