

CNN-Based Large Area Pixel-Resolution Topography Retrieval From Single-View LROC NAC Images Constrained With SLDEM

Hao Chen , Xuanyu Hu , Philipp Gläser , Haifeng Xiao , Zhen Ye , Hanyue Zhang, Xiaohua Tong , *Senior Member, IEEE*, and Jürgen Oberst

Abstract—The production of high-resolution digital terrain models (DTMs) based on images is often hampered by the lack of appropriate stereo observations. Here, we propose a deep learning-based reconstruction of pixel-resolution DTMs from Lunar Reconnaissance Orbiter (LRO) single-view narrow angle camera (NAC) images, constrained by Selenological and Engineering Explorer and LRO LOLA Elevation Models (SLDEM). The procedure is carried out for a set of adjacent images, and the mosaicking of a contiguous large-area DTM is demonstrated. The approach is applied to the CE-3 and CE-4 landing sites, involving six multiple coverage and eight adjacent NAC L/R image pairs, respectively. For the DTM reconstruction, we use an improved convolutional neural network architecture with a weighted sum loss function involving three loss terms. We demonstrate that our method is robust and can deal with images acquired under varying illumination conditions. The DTM mosaic (1.5 m pixel size) covering the CE-4 landing area (72.8 × 30.3 km) is without apparent seams between the individual image boundaries and consistent with the SLDEM (60 m pixel size) in terms of overall elevation, trend, and scale, but is showing considerably more morphologic detail.

Index Terms—Deep learning-based reconstruction, digital terrain model (DTM) mosaics, pixel-resolution DTM, single-view narrow angle camera (NAC) images, Selenological and Engineering Explorer and LRO LOLA Elevation Model (SLDEM).

I. INTRODUCTION

DIGITAL terrain models (DTMs) play a ubiquitous role in planetary exploration, e.g., in support of mission planning [1], landing site selection [2], and autonomous landing and hazard avoidance [3]. Terrain models are also prerequisite for the investigation of planetary (sub-)surface environment, such as illumination conditions [4], material properties [5], and volatile distribution [6], and thus lay the groundwork for achieving key scientific objectives. Most recent and future manned or unmanned missions to the Moon will focus on lander, rover, and sample return exploration [2], [7], [8]. To ensure a safe and successful landing, DTMs are used to investigate terrain slope, illumination conditions, Earth communication, and obstacles (i.e., craters and rocks) distribution for the large candidate landing areas to select suitable landing sites [9]. Currently, the landing site selection work is mainly based on coarse-resolution DTMs derived from laser altimetry measurements and stereo photogrammetric models, e.g., 20 m resolution (CE-3 landing area) and 30 m resolution (CE-4 landing area) DTMs derived from the integration processing of Lunar Orbiter Laser Altimeter (LOLA) measurements and CE-2 stereo imagery collected during the circular orbit [10], [11]. Nevertheless, high resolution, i.e., meter-scale or lander/rover footprint scale, DTMs displaying more refined landing area environment provides a more detailed and faithful context for operations and investigations. On the other hand, 1.5-m resolution digital orthophoto maps (DOMs) were produced for the landing areas of CE-4 and CE-5 missions to extract craters and rocks that pose a safety hazard for the lander [11], [12]. Due to the lack of the same resolution topographic modeling to provide 3-D spatial information, some works only used the distribution of obstacles to guide landing site selection [11]. Actually, the volume of craters and rocks, especially ones on a meter scale or larger, which needs to be calculated from high-resolution DTMs, is also an important criterion for the evaluation of landing safety [10], [13], [14]. Overall, it is highly desirable to develop effective new techniques for large-area high-resolution lunar surface DTM generation,

Manuscript received 29 June 2022; revised 14 September 2022; accepted 1 October 2022. Date of publication 17 October 2022; date of current version 9 November 2022. The work of Hao Chen was supported by the China Scholarship Council (CSC) for his Ph.D. study at Technische Universität Berlin, Berlin, Germany. This work was supported in part by the National Natural Science Foundation of China under Grant 42101447, in part by the Shanghai Sailing Program under Grant 21YF1448800, and in part by the Open Access Publication Fund of TU Berlin. (*Corresponding authors: Xuanyu Hu; Zhen Ye.*)

Hao Chen is with the Institute of Geodesy and Geoinformation Science, Technische Universität Berlin, 10553 Berlin, Germany, also with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China, and also with the Shanghai Key Laboratory of Space Mapping and Remote Sensing for Planetary Exploration, Shanghai 200092, China (e-mail: hao.chen.2@campus.tu-berlin.de).

Xuanyu Hu, Philipp Gläser, and Haifeng Xiao are with the Institute of Geodesy and Geoinformation Science, Technische Universität Berlin, 10553 Berlin, Germany (e-mail: x.hu@tu-berlin.de; philipp.glaeser@tu-berlin.de; haifeng.xiao@campus.tu-berlin.de).

Zhen Ye is with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China, and also with the Shanghai Key Laboratory of Space Mapping and Remote Sensing for Planetary Exploration, Shanghai 200092, China (e-mail: 89_yezhen@tongji.edu.cn).

Hanyue Zhang is with the Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing 100083, China (e-mail: zhanghanyue@bjfu.edu.cn).

Xiaohua Tong is with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: xhtong@tongji.edu.cn).

Jürgen Oberst is with the Institute of Geodesy and Geoinformation Science, Technische Universität Berlin, 10553 Berlin, Germany, and also with the Institute of Planetary Research, German Aerospace Center (DLR), 12489 Berlin, Germany (e-mail: juergen.oberst@tu-berlin.de).

Digital Object Identifier 10.1109/JSTARS.2022.3214926

which generally requires several adjacent high-resolution orbiter images to cover.

The two narrow angle cameras (NACs), designated as NAC-left (NAC-L) and NAC-right (NAC-R), onboard the Lunar Reconnaissance Orbiter (LRO), operated in push-broom mode, have collected a huge number of images with an unprecedented pixel resolution up to 0.5 m, which are the primary data source for high-resolution lunar surface 3-D reconstruction [15]. At the nominal orbit height of 50 km, the two NACs cover a surface area of 5–25 km [16]. Large areas, such as the landing areas of China’s Lunar Exploration Program, usually require multiple adjacent NAC images for complete coverage. Stereo-model-based photogrammetry is the most widely used technique to retrieve high-resolution lunar surface DTMs from stereo or multiview NAC observations [15], [17]. While NACs have collected enough measurements to almost completely cover the lunar surface at least once, stereo pairs suitable for retrieving terrain models (which require a tilting of the spacecraft in consecutive orbits) are still spatially limited [17], [18]. Therefore, the construction of large-area DTMs from multiple stereo NAC images has been a long-standing challenge and issue. For example, even though the single-view NAC images cover the whole CE-4 landing area with approximately 72.8×30.3 km ($177.6^\circ\text{E} \pm 1.2^\circ$, $45.5^\circ\text{S} \pm 0.5^\circ$), there is only one stereo pair of NAC images covering a small portion of this area that could be used to generate a photogrammetric DTM [11]. Most existing photogrammetric DTMs are also only derived from a single stereo pair [15]. On the other hand, shape from shading (SFS) methods, constrained by coarse-resolution DTMs, have been successfully applied for the fine-scale topographic retrieval of local lunar surface from single NAC images [19], [20]; for example, the landing site of CE-4 [11]. Nevertheless, SFS is time-consuming in the case of large-scale mapping [21], and it is often to be less accurate in the cross-sun direction [22]. In addition, the results of the aforementioned techniques are often sensitive to image quality and model parameterization in practice. Consequently, the quality control of the DTM demands substantial empirical expertise and experience on an ad hoc basis [21], [23]. This, to a certain extent, also limits the applicability of the photogrammetric and SFS techniques to large-area DTM generation.

Nowadays, convolutional neural network (CNN)-based deep learning methods have been applied to DTM generation from single-view orbiter imagery for the lunar and Martian surfaces [23], [24], [25]. Chen et al. [24] applied CNN-based deep learning method to Martian surface 3-D reconstruction from single orbiter images. Their proposed network consists of two cascaded subnetworks. The first subnetwork is designed to convert input images to Lambertian images, and the second to predict DTMs from the Lambertian images. Tao et al. [23] developed a single-image-input Martian surface 3-D reconstruction system and produced three mapping products over the large activity area of the planned ExoMars 2022 Rosalind Franklin rover mission. Chen et al. [25] used the altimetric DTMs as part of the model input to constrain model learning. They proposed a dual-branch CNN architecture to predict pixel-resolution DTMs for the lunar surface, with one branch devised specifically to process the coarse-resolution DTM alongside single high-resolution images

as input. This way, the elevation constraint has been intrinsically assimilated in the learning process, which was demonstrated to significantly enhance the accuracy of model prediction. Furthermore, the deep learning methods can predict DTMs in a fast and user-friendly way without a priori and expert knowledge [23], and therefore may have great potential for large-area high-resolution lunar topographic modeling.

In this article, we propose a deep learning-based reconstruction for large-area pixel-resolution lunar surface DTMs from single-view NAC L/R image pairs and the Selenological and Engineering Explorer and LRO LOLA Elevation Model (SL-DEM). This is in contrast with photogrammetric technique, which requires stereo or multiview overlapping images as input to derive DTMs. Specifically, we introduce an improved CNN architecture with a weighted sum loss function accounting for three types of errors, which yield more accurate DTMs than in our earlier efforts [25]. The input data are first coregistered and splitted into small tiles in normalized scale. Once trained, our method reconstructs DTMs from each tiles, and then, scale recovery and mosaicking of DTM tiles inner and inter adjacent image pairs are carried out.

Here, we present two model applications for the purpose of validation as well as illustration. In the first case, we process six NAC image L/R pairs, all of which provide a reasonable, and thus repeated, coverage of the CE-3 landing site. This allows us to rigorously verify the consistency of the reconstructed DTMs under varying illumination conditions. In another, more challenging application, the tested model is applied to the much wider, whole landing area of CE-4. The full coverage of the region requires multiple adjacent observations with partial or even marginal overlap. We present the output DTM of 1.5 m resolution of the CE-4 landing area based on eight neighboring NAC pairs, which demonstrates not only the performance of the proposed approach but also its imminent practicality for large-area application.

The rest of this article is organized as follows. In Section II, we present in detail our reconstruction pipeline and improved deep learning network for large-area pixel-resolution lunar surface DTM generation. In Section III, the experimental results and analyses are provided to exhibit the capability of the presented method. Finally, Section IV concludes this article.

II. DEEP LEARNING-BASED LARGE-AREA LUNAR SURFACE RECONSTRUCTION

A. Reconstruction Pipeline

The entire reconstruction pipeline based on our improved CNN architecture to retrieve the lunar surface large-area pixel-resolution DTM is shown in Fig. 1. First, data preprocessing is performed on NAC images and SLDEM to eliminate the geometric inconsistency between them, and clipped into small normalized tiles to meet the input requirements of our deep learning network. Then, we use the well-trained model of our proposed CNN architecture to estimate the DTMs from the input tiles. The CNN architecture and model training are the core of this work, which will be introduced in Section II-B. Finally, we

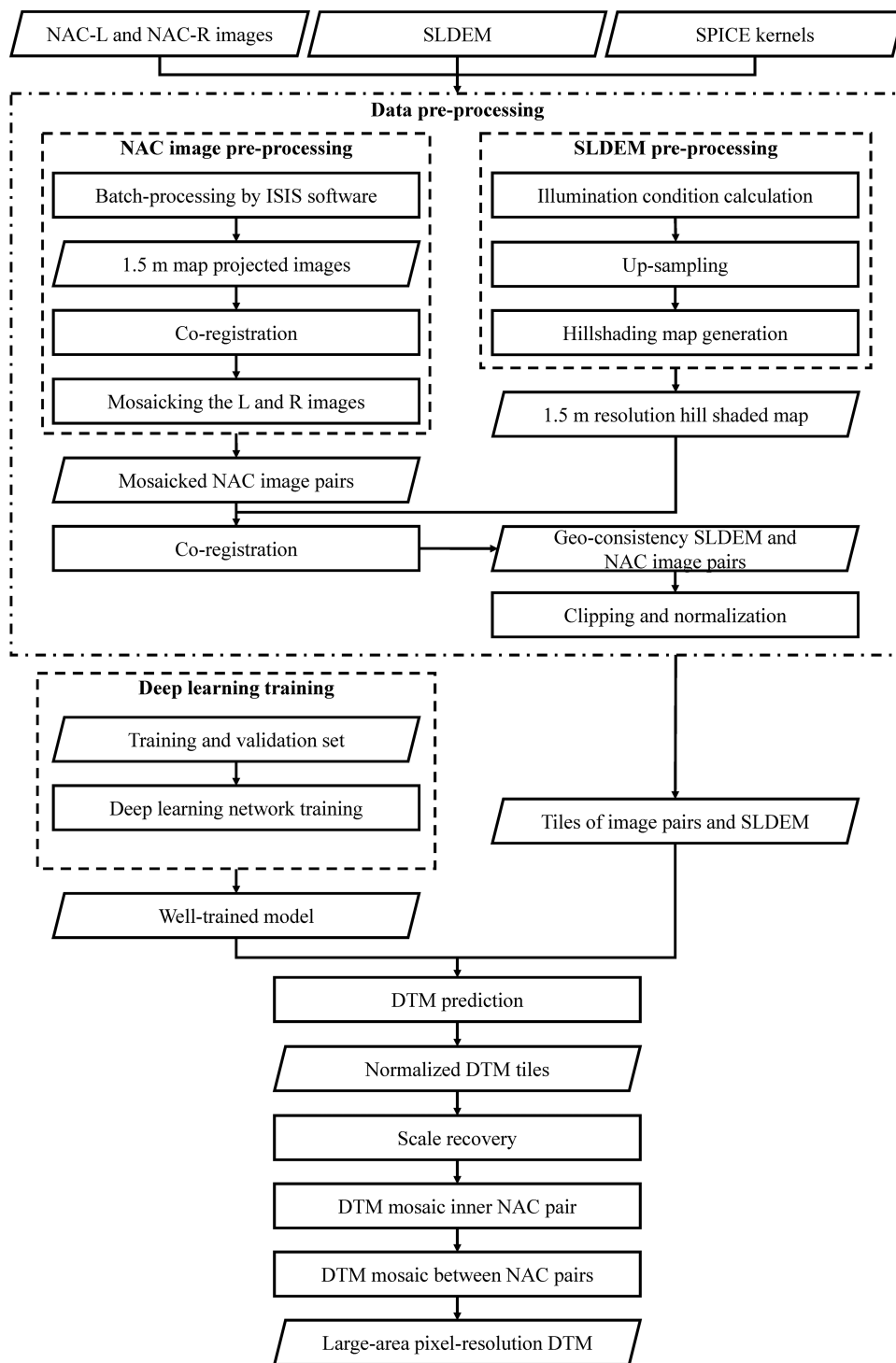


Fig. 1. Entire reconstruction pipeline for the large-area pixel-resolution DTM from NAC images and SLDEM.

recover the predicted, normalized DTM tiles to the real scale, which are subsequently mosaicked to large areas.

1) *Data Preprocessing*: The reconstruction pipeline needs only single-view high-resolution images and corresponding coarse-resolution DTMs as input to derive pixel-resolution DTMs, i.e., with the same ground sample distance of input images. The NAC images feature the highest resolution of orbiter images available and offer almost complete coverage of the lunar

surface in single view [12]. The NAC-L and NAC-R images are acquired simultaneously with ~ 135 pixels overlapping in the cross-track direction of the LRO orbit [14]. As our approach requires no prior knowledge of camera models, observations through the two lenses with independent camera models can be mosaicked together directly to increase the swath coverage. The NAC images used for DTM reconstruction in this article are all the mosaicked images from NAC-L and NAC-R (hereafter

designated as “NAC image pairs”). On the other hand, the SLDEM is a combined product of LOLA measurements and photogrammetric models derived from SELENE terrain camera images covering latitudes within $\pm 60^\circ$ [26].¹ While it has a relatively low spatial resolution of 60 m, the SLDEM is more accurate and suffers from fewer artifacts than the LOLA DTM [26]. We select it as the input coarse-resolution DTM.

The NAC experimental data record images can be downloaded from NASA’s Planetary Data System and batch processed by Integrated System for Imagers and Spectrometers software to attach the SPICE kernels (*spiceinit* tool) [27], perform radiometric corrections (*Ironacal* tool), remove echo effects (*Ironacecho* tool), and obtain the map-projected products with 1.5 m resolution (*cam2map* tool).

An affine transformation model [28], displayed in the following, is used to coregister the images to remove geometric discrepancies between the NAC observations in both intra and intertrack cases as well as the SLDEM:

$$\begin{aligned} r_r &= a_0 + a_1 c_s + a_2 r_s \\ c_r &= b_0 + b_1 c_s + b_2 r_s \end{aligned} \quad (1)$$

where (r_r, c_r) is the geographic location of the reference image, (r_s, c_s) is the geographic location of the source image, $a_0, a_1, a_2, b_0, b_1,$ and b_2 are affine transformation parameters, which are calculated from the tie points between the reference and source images. The SLDEM is the reference for the whole coregistration process. The NAC-L images are the reference for the NAC-R images. First, all the NAC-R images are coregistered to their NAC-L counterparts before mosaicking to NAC image pairs. The NAC images and SLDEM have a large difference in resolution and represent different modalities of surface, which makes it difficult to coregister them directly. To overcome this issue, the SLDEM is resampled to match the resolution of the NAC images. Based on the upsampled SLDEM, a simulated image is generated using hill shading techniques [29] to resembling the illumination conditions at the times of the NAC image acquisition. The NAC image pair in the central part of the whole area is first registered to the simulated image of SLDEM and then used as the reference for the neighboring images. Then, the new registered image pair becomes new reference for subsequent adjacent pairs until all are registered. If the overlap between the registered image and the neighboring image is large enough, the registered image will become the new reference for ensuing registration with the adjacent image; otherwise, the image simulated with SLDEM is used as a complementary reference.

The size of the input data for the deep learning network is dependent on the model architecture and constrained by machine memory. The present study uses a single Nvidia RTX3060 graphic card with a 12 GB memory. The NAC image pairs and corresponding SLDEMs are clipped into small overlapping tiles (256×320 pixels per tile, with 100 overlapping pixels in the horizontal and vertical directions). We normalize the gray values of the image tiles to $[0, 1]$ [see (2)], and accordingly, the heights

of SLDEM tiles to standard normal distribution, and subtract the minimum value for each normalized SLDEM tiles [see (3)] to ensure that the distribution of our input is consistent with the dataset used to training the deep learning model

$$\hat{I} = (I - \min(I)) / (\max(I) - \min(I)) \quad (2)$$

$$\hat{D} = (D - \mu) / \sigma - \min((D - \mu) / \sigma) \quad (3)$$

where I and \hat{I} represent the original and normalized images, respectively, D and \hat{D} represent the original and normalized SLDEMs, respectively, and μ and σ represent the mean and standard deviation of the original SLDEM heights.

2) *Construction of Large-Area Pixel-Resolution DTM*: The preprocessed small tiles of NAC image pairs and SLDEM are input to the well-trained model of our improved CNN architecture (see Section II-B for detailed description) to predict the pixel-resolution DTM tiles. The DTM tiles are in normalized scale with small size. After DTM prediction, the scale recovery and DTM mosaicking are carried out to construct the large-area DTM.

First, we use the mean and standard deviation value of reference SLDEM elevations to recover the initial scale of the normalized predicted DTMs

$$\text{DTM}_{\text{init_scale}} = \sigma * \text{DTM}_{\text{norm_scale}} + \mu \quad (4)$$

where $\text{DTM}_{\text{init_scale}}$ is the initial recovered DTM, and $\text{DTM}_{\text{norm_scale}}$ is the predicted DTM in normalized scale.

In (3), we add a bias to ensure that the heights of normalized SLDEM are all greater than or equal to 0, the initial recovered DTM is missing this bias. Besides, the pixel-resolution predicted DTMs derived from the network contain abundant high-frequency elevation details. In contrast, the coarse-resolution reference only provides low-frequency information. As a result, there are still varying degrees of vertical deviation between initial recovered DTMs and the reference. We use the *pc_align* tool in Ames Stereo Pipeline (ASP) to align the initial recovered DTMs to SLDEM in order to make the elevation between adjacent tiles consistent [30].

Then, we adopt the *dem_mosaic* tool in ASP to mosaic the DTM tiles to large area. First, we mosaic the DTM tiles from inner NAC pairs to derive DTM mosaics for the individual NAC image pairs. Then, we mosaic the DTM mosaics of adjacent NAC pairs to derive the whole large-area DTM mosaic. Fig. 2 shows that the *pc_align* step can effectively suppress artifacts caused by vertical disparities between neighboring DTM tiles. The DTM prediction, scale recovery, and DTM mosaicking can be implemented simultaneously between different NAC image pairs.

B. Deep Learning Method for DTM Reconstruction

1) *Improved Network Architecture*: As the high-resolution orbiter imagery and coarse-resolution DTM characterize different modalities of the lunar surface, Chen et al. [25] proposed a dual-branch CNN architecture with a weighted sum loss function consisting of two loss terms to estimate pixel-resolution DTMs

¹Downloaded from <http://imbrium.mit.edu/EXTRAS/SLDEM2015/>

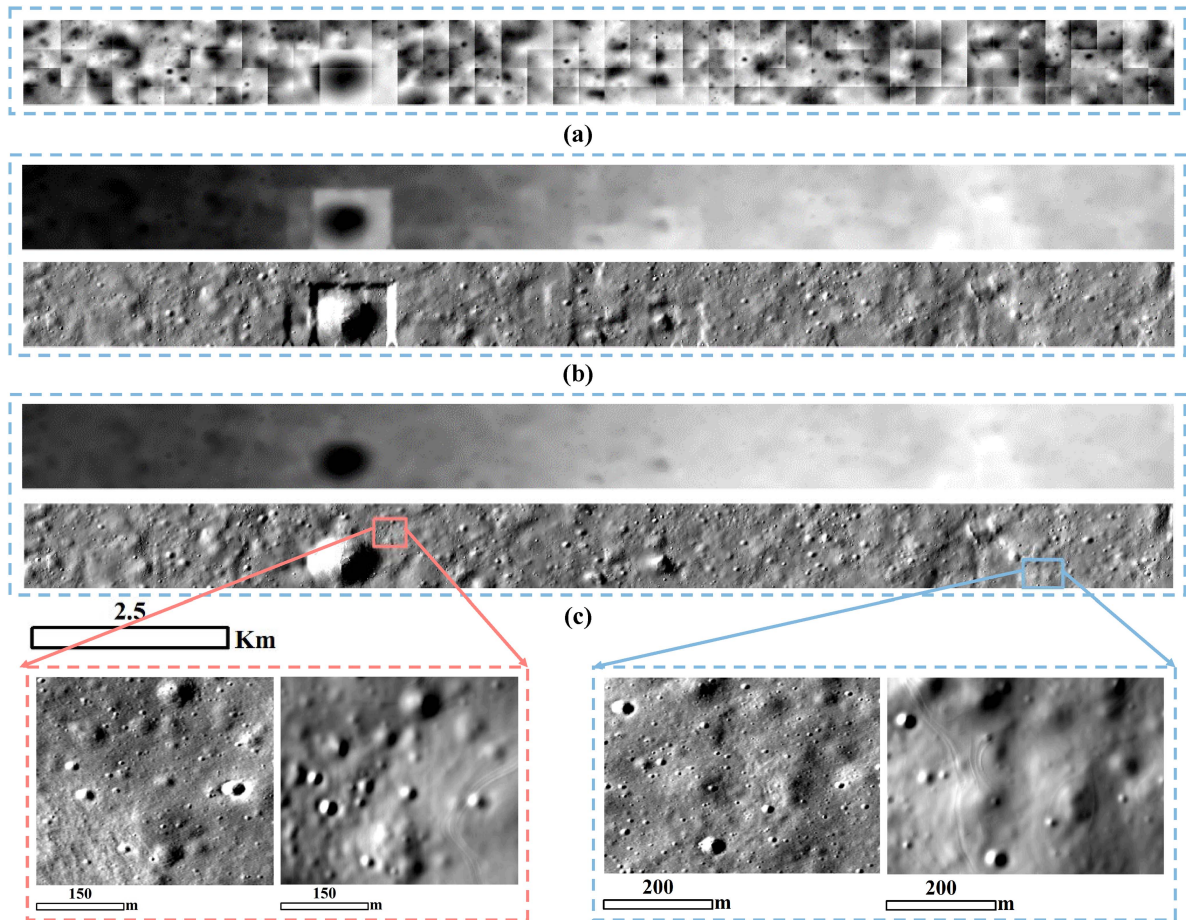


Fig. 2. Example of DTM mosaic. (a) Initial recovered DTM tiles. (b) DTM mosaic and the derived hill shaded map without applying *pc_align*. (c) DTM mosaic and the derived hill shaded map after applying *pc_align*. The bottom panels present the zoom-in views of original NAC images and the hill shaded maps after applying *pc_align* in two local areas.

from single-view images and coarse-resolution DTMs. In its network, two independent encoder branches, high-resolution image encoder branch (HIEB) and coarse-resolution DTM encoder branch (CDEB), were established. As the input images feature fine texture details, the HIEB is built based on the ResNet50 architecture [31], which has been broadly employed in deep learning-based depth estimation from high-resolution single images to extract abundant textural features [32]. By contrast, the CDEB is built with five stacked simple convolution blocks to extract the vertical features from input DTMs. The outputs from the encoder branches are merged before being processed by the decoder module, in which five stacked upsampling blocks are designed in order to restore the resolution of the predicted DTM at the same level as the input. The first four are upprojection blocks [33], and the last one is simple nearest upsampling block [25]. At the same time, in order to take the full advantage of multiscale terrain features extracted from the HIEB to assist in predicting DTMs with finer terrain details, the input for the last four blocks of decoder module is concatenated by the output from the previous block and the counterpart in the HIEB with the same spatial resolution. The final step includes a convolution layer and a rectified linear unit activation function [34] to

produce a single-channel output, which will be the predicted DTM.

Compared with networks with image-only input or single-branch architecture with coarse-resolution DTM as constraint, the method of Chen et al. [25] provides a higher elevation accuracy and sensitivity to terrain details. However, when it comes to ultradetails, e.g., extremely small-scale craters with a few pixels, the model capability is still limited and may suffer from the occurrence of perceptible noises in certain areas. This article aims at two objectives: 1) the first convolution layer of the HIEB in [25] contains more specific texture features than the layers that follow. The output of this layer is directly connected to the last block of the decoder module to enable the model to predict more small-scale terrain details. Here, we reduce the convolution kernel of this layer from 7×7 pixels to 3×3 pixels to enable the model to extract more local details. 2) The deconvolution layers in upprojection blocks for spatial upsampling is prone to generating excessive noise [24], [35]. Here, we adopt the nearest neighbor interpolation algorithm instead of the deconvolution operation for spatial upsampling. The improved network architecture is shown in Fig. 3.

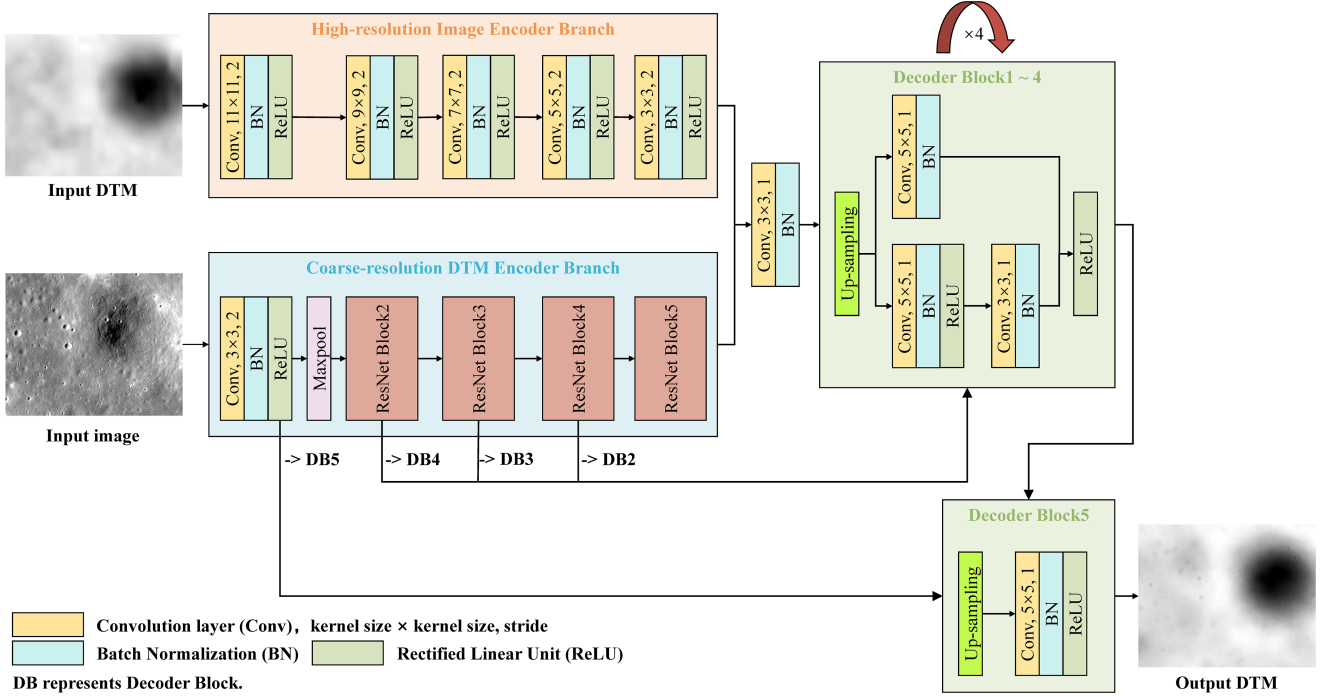


Fig. 3. Our improved network architecture.

2) *Loss Function*: A suitable loss function is a necessity for the learning process, which measures the error of model prediction and, thus, offers the criteria for optimization. The design and implementation of loss functions has significant impact on the overall DTM prediction performance. Many studies in the literature sought to define a loss function that minimizes the residuals of the elevation values between the ground truth DTMs (z) and the predicted DTMs (\hat{z}) while also penalizing distortion of high-frequency details on the surface [24], [25], for example, craters, which are the most common terrain features on the lunar surface [10].

For training the network, we define the loss function (l) between z and (\hat{z}) as the weighted sum of three loss terms

$$l(z, \hat{z}) = \rho l_{\text{elev}}(z, \hat{z}) + \varphi l_{\text{grad}}(z, \hat{z}) + \omega l_{\text{norm}}(z, \hat{z}) \quad (5)$$

where ρ , φ , and ω are the weight factors. The first two loss terms are the same as in [25].

l_{elev} measures the mean absolute error between z and (\hat{z})

$$l_{\text{elev}} = \frac{1}{n} \sum_{p=1}^n |z_p - \hat{z}_p| \quad (6)$$

where z_p and \hat{z}_p represent the pixel elevation values of the predicted DTM and the ground truth, respectively, n is the number of pixels in the ground truth and predicted DTM, and l_{grad} measures the sum of mean absolute gradient differences between z and (\hat{z}) in the x and y directions

$$l_{\text{grad}} = \frac{1}{n} \sum_{p=1}^n (|\hat{g}_x(p) - g_x(p)| + |\hat{g}_y(p) - g_y(p)|) \quad (7)$$

where \hat{g}_x and \hat{g}_y denote the gradients in the x and y directions of the ground truth, and g_x and g_y denote the gradients in the x and y directions of the predicted DTM.

To further improve the performance and predict more accurate terrain details, we consider yet another loss term l_{norm} for training

$$l_{\text{norm}} = \frac{1}{n} \sum_{p=1}^n \left(1 - \frac{\langle \hat{n}_p, n_p \rangle}{\sqrt{\langle \hat{n}_p, \hat{n}_p \rangle} \sqrt{\langle n_p, n_p \rangle}} \right) \quad (8)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of vector, $\hat{n}_p = [-\hat{g}_x(p), -\hat{g}_y(p), 1]^T$, and $n_p = [-g_x(p), -g_y(p), 1]^T$.

l_{norm} measures the accuracy of the normal to the surface of the predicted DTM with respect to its ground truth. Although it is also computed and derived from depth gradients, it measures the angle between two surface normals of the predicted DTM and ground truth, and thus is sensitive to small surface undulations [36].

3) *Network Training Details*: The photogrammetric DTMs and orthorectified images (ORIs) derived from NAC stereo observations as well as SLDEM are used to construct the dataset for training and validating our improved network.² In order to objectively evaluate the impact of network modifications on DTM prediction performance, we use the same dataset with 30 000 samples, 90% as training set and 10% as validation set, as in [25]. In order to ensure convergence and enhance the effectiveness of our model, we normalized our samples using (2) and (3). To ensure that the proposed CNN architecture is robust to more variable illumination conditions, the training set was expanded by flipping images with 0.5 probability in

²Downloaded from https://wms.lroc.asu.edu/lroc/rdr_product_select

TABLE I
QUANTITATIVE EVALUATION OF METHODS PROPOSED IN [25] AND THIS STUDY IN TERMS OF $\delta_{0.1}$, $\delta_{0.25}$, $\delta_{0.5}$, REL, AND MSE

Method	$\delta_{0.1}\uparrow$	$\delta_{0.25}\uparrow$	$\delta_{0.5}\uparrow$	REL \downarrow	MSE \downarrow
Chen et al. [25]	0.357	0.623	0.798	0.184	0.340
Disabling modification 1 in Section II-B1	0.343	0.621	0.804	0.190	0.335
Disabling modification 2 in Section II-B1	0.352	0.620	0.799	0.193	0.350
Loss function without l_{norm}	0.365	0.626	0.800	0.192	0.382
Loss function without l_{grad}	0.360	0.627	0.805	0.194	0.370
This study	0.363	0.630	0.805	0.173	0.318

The best results are bolded.

the vertical and horizontal directions. This procedure, thus, introduces supplementary illumination conditions, with contrary solar azimuths, for the training. We trained the model for 20 epochs with a batch size of four. We initiate with a learning rate of 0.0001, which is reduced to its 10% after every five epochs. The weights in (5) are set empirically as $\rho = 1$, $\varphi = 0.5$, and $\omega = 0.5$, based on experiments.

For the assessment of model performance, we used the following metrics.

1) *Average relative error (REL)*:

$$\frac{1}{n} \sum_{p=1}^n \frac{|z_p - \hat{z}_p|}{z_p}. \quad (9)$$

2) *Mean squared error (mse)*:

$$\frac{1}{n} \sum_{p=1}^n (z_p - \hat{z}_p)^2. \quad (10)$$

3) *Threshold accuracy (δ_i)*:

$$\% \text{ of } z_p \text{ s.t. } (z_p - \hat{z}_p) = \delta_j < j \text{ for } j = 0.1, 0.25, 0.5. \quad (11)$$

In general, lower REL and mse or higher δ_i indicate better DTM prediction accuracy [32], [33]. The training proceeds to determine an optimal model, i.e., one with the highest accuracy when applied to the validation set, which was then used for DTM reconstruction.

Compared with [25], our improved network improves all five of these metrics on the validation set, as shown in the first and last rows of Table I. Besides, we test the effects of disabling one of the three changes in this study with respect to [25]. The second and third rows of Table I present the results disabling one of the two modifications in our model introduced in Section II-B1. The fourth and fifth rows present the results without l_{grad} or l_{norm} in the loss function to train the model. Overall, all three changes make contributions to improve the DTM prediction accuracy, and our improved model can achieve the highest accuracy in the five metrics, except for $\delta_{0.1}$. Nevertheless, there is only a marginal degradation in $\delta_{0.1}$ compared with the case of loss function without l_{norm} . We use a 5-m resolution photogrammetric DTM tile with 256×320 pixels as ground truth [see Fig. 4(b)],³ the same resolution ORI corresponding to the tile

[see Fig. 4(a)] and SLDEM as the input to predict 5-m resolution DTMs from [25] and this study method, respectively. Fig. 4 displays the resultant DTMs. From the hill shaded maps, we can see that the result derived from [25] can retrieve the terrain details, but it suffers from noticeable noise that should be smooth, e.g., the areas indicated by the yellow arrows in Fig. 4(e). In contrast, our method, as shown in Fig. 4(g), can retrieve more detailed and accurate terrain structure information, and eliminate or alleviate the noise present in [25]. Our result is more consistent with the photogrammetric DTM. The results from deep learning method tend to be oversmoothed [25], and we further utilize the method introduced by Bland et al. [37] to evaluate the effective resolution of our results. By comparing the input DOM [see Fig. 4(a)], the hill shaded maps derived from photogrammetric DTM [see Fig. 4(c)] and our result [see Fig. 4(g)], our method can extract the terrain features larger than 6 pixels. Therefore, we use the well-trained model of our improved network to predict DTMs for large-area DTM generation, see Section II-A2.

III. RESULTS AND ANALYSIS

A. Study Area and Experimental Data

For testing the performance of our proposed deep learning-based reconstruction for large-area pixel-resolution lunar surface DTMs, two Chinese lunar exploration landing areas, located on the near side (CE-3) and far side (CE-4) of the Moon, were used as the study areas. With their touchdown at 44.119°N , 19.513°W on the Mare Imbrium on December 14, 2013, the CE-3 lander and the accompanying rover, ‘‘Jade Rabbit,’’ achieved the first soft landing on the lunar surface since Luna 24 in 1976 [10], [38]. Robustness to images acquired under different conditions, especially the varying illumination conditions, is a prerequisite for the ability to perform terrain construction over large areas. We selected six NAC image pairs, repeated coverage of the CE-3 landing site, with different resolutions, slew angles, and illumination conditions to verify the robustness of the presented algorithm. The successful landing of the CE-4 mission in Von Kármán crater in the northwestern South Pole-Aitken basin on January 3, 2019, was the very first to descend on the far side of the Moon [1]. In total, eight adjacent NAC image pairs, which can cover the whole CE-4 landing area ($177.6^\circ\text{E} \pm 1.2^\circ$, $45.5^\circ\text{S} \pm 0.5^\circ$), were selected to verify the performance of our presented method on the construction of

³Downloaded from https://wms.lroc.asu.edu/lroc/view_rdr/NAC_DTM_A12LMAS

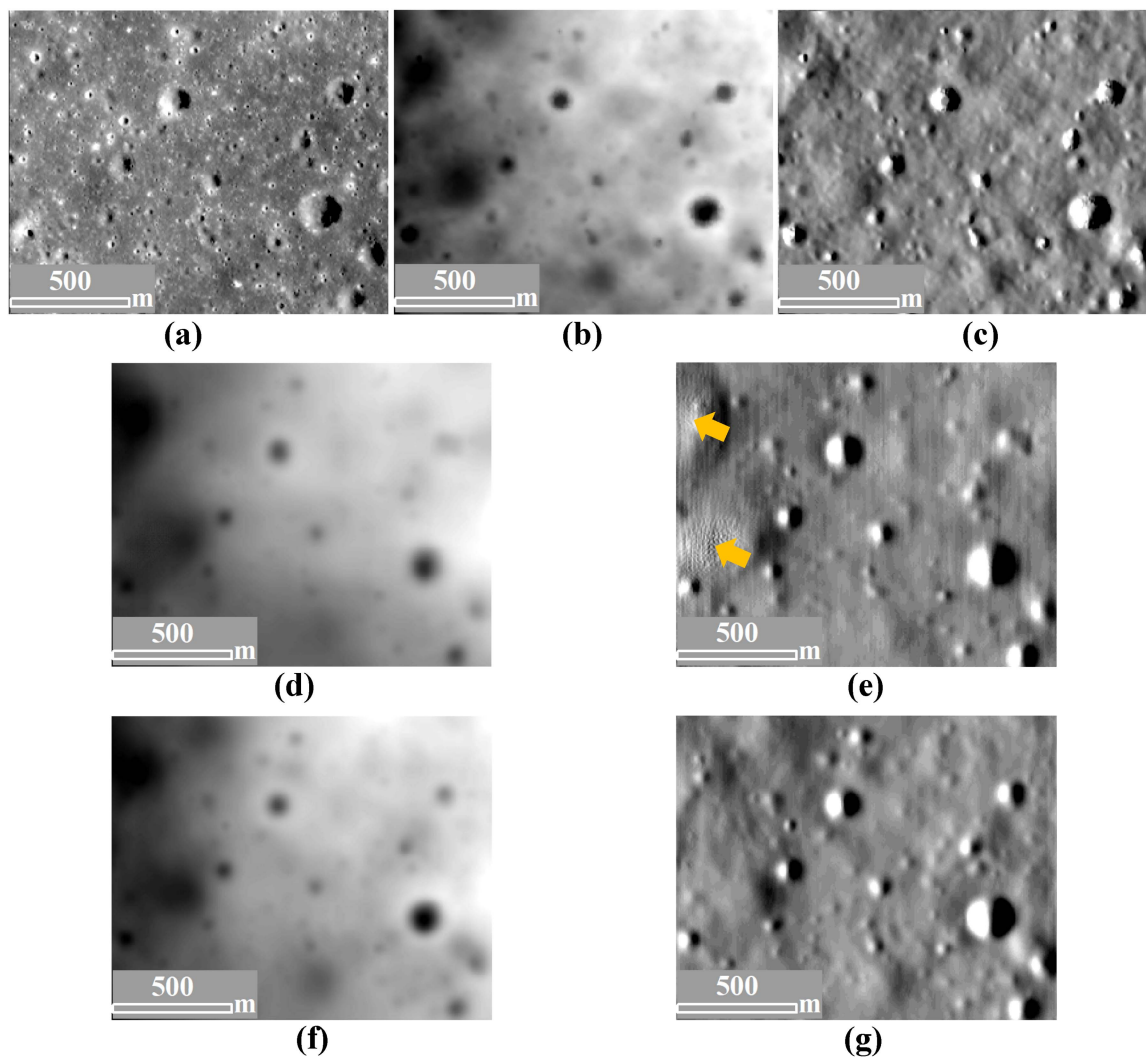


Fig. 4. Example of DTM prediction result with 256×320 pixels. (a)–(c) Show the 5-m resolution photogrammetric DOM, DTM, and the derived hill shaded map. (d) and (e) are the 5-m resolution DTM and the hill shaded map derived from Chen et al. [25]. (f) and (g) show the 5-m resolution DTM and the hill shaded map derived from this study.

large-area high-resolution DTM. The image names and their main parameters, such as pixel resolution, slew angle, emission angle, solar azimuth, and altitude, are listed in Table II. The layouts of the ground coverage of all images used in the study are shown in Fig. 5.

B. Results and Analysis of CE-3 Landing Site

1) *Retrieved DTMs Under Varying Illumination Conditions:* In this experiment, we describe the performance and robustness of the proposed method to generate DTMs from single-view NAC image pairs acquired under different illumination conditions. The typical photogrammetric DTM [see Fig. 6(a)] derived by NAC stereo observations (M1144936321 and M1144922100)⁴ and SLDEM [see Fig. 6(b)] are used as reference. The original resolutions of the six used NAC image pairs (12 L/R images, multiple coverage of CE-3 landing site) are

⁴Downloaded from https://wms.lroc.asu.edu/lroc/view_rdr/NAC_DTM_CHANGE3

from 0.92 to 1.48 m with different solar azimuths and altitudes. We first downsample all images to the same resolution of 1.5 m. Fig. 6(d)–(i) shows the six mosaicked image pairs (left-hand side), 1.5-m resolution DTMs derived from our method (middle), and the derived hill shaded maps based on the DTMs (right-hand side). The generated DTMs almost extend to the total coverage, i.e., the outermost bounds, of the corresponding NAC pairs. The original M1154358210 image pair features the lowest resolution (1.48 m), and the generated DTM covers an area of 1, 505 km². The original M1376551591 image pair features the highest resolution (0.92 m), and the generated DTM covers an area of 612 km². From Fig. 6, we observe that our six derived DTMs can reveal robust, correct, and consistent terrain trends and features from different single-view NAC image pairs under varying illumination conditions. The generated terrain patterns and distinct features, such as craters from hundreds of meters in size, correspond well to those in the original NAC image pairs. Even for the mosaicked NAC image pair which has obvious image intensity differences in the NAC-L

TABLE II
PARAMETERS OF THE USED NAC IMAGES FOR THE CE-3 AND CE-4 LANDING AREAS

Landing area	Image name (L/R)	Resolution (m)	Slew angle (°)	Emission angle (°)	Solar azimuth (°)	Solar altitude (°)
CE-3	M1154358210	1.48	0.010	1.78	106.444	15.511
	M1157862508	1.46	-3.009	2.06	253.066	14.823
	M1180251193	1.45	13.046	12.33	159.498	44.732
	M1328449896	1.08	-0.011	1.74	206.784	43.862
	M1348370554	1.11	-0.007	1.75	118.591	26.734
	M1376551591	0.92	-0.010	1.73	133.308	35.042
CE-4	M1298909402	0.81	-0.007	1.72	349.715	43.038
	M1301264579	0.84	0.001	1.72	28.331	40.551
	M1301271610	0.84	-0.009	1.72	27.095	40.876
	M1303612811	0.87	-0.004	1.72	60.078	26.719
	M1303640934	1.15	-39.875	43.92	56.481	29.092
	M1324774291	0.95	0.003	1.73	294.966	20.576
	M1329462206	0.96	-0.010	1.73	1.783	43.669
	M1383485778	1.60	0.007	1.74	292.257	18.527

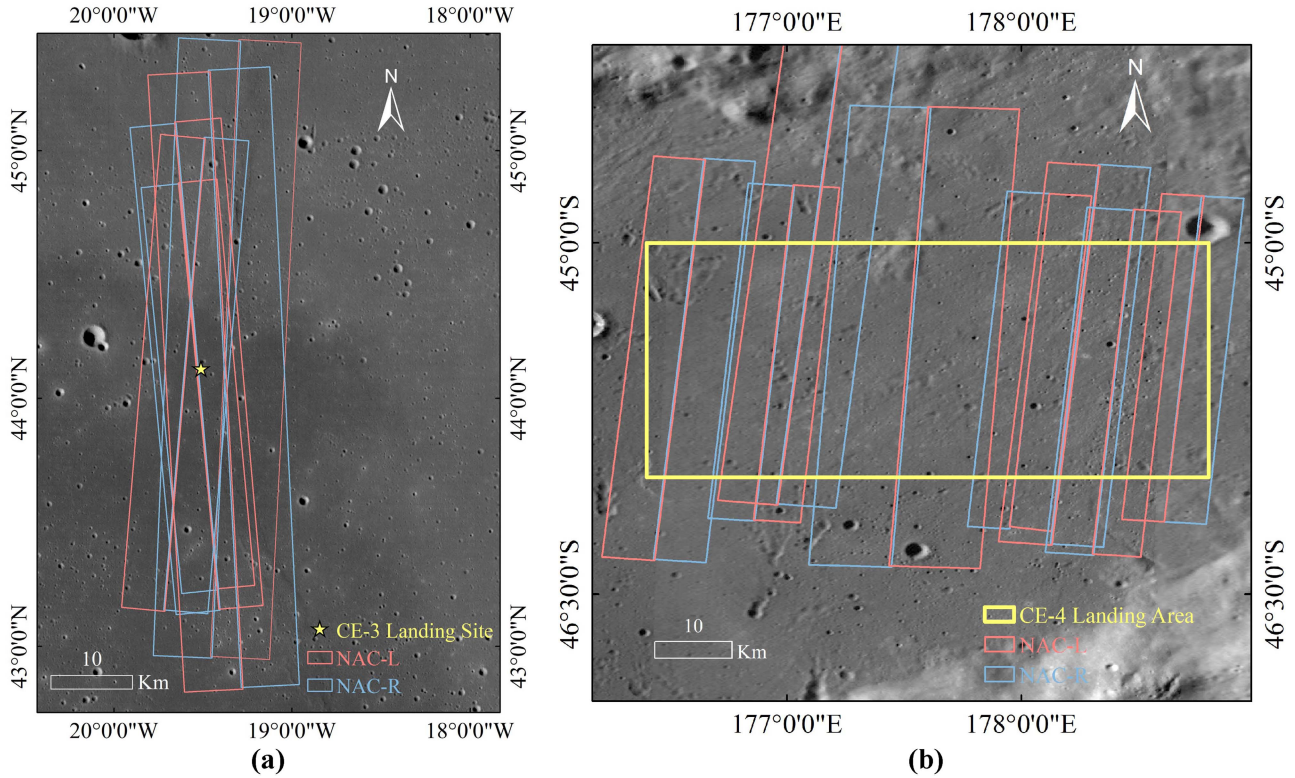


Fig. 5. Layouts of the ground coverage of (a) six NAC image pairs of the CE-3 landing site and (b) eight adjacent NAC image pairs of the CE-4 landing area.

and NAC-R parts, our method still yields topographically continuous result [see Fig. 6(i)]. Besides, we compare the two reference DTMs and our six derived DTMs through profile analysis [see Fig. 6(c)]. The comparative results show, in general, that our generated DTMs are more consistent with the SLDEM than the photogrammetric DTM. The elevations are generally aligned for different depths of craters between the references and our DTMs. From the enlarged part of the

profiles, we can see our six DTMs are highly aligned with each other.

Furthermore, we select two local areas, which cover the intensity change boundary in Fig. 6(i), to present detailed profile comparisons of small-scale features among the reference photogrammetric DTM, SLDEM, and our six generated DTMs (see Fig. 7). From the profile analysis results, we see the profile of DTM derived from NAC image M1376551591 (the green bolded one) is continuous and exists no subtle artefacts during the

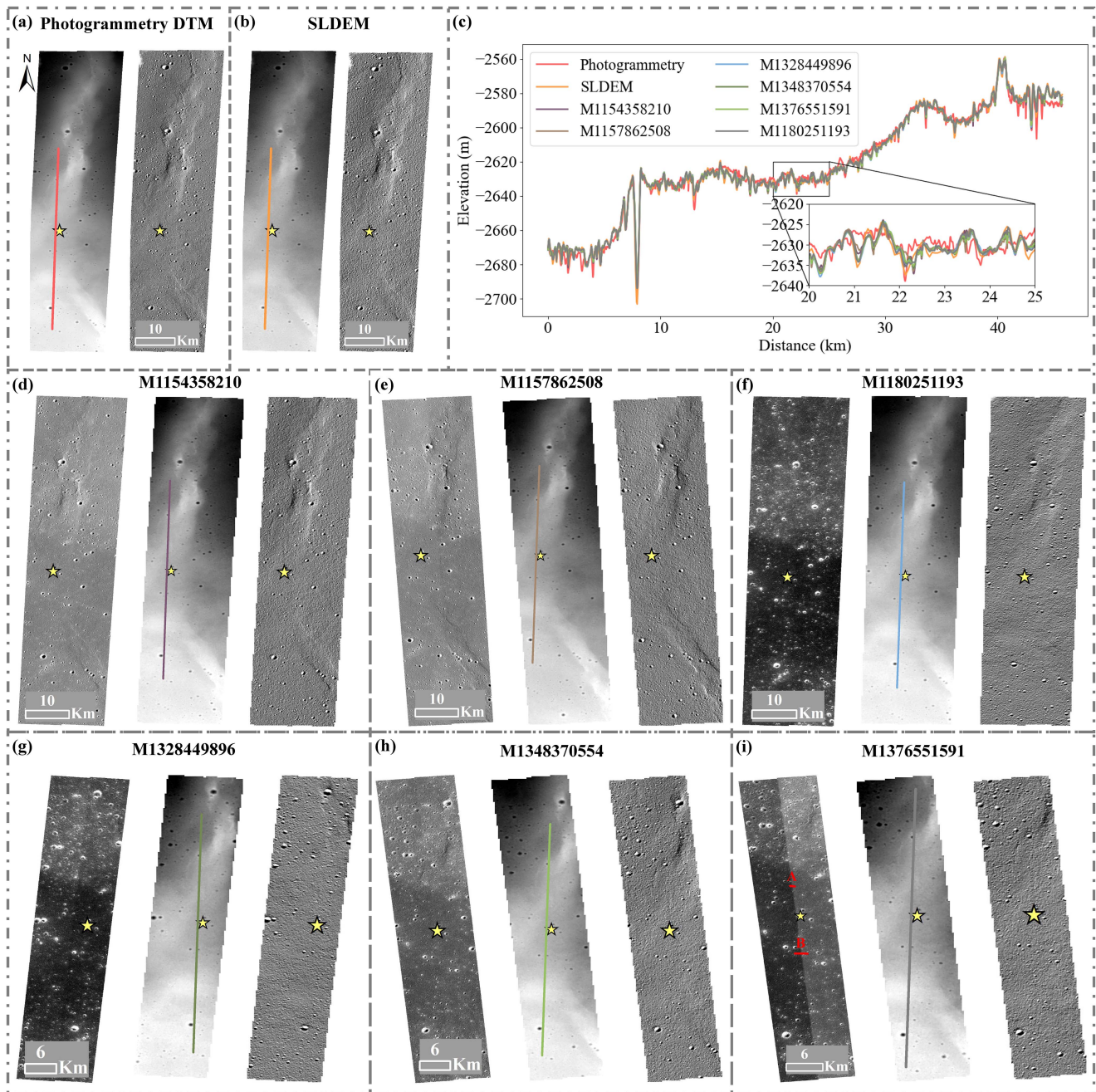


Fig. 6. Comparison of photogrammetric DTM, SLDEM, and our six generated DTMs in the CE-3 landing site. (a) Photogrammetry-derived DTM and the corresponding hill shaded map. (b) SLDEM and the corresponding hill shaded map. (c) Profile analysis of our DTMs, photogrammetric DTM, and SLDEM; the color lines in (a), (b), and (d)–(i) represent the locations of the profiles. (d)–(i) show the six used NAC image pairs, the generated DTMs, and the hill shaded maps, respectively. The solar azimuth and altitude of the hill shaded maps derived from our DTMs are consistent with the original image pairs. The illumination conditions of hill shaded maps derived from the reference DTMs are consistent with M1154358210.

intensity change boundary. The profiles of our six reconstructed DTMs are still self-consistent even in local areas. They are able to recover the topographical undulations caused by the small-scale craters.

Furthermore, the statistics (mean and standard deviations) of the absolute difference among the two reference DTMs and our six derived DTMs are shown in Fig. 8. The absolute difference between the photogrammetric DTM and SLDEM shows the largest mean and standard deviation errors, which

are 3.02 and 2.46 m, respectively. As our method is constrained by SLDEM, the absolute difference between our six generated DTMs and SLDEM is less than the absolute difference with the photogrammetric DTM. The mean absolute difference between our six generated DTMs with SLDEM is 1.56 (mean) and 1.57 m (standard deviation), about 1 pixel in the image space. The mean absolute difference between our six generated DTMs with photogrammetric DTM is 2.80 (mean) and 2.30 m (standard deviation).

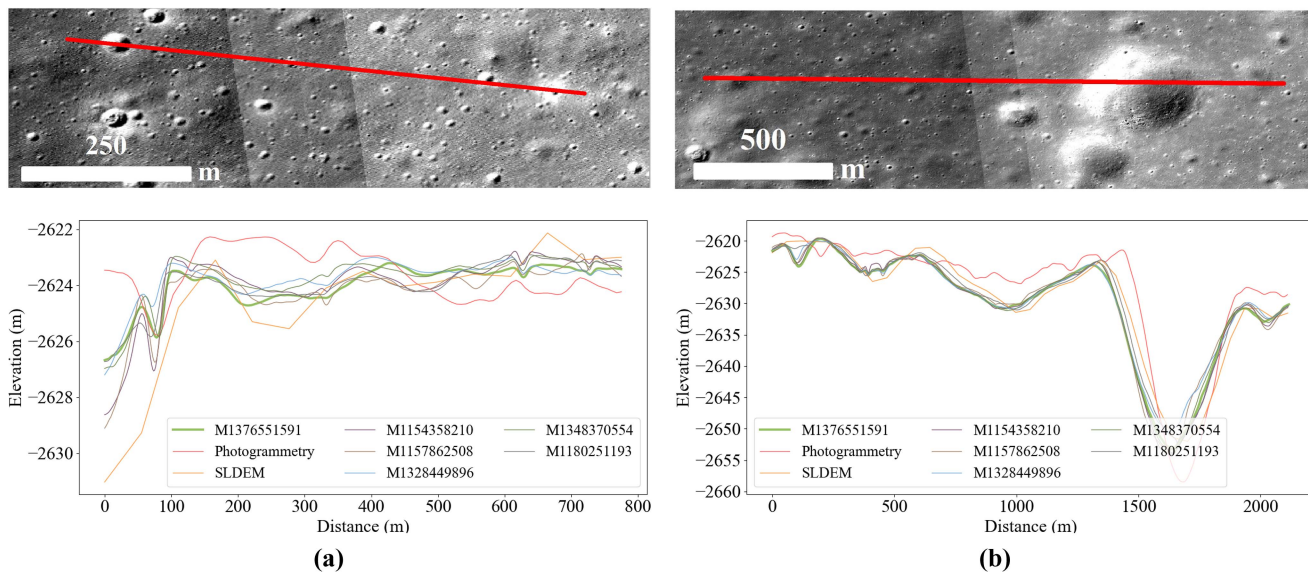


Fig. 7. Profile analysis of two local areas covering the intensity change boundary in Fig. 6(i). (a) Profile A in Fig. 6(i). (b) Profile B in Fig. 6(i). The top panel displays the NAC image pair: M1376551591, and the red lines represent the profiles. The bottom panel shows the profile analysis results of the photogrammetric DTM, SLDEM, and our generated DTMs.

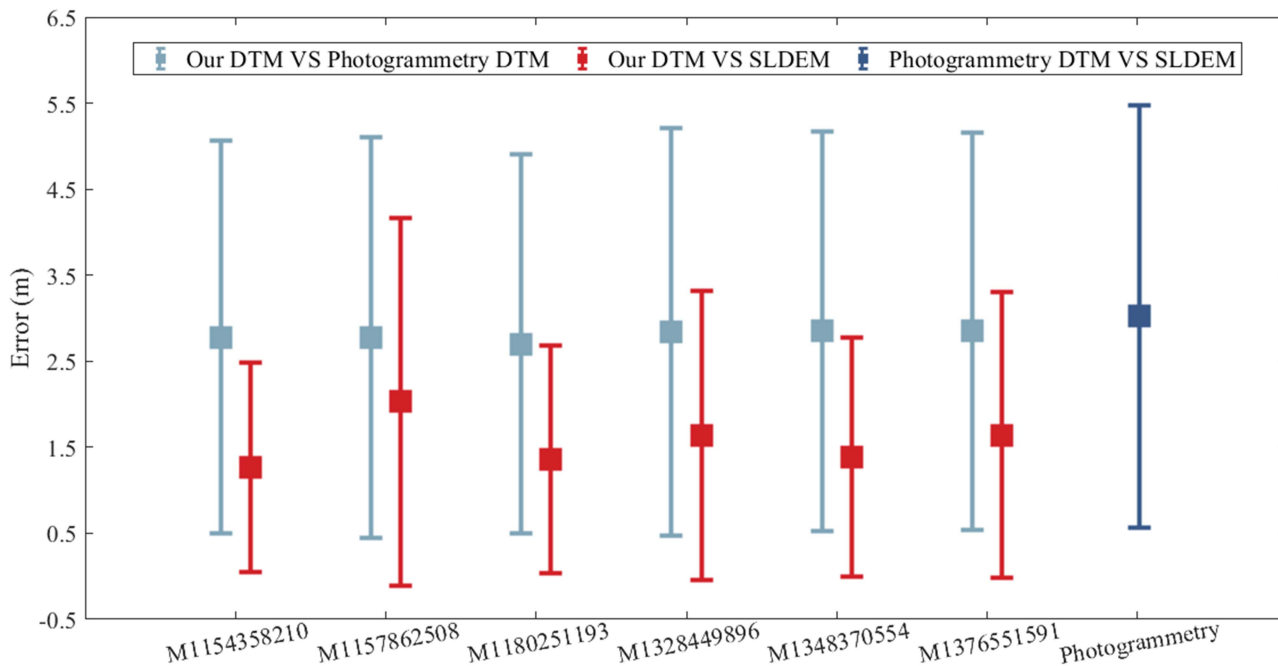


Fig. 8. Statistics (mean and standard deviations) of the absolute difference among our generated DTMs, photogrammetric DTM, and SLDEM. The squares at the middle of the error lines represent the means, and the distance above or below the squares represents the standard deviation.

Among the six used NAC image pairs in the CE-3 landing site, M1180251193 and M1376551591 feature large slew angle difference and small illumination difference, respectively. We, therefore, select M1180251193RE and M1376551591RE to generate photogrammetric DTM with 206 km² by ASP for further comparative analysis. We use the *bundle_adjust* tool to correct the position and orientation of these images and select Smooth semiglobal matching algorithm [39] with 7 × 7 correlation kernel for dense matching, and the derived DTM is

aligned to SLDEM by using the *pc_align* tool. Fig. 9(a) shows the generated DTM and the derived hill shaded map. We see that it can retrieve the terrain trends and large-scale features, but compared with the reference photogrammetric DTM [see Fig. 6(a)], it suffers from clear artifacts, for example, the obvious striping phenomenon. Furthermore, the zoom-in views of two local areas are shown in Fig. 9(b) and (c). In addition to the hill shaded maps of the reference photogrammetric DTM, our ASP-derived photogrammetric DTM and our generated

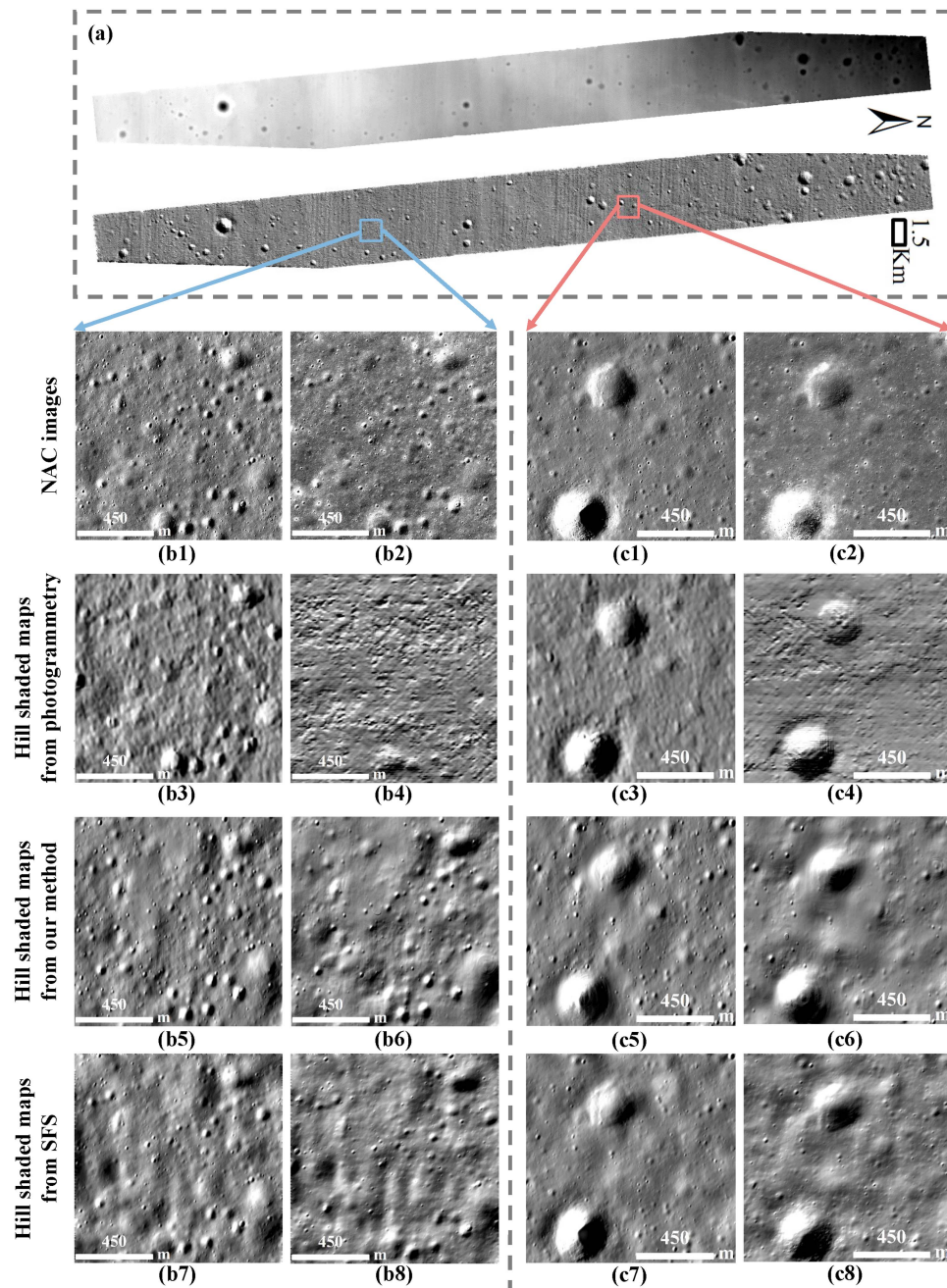


Fig. 9. Detailed comparison between the photogrammetric, SFS, and our generated DTMs. (a) Resultant photogrammetric DTM derived from ASP and the corresponding hill shaded map. (b), (c) Zoom-in views of two local areas with about 1.2×1.2 km in size, where (1) and (2) represent two original NAC images: M1154358210 and M1376551591, respectively; (3) and (4) represent the hill shaded maps of the reference photogrammetric DTM and ASP-derived photogrammetric DTM; (5) and (6) represent the hill shaded maps of our deep learning DTMs derived from (1) and (2); (7) and (8) represent the hill shaded maps of SFS DTMs derived from (1) and (2). The illumination conditions in (3), (5), and (7) are the same as in (1), and (4), (6), and (8) are the same as in (2).

deep learning DTMs, we also produced the 1.5-m resolution SFS DTMs of these local areas based on ASP by using M1180251193RE and M1376551591RE as input constrained by SLDEM, respectively. Their hill shaded maps are illustrated in the last row of Fig. 9 for comparison. It can be noticed that our generated photogrammetric DTM [see Fig. 9(b4) and (c4)] only clearly reflect the large-scale craters, e.g., the craters with diameters of about 300–400 m in Fig. 9(c4). There are ubiquitous artifacts in its derived hill shaded map. It is completely unable to reveal the topographical details. The reference photogrammetric

DTM derived from suitable stereo observations [see Fig. 9(b3) and (c3)] can reflect more terrain details, but it also shows clear matching errors and fail to retrieve the small-scale craters due to the limitation of its own coarse resolution (about three times the resolution of the original images). In contrast, our deep learning derived pixel-resolution DTMs retrieve the small-scale craters. Clearly, our generated DTMs are more consistent with the original NAC images compared with the photogrammetric DTMs. Furthermore, with decreasing solar altitude, shading caused by surface topography becomes more significant and textures

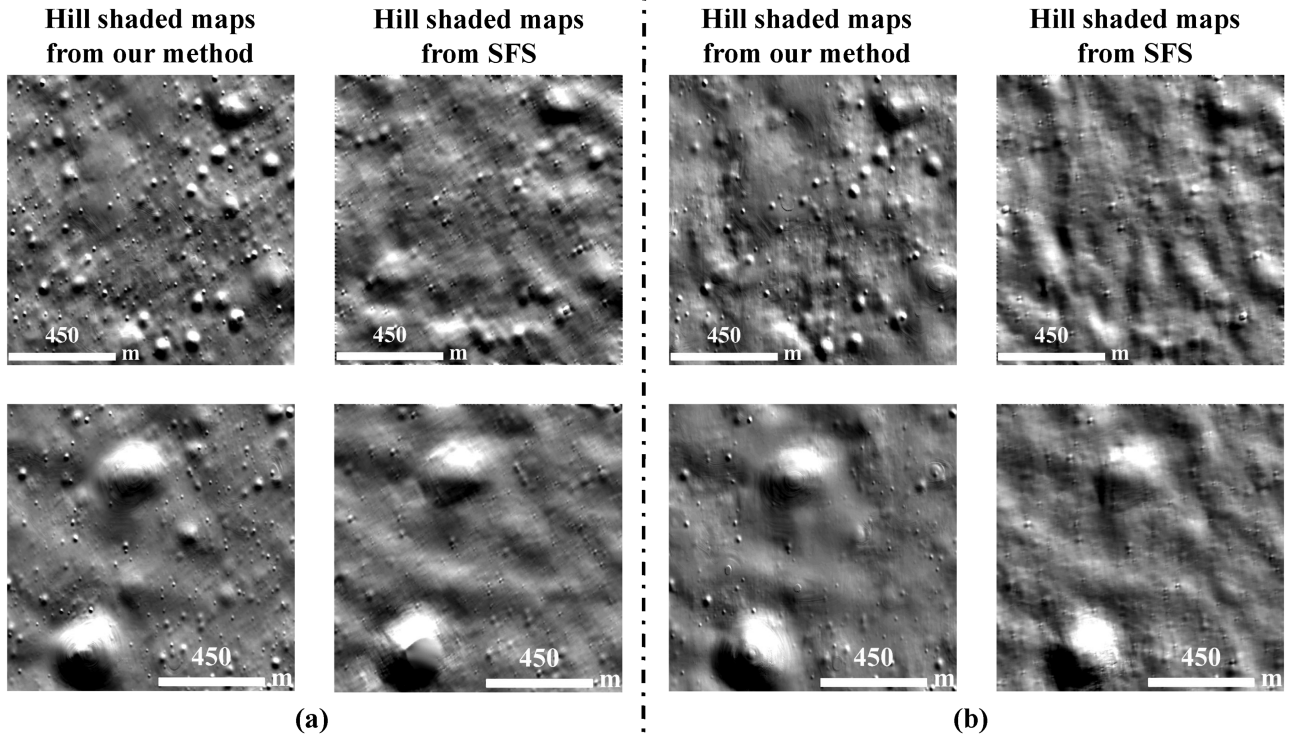


Fig. 10. Hill shaded maps of two local areas in Fig. 9 derived from our generated deep learning DTMs and SFS DTMs in cross-sun directions. (a) Results derived from an NAC image: M1154358210. (b) Results derived from an NAC image: M1376551591.

reflecting the terrain undulations in the NAC images become more obvious. From the results, we see that the lower the solar altitude, the finer the extracted terrain details. For example, some terrain features are clearly recovered in Fig. 9(b5) and (c5) with 15.511° solar altitude. By contrast, they cannot be recovered in Fig. 9(b6) and (c6) with 35.042° solar altitude. From the SFS results, it can be noticed that they are able to retrieve finer local details along the sun direction than deep learning derived DTMs. The SFS DTMs are able to extract the terrain features with 6 pixels; in contrast, our results can only extract features large than 10 pixels.

In addition, we derive the hill shaded maps of our generated deep learning DTMs and SFS DTMs at cross-sun directions to further test the results of both methods, as shown in Fig. 10. It can be seen that our DTMs are able to retrieve robust and consistent terrain details along and cross the sun direction (the results along the sun direction please see the penultimate row of Fig. 9). In contrast, the results of SFS method tend to be less accurate in the cross-sun directions. The hill shaded maps of SFS DTMs fail to display some terrain details that are obvious in those derived along the sun direction and in the counterparts of our DTMs.

2) *Analysis of the Potential of Large-Area High-Resolution DTM Generation:* Large-area high-resolution DTM reconstruction for the lunar surface has been one of the most pressing research issues in the planetary mapping community. The landing areas for the past and future exploration missions, generally covered by multiple adjacent NAC image pairs, are the most studied regions of the Moon, holding great promise for high-resolution 3-D terrain reconstruction. Photogrammetric techniques require

at least stereo observations with optimal slew angles, reasonable overlapping coverage, and similar illumination conditions and resolutions to produce quality DTMs. However, NAC is not a camera with built-in stereo capability, and in general, photogrammetric DTM production suffers from input images with varying illumination conditions. Hence, there is a great lack of redundant stereo NAC images for large-area DTM generation through the photogrammetric technique. For example, while all six used NAC image pairs are capable of multiple covering the CE-3 landing site with reasonable overlapping coverage, not all the images were taken under geometry and illumination conditions suited for stereo reconstruction. Only the convergence angle between M1180251193 and other NAC images is greater than 10° . The differences in solar azimuth between some images are too large, even larger than 100° (e.g., M1154358210 and M1157862508). Therefore, while they share similar incidence and emission angles, the matching technique still fails. The large variation in solar altitude (or incidence angle) also limits the generation of high-quality DTMs [15]. Only parts of the CE-3 and CE-4 landing areas have suitable stereo NAC observations [10], [11]. Besides, the stereo coverage tends to have even smaller coverage than the original NAC images, which further restricts the coverage of the derived photogrammetry DTMs. At present, common photogrammetric stereo-techniques have limited use for large-area lunar surface reconstruction.

In contrast, our presented deep learning-based reconstruction method can generate pixel-resolution DTMs from single-view images constrained by coarse-resolution DTMs. These aforementioned issues specific to photogrammetric techniques can be avoided, because we require only single-view images

as input. The generated DTM coverage is nearly the same as the whole NAC image pair mosaicked from NAC-L and NAC-R parts, which takes full advantage of the dual-NAC configuration. While the derived DTMs from our method tend to be oversmoothed compared to the SFS method, our results show consistent and fine terrain features both along and cross the sun direction. Furthermore, our method can generate robust results under varying illumination conditions, which allows us to select more suitable images for DTM reconstruction with minimal effort, e.g., far less than would usually be required by photogrammetry. Currently, the coarse-resolution DTM and single-view NAC images can cover basically any area of interest. Therefore, the presented method is capable of making full use of these data for large-area DTM generation. Compared with photogrammetric and SFS techniques, the deep learning method can rapidly predict DTMs without setting any parameters for the new input images [21]. Nevertheless, the quantitative influence of the varying illumination geometry of the NAC images on the results as well as the high-precision automatic alignment of the images with geometric inconsistencies and different illuminations are the limitations to be considered for larger-area/near-global high-resolution DTM constructions.

C. Results and Analysis of CE-4 Landing Area

1) Retrieved DTMs Covering the Whole Landing Area:

In this section, we take the whole CE-4 landing area with 72.8×30.3 km as an example to verify the performance of the proposed method for large-area high-resolution DTM generation. In total, eight NAC image pairs (16 L/R images) that can cover the whole landing area are used to yield 1.5-m resolution DTM. The result is shown in Fig. 11. The eight image pairs are acquired under different illumination conditions, and they show unbalanced image grayscale and contrast [see Fig. 11(a)]. However, the reconstructed result is not influenced by these factors. From Section III-B, we know that our present approach is robust under different illumination conditions. Besides, our method is constrained by SLDEM. We design a CDEB module in the network to assimilate the elevation constraint in SLDEM to improve the DTM prediction accuracy. We also eliminate or mitigate elevation inconsistencies between different DTM tiles during the scale recovering stage. As a result, the final derived CE-4 landing area DTM is seamless [see Fig. 11(b)], which is mosaicked from tens of thousands of small DTM tiles. In addition, our generated DTM retrieves the terrain trends and features correctly. It can be noticed that the different-scale craters, i.e., the most common regular terrain features, in the landing area are obviously retrieved, as shown in Fig. 11(c). Furthermore, we see the irregular topographical features in the upper left and upper center corners of the landing area, which are marked by the pink ellipses in Fig. 11(c), are also successfully recovered.

Fig. 12(a) and (b) shows the zoom-in results of the pink squares A and B in Fig. 11(a), both of them passing through the seam of two adjacent NAC images [see Fig. 12(a1) and (b1)]. It can be noticed that the derived DTMs [see Figs. 12(a2) and (b2)] and the corresponding hill shaded maps [see Fig. 12(a3) and

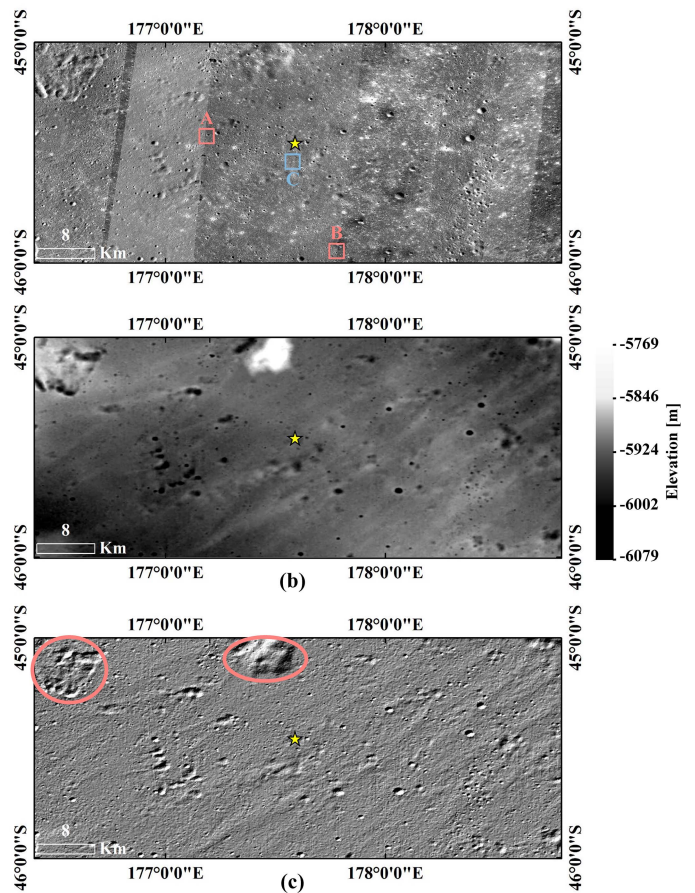


Fig. 11. Results of the CE-4 landing area. (a) Original NAC image pairs. (b) Generated 1.5-m resolution DTM. (c) Hill shaded map derived from (b) with 315° solar azimuth and 45° solar elevation. The yellow five-pointed star represents the CE-4 landing site. The pink ellipses mark the irregular topographical features. The zoom-in results of the pink squares (A and B) and blue square (C) are shown in Fig. 12.

(b3)] show no artifacts during the seams, and the terrain is continuous. Fig. 12(c) shows the zoom-in results of the blue square C in Fig. 11(a). In this area, we compare our generated DTM with a typical 5-m resolution photogrammetry DTM.⁵ From the derived hill shaded maps, we see that the photogrammetry DTM retrieves a clearer shape of some craters, but it cannot reflect small-scale craters. On the other hand, the result is affected by the matching performance [see Fig. 12(c2)]. In contrast, our generated DTM recovers more topographic details, such as the ubiquitous small-scale craters [see Fig. 12(c3)].

2) Performance Comparison With Other Large-Area DTMs:

Currently, the existing large-area DTMs covering the CE-4 landing area are mostly derived from altimetric measurements or stereo-photogrammetric models with coarse resolutions, e.g., the 20-m resolution model derived from CE-2 stereo observations (CE-2 DTM) [40], the 30-m resolution DTM interpolated from LOLA profiles (LOLA DTM) [41], 60-m resolution SLDEM [26], and 100-m resolution Global Lunar DTM (GLD100) [42] produced from cross-track stereo LROC wide

⁵Downloaded from https://wms.lroc.asu.edu/lroc/view_rdr_product/NAC_DTM_CHANGE4

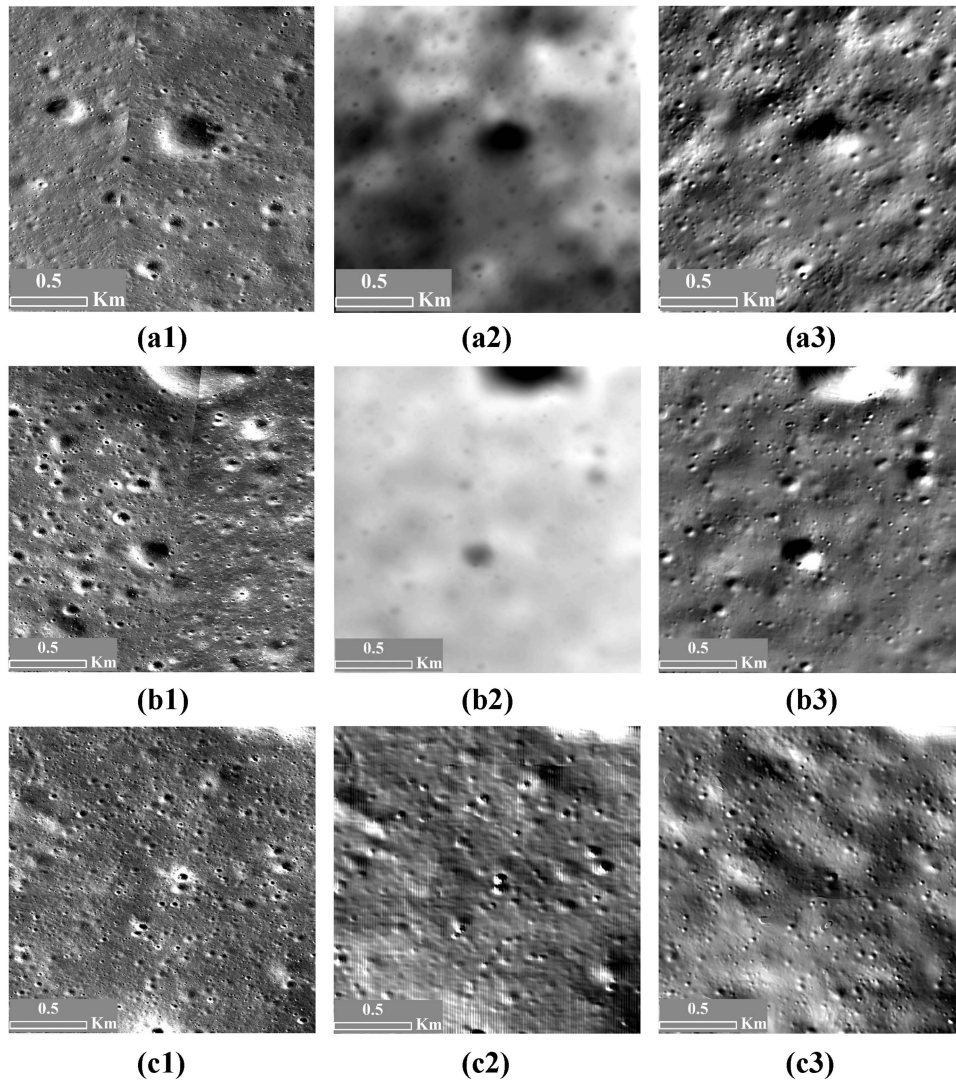


Fig. 12. Zoom-in views of the hill shaded maps in three typical local areas of CE-4 landing area. (a), (b) Areas within the pink squares A and B in Fig. 11(a), respectively. (c) Blue square C in Fig. 11(a). (a1), (b1), (c1) Original NAC images. (a2), (b2) Our derived DTMs. (c2) Hill shaded map derived from photogrammetry DTM. (a3), (b3), (c3) Hill shaded maps derived from our generated DTM.

angle camera imagery. We compare our 1.5-m resolution result with these DTMs to verify the performance of our presented approach. Fig. 13 shows these five DTMs and their derived hill shaded maps. It can be noticed that artifacts are present in the LOLA DTM [see Fig. 13(a)] see, most likely introduced by the interpolation of sparse profiles. The parallel but irregularly spaced tracks are outliers due to errors in the reconstructed orbits, spacecraft pointing, or time stamps. The hill shaded map derived from LOLA DTM shows that significant texture errors and the terrain features are smoothed out. As the low effective resolution of GLD100 model, ~ 1 km [42], only the large-scale hills in the upper center of the landing area are clearly retrieved by the stereo-photogrammetric processing [see Fig. 13(b)]. The model shows obvious incorrect high-frequency, bumpy texture, and a large number of multiscale craters are failed to be retrieved. Compared with LOLA DTM and GLD100 model with obvious errors, the CE-2 DTM [see Fig. 13(c)] and SLDEM [see Fig. 13(d)] display more topographical features

and correct terrain trends, which are consistent with the texture characteristics in the NAC images [see Fig. 11(a)]. However, there are significant horizontal positional deviations between the CE-2 model and SLDEM, as indicated by the position of the same crater marked by the white arrows in Fig. 14(d) and (e). By contrast, our 1.5-m resolution DTM can retrieve terrain features [see Fig. 13(e)], especially small-scale terrain features that cannot be captured in coarse-resolution DTMs. Besides, we use the SLDEM as reference model, and compute the root-mean-square error (RMSE) values between other DTMs (LOLA DTM, GLD100, CE-2 DTM, and our generated DTM) and SLDEM. We provide the RMSE values before and after the alignment of LOLA DTM, GLD100, and CE-2 DTM with respect to SLDEM for an elaborate comparison, which is presented in Table III. The GLD100 model before and after the alignment demonstrates the largest RMSE value, 21.398 and 10.128 m, respectively. The RMSE between CE-2 DTM and SLDEM is larger than that between LOLA DTM and SLDEM

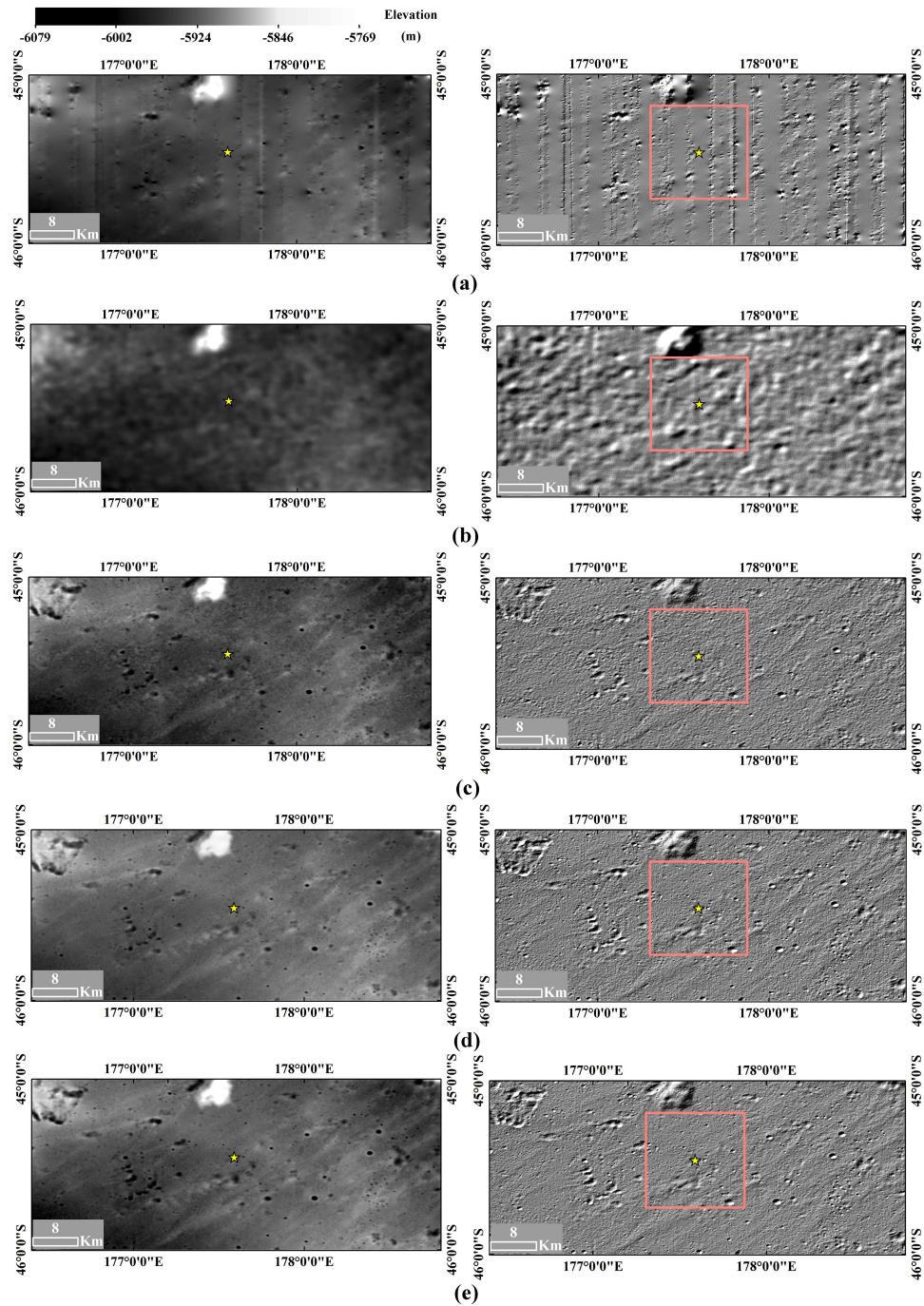


Fig. 13. Five different large-area DTMs covering the CE-4 landing area. Left-hand column presents the DTMs. Right-hand column shows the derived hill shaded maps with 315° solar azimuth and 45° solar elevation. (a) displays 30-m resolution LOLA DTM. (b) displays 100-m resolution GLD100 model. (c) displays 20-m resolution CE-2 DTM. (d) displays 60-m resolution SLDEM. (e) displays 1.5-m resolution our generated DTM.

TABLE III
RMSES OF LOLA DTM, GLD100, CE-2 DTM, AND OUR GENERATED DTM WITH RESPECT TO SLDEM

DTM	LOLA DTM	GLD100	CE-2 DTM	Our DTM
RMSE before alignment (m)	8.986	21.398	11.459	-
RMSE after alignment (m)	7.750	10.128	7.322	1.254

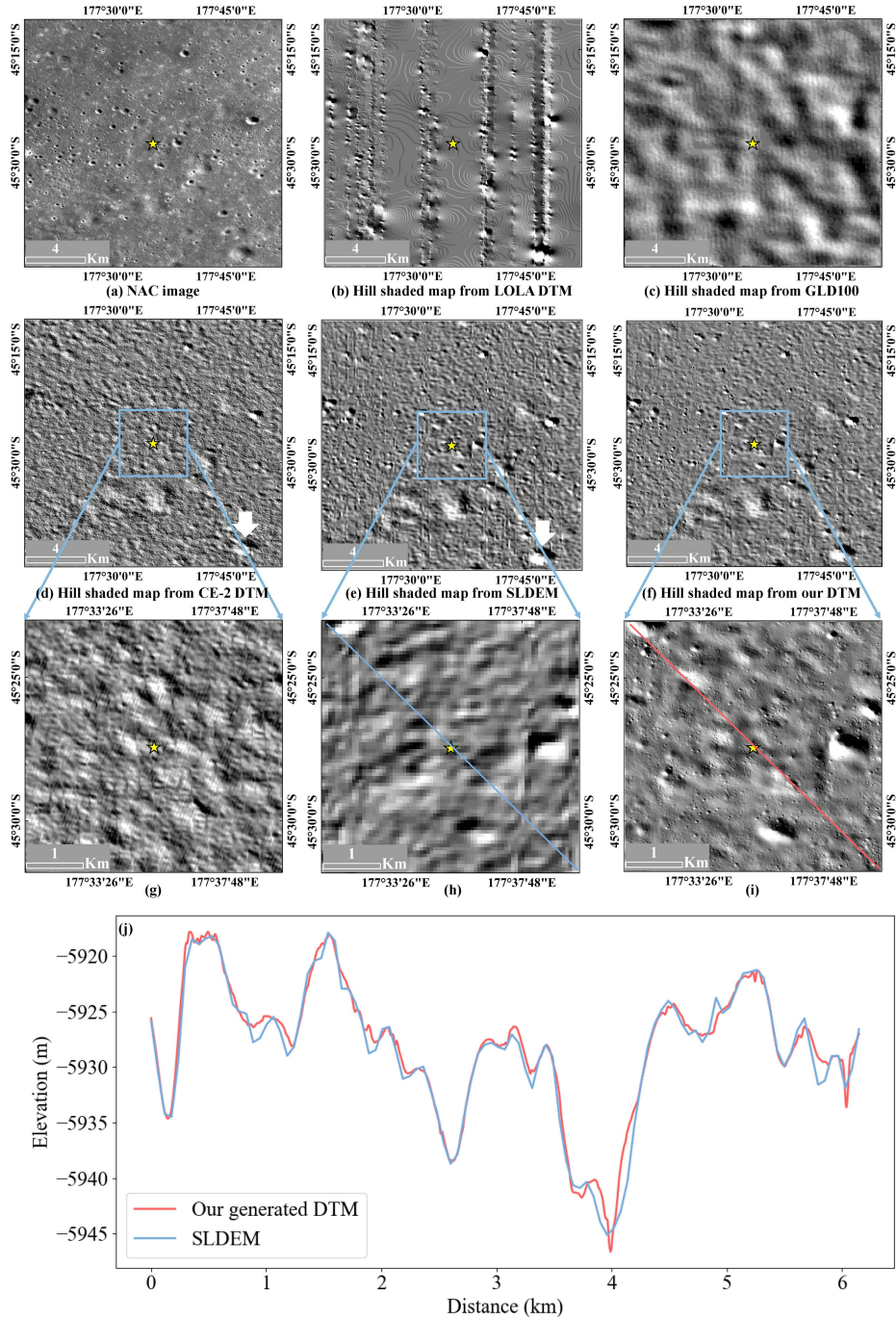


Fig. 14. Zoom-in views of hill shaded maps of the five DTMs in Fig. 13. (a) NAC image: M1303640934. (b)–(f) Hill shaded maps of LOLA DTM, GLD100, CE-2 DTM, SLDEM, and our generated DTM covering the pink boxes in Fig. 13, respectively. (g)–(i) Blue boxes area in (d)–(f). (j) Profile analysis results between our generated DTM and SLDEM. The pink and blue lines in (h) and (i) are the profiles.

before alignment. In contrast, as the CE-2 DTM recovers the correct topographic trends and more terrain features, the aligned RMSE value (7.322 m) is smaller than that of the LOLA DTM (7.750 m). Since our CNN method is constrained by the SLDEM, the RMSE value between our result and SLDEM is only 1.254 m, which is far less than other three DTMs.

Fig. 14 further displays zoom-in views of the hill shaded maps derived from the five DTMs in Fig. 13. For the sake of comparison, their solar azimuth and altitude are consistent with the NAC

image: M1303640934 [see Fig. 14(a)]. We see that the LOLA DTM and GLD100 models fail to retrieve the terrain features in the local area. Subject to the coarse resolution, the CE-2 model and SLDEM cannot show small-scale craters. In contrast, apart from large-scale terrain features, our 1.5-m resolution DTM can retrieve the small-scale, from meters to tens of meters, craters as well [see Fig. 14(i)]. To reduce the effects of the artefacts on the SLDEM, only the output with global information of the CDEB module is fed into the subsequent decoder module. In contrast,

to enable the model to predict more terrain features accurately, the multiscale outputs from the blocks in the HIEB module are fed into the counterparts in the decoder module, and two types of loss terms (l_{grad} and l_{norm}) considering the relationships among neighbor pixels are designed. From the profile analysis results [see Fig. 14(j)], we see that the elevation trends of our generated DTM are highly consistent with SLDEM. Nevertheless, our result is more continuous, even at the abrupt elevation changes in SLDEM. The elevation variations of many craters are able to be reflected in the profiles of our results. Overall, compared with other large-area DTMs, our high-resolution DTM not only shows considerably more details of the terrain, especially the small-scale features, but the vertical direction is also consistent with the SLDEM.

IV. CONCLUSION

In this article, we propose a deep learning-based reconstruction for large-area pixel-resolution DTM from multiple adjacent single-view NAC images and SLDEM. An improved CNN architecture is presented to predict DTMs with more detailed and accurate terrain features. We use the CE-3 and CE-4 landing areas to test the robustness and performance. The results show that our approach is robust under varying illumination conditions. Compared with a photogrammetric technique, it has great potential for large-area DTM generation from multiple single-view NAC images constrained by SLDEM. Our generated 1.5-m resolution DTM for the large CE-4 landing area features the best elevation consistency with respect to SLDEM in comparison to other existing DTMs. Due to the great increase in resolution, our results can show finer topographical details, especially small-scale craters, which are not reflected in the other existing large-area DTMs. In future, we will investigate specifically and quantitatively the effect of illumination geometry and the constraining DTM on model performance. We will also focus on the global-scale high-resolution DTM generation.

ACKNOWLEDGMENT

The authors would like to thank the LROC SOC for their continuing work in deriving and maintaining the data used in this paper.

REFERENCES

- [1] Y. Jia, Y. Zou, J. Ping, C. Xue, J. Yan, and Y. Ning, "The scientific objectives and payloads of Chang'E-4 mission," *Planet. Space Sci.*, vol. 162, pp. 207–215, Nov. 2018.
- [2] D. D. Rosa et al., "Characterisation of potential landing sites for the European space agency's lunar lander project," *Planet. Space Sci.*, vol. 74, no. 1, pp. 224–246, Dec. 2012.
- [3] B. Wu, H. Hu, and J. Guo, "Integration of Chang'E-2 imagery and LRO laser altimeter data with a combined block adjustment for precision lunar topographic modeling," *Earth Planet. Sci. Lett.*, vol. 391, pp. 1–15, Apr. 2014.
- [4] P. Gläser, J. Oberst, G. A. Neumann, E. Mazarico, E. J. Speyerer, and M. Robinson, "Illumination conditions at the lunar poles: Implications for future exploration," *Planet. Space Sci.*, vol. 162, pp. 170–178, Nov. 2018.
- [5] J. Lai et al., "First look by the Yutu-2 rover at the deep subsurface structure at the lunar farside," *Nature Commun.*, vol. 11, no. 1, pp. 1–9, Jul. 2020.
- [6] V. Lohse, C. Heipke, and R. L. Kirk, "Derivation of planetary topography using multi-image shape-from-shading," *Planet. Space Sci.*, vol. 54, no. 7, pp. 661–674, Jul. 2006.

- [7] J. Flahaut et al., "Regions of interest (ROI) for future exploration missions to the lunar south pole," *Planet. Space Sci.*, vol. 180, Jan. 2020, Art. no. 104750.
- [8] C. Zhou et al., "Scientific objectives and payloads of the lunar sample return mission—Chang'E-5," *Adv. Space Res.*, vol. 69, no. 1, pp. 823–836, Jan. 2022.
- [9] M. A. Ivanov et al., "Landing site selection for Luna-Glob mission in crater Boguslawsky," *Planet. Space Sci.*, vol. 117, pp. 45–63, Nov. 2015.
- [10] B. Wu et al., "Topographic modeling and analysis of the landing site of Chang'E-3 on the moon," *Earth Planet. Sci. Lett.*, vol. 405, pp. 257–273, Nov. 2014.
- [11] B. Wu et al., "Topographic and geomorphological mapping and analysis of the Chang'E-4 landing site on the far side of the moon," *Photogrammetric Eng. Remote Sens.*, vol. 86, no. 4, pp. 247–258, Apr. 2020.
- [12] K. Di et al., "High-resolution large-area digital orthophoto map generation using LROC NAC images," *Photogrammetric Eng. Remote Sens.*, vol. 85, no. 7, pp. 481–491, Jul. 2019.
- [13] M. Golombek et al., "Selection of the Mars science laboratory landing site," *Space Sci. Rev.*, vol. 170, no. 1, pp. 641–737, Jul. 2012.
- [14] X. Jiang, L. Shuang, and T. Ting, "Innovative hazard detection and avoidance strategy for autonomous safe planetary landing," *Acta Astronautica*, vol. 126, pp. 66–76, Sep. 2016.
- [15] M. R. Henriksen et al., "Extracting accurate and precise topography from LROC narrow angle camera stereo observations," *Icarus*, vol. 283, pp. 122–137, Feb. 2017.
- [16] M. S. Robinson et al., "Lunar reconnaissance orbiter camera (LROC) instrument overview," *Space Sci. Rev.*, vol. 150, no. 1, pp. 81–124, Mar. 2010.
- [17] K. Di, J. Oberst, I. Karachevtseva, and B. Wu, "Topographic mapping of the Moon in the 21st century: From hectometer to millimeter scales," *ISPRS Arch.*, vol. 43, no. B3, pp. 1117–1124, Aug. 2020.
- [18] I. Haase, P. Gläser, and J. Oberst, "Bundle adjustment of spaceborne double-camera push-broom imagers and its application to LROC NAC imagery," *ISPRS Arch.*, vol. 42, no. W13, pp. 1397–1404, Jun. 2019.
- [19] B. Wu, W. Liu, A. Grumpe, and C. Wöhler, "Construction of pixel-level resolution DEMs from monocular images by shape and albedo from shading constrained with low-resolution DEM," *ISPRS J. Photogrammetry Remote Sens.*, vol. 140, pp. 3–19, Jun. 2018.
- [20] W. Liu and B. Wu, "An integrated photogrammetric and photogrammetric approach for illumination-invariant pixel-resolution 3D mapping of the lunar surface," *ISPRS J. Photogrammetry Remote Sens.*, vol. 159, pp. 153–168, Jan. 2020.
- [21] Y. Tao et al., "Rapid single image-based DTM estimation from ExoMars TGO CaSSIS images using generative adversarial U-Nets," *Remote Sens.*, vol. 13, no. 15, p. 2877, Jul. 2021.
- [22] O. Alexandrov and R. A. Beyer, "Multiview shape-from-shading for planetary images," *Earth Space Sci.*, vol. 5, no. 10, pp. 652–666, Sep. 2018.
- [23] Y. Tao, J. P. Muller, S. J. Conway, and S. Xiong, "Large area high-resolution 3D mapping of Oxia Planum: The landing site for the ExoMars Rosalind Franklin rover," *Remote Sens.*, vol. 13, no. 16, p. 3270, Aug. 2021.
- [24] Z. Chen, B. Wu, and W. Liu, "Mars3DNet: CNN-based high-resolution 3D reconstruction of the Martian surface from single images," *Remote Sens.*, vol. 13, no. 5, Feb. 2021, Art. no. 839.
- [25] H. Chen, X. Hu, and J. Oberst, "Pixel-resolution DTM generation for the lunar surface based on a combined deep learning and shape-from-shading (SFS) approach," *ISPRS Arch.*, vol. V-3-2022, pp. 511–516, Mar. 2022.
- [26] M. K. Barker, E. Mazarico, G. A. Neumann, M. T. Zuber, and J. Haruyama, "A new lunar digital elevation model from the lunar orbiter laser altimeter and SELENE terrain camera," *Icarus*, vol. 273, pp. 346–355, Jul. 2016.
- [27] C. H. Acton Jr., "Ancillary data services of NASA's navigation and ancillary information facility," *Planet. Space Sci.*, vol. 44, no. 1, pp. 65–70, Jan. 1996.
- [28] X. Xin, B. Liu, K. Di, M. Jia, and J. Oberst, "High-precision co-registration of orbiter imagery and digital elevation model constrained by both geometric and photometric information," *ISPRS J. Photogrammetry Remote Sens.*, vol. 144, pp. 28–37, Oct. 2018.
- [29] B. K. P. Horn, "Hill shading and the reflectance map," *Proc. IEEE*, vol. 69, no. 1, pp. 14–47, Jan. 1981.
- [30] R. Beyer et al., "NeoGeographyToolkit/StereoPipeline 2.7.0 (2.7.0)," Zenodo, Geneva, Switzerland, 2020, doi: [10.5281/zenodo.3963341](https://doi.org/10.5281/zenodo.3963341).
- [31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

- [32] X. Cheng, P. Wang, and R. Yang, "Depth estimation via affinity learned with convolutional spatial propagation network," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 103–119.
- [33] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, "Deeper depth prediction with fully convolutional residual networks," in *Proc. Int. Conf. 3D Vis.*, 2016, pp. 239–248.
- [34] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
- [35] S. S. Shivakumar, T. Nguyen, I. D. Miller, S. W. Chen, V. Kumar, and C. J. Taylor, "DfuseNet: Deep fusion of rgb and sparse depth information for image guided dense depth completion," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2019, pp. 13–20.
- [36] J. Hu, M. Ozay, Y. Zhang, and T. Okatani, "Revisiting single image depth estimation: Toward higher resolution maps with accurate object boundaries," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2019, pp. 1043–1051.
- [37] M. T. Bland, R. L. Kirk, D. M. Galuszka, D. P. Mayer, and R. A. Beyer, and R. L. Fergason, "How well do we know Europa's topography? An evaluation of the variability in digital terrain models of Europa," *Remote Sens.*, vol. 13, no. 24, Dec. 2021, Art. no. 5097.
- [38] L. Xiao, "China's touch on the moon," *Nature Geosci.*, vol. 7, no. 6, pp. 391–392, Jun. 2014.
- [39] G. Facciolo, C. D. Franchis, and E. Meinhardt, "MGM: A significantly more global matching for stereovision," in *Proc. Brit. Mach. Vis. Conf.*, 2015, pp. 1–12.
- [40] C. Li et al., "Lunar global high-precision terrain reconstruction based on Chang'E-2 stereo images," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 43, no. 4, pp. 485–495, Apr. 2018.
- [41] D. E. Smith et al., "The lunar orbiter laser altimeter investigation on the lunar reconnaissance orbiter mission," *Space Sci. Rev.*, vol. 150, no. 1, pp. 209–241, 2010.
- [42] F. Scholten et al., "GLD100: The near-global lunar 100 m raster DTM from LROC WAC stereo image data," *J. Geophys. Res., Planets*, vol. 117, Mar. 2012, Art. no. E12.



Hao Chen was born in 1994. He received the M.Sc. degree in geographic information system from Tongji University, Shanghai, China, in 2020. He is currently working toward the Ph.D. degree in surveying and mapping engineering with the Technical University of Berlin, Berlin, Germany.

From July 2020 to January 2021, he was a Research Assistant with the College of Surveying and Geoinformatics, Tongji University. His research interests include spatial data processing, planetary mapping (mainly for lunar surface and asteroids), and deep

learning.

Xuanyu Hu, photograph and biography not available at the time of publication.

Philipp Gläser, photograph and biography not available at the time of publication.



Haifeng Xiao received the M.S. degree in surveying and mapping engineering from Tongji University, Shanghai, China, in 2017. He is currently working toward the Ph.D. degree in planetary geodesy with the Chair of Planetary Geodesy, Institute of Geodesy and Geoinformation Science, Technical University of Berlin, Berlin, Germany.

His research interests include planetary laser altimetry, Martian seasonal polar caps, body tidal deformation of Moon and Mercury, planetary rotation, interior structures, and evolution.



Zhen Ye received the B.E. and Ph.D. degrees in photogrammetry and remote sensing from Tongji University, Shanghai, China, in 2011 and 2018, respectively.

From 2018 to 2020, he was a Postdoctoral Researcher with the Chair of Photogrammetry and Remote Sensing, Technical University of Munich, Munich, Germany. He is currently an Associate Professor with the College of Surveying and Geoinformatics, Tongji University. His research interests include photogrammetry and remote sensing, high-accuracy image registration, and high-resolution satellite image processing.



Hanyue Zhang was born in 1994. She is currently working toward the Ph.D. degree in forestry with Beijing Forestry University, Beijing, China.

Her research interests include spatial data processing, forest management, and data mining.



Xiaohua Tong (Senior Member, IEEE) received the Ph.D. degree in geographic information system from Tongji University, Shanghai, China, in 1999.

Between 2001 and 2003, he was a Postdoctoral Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. He was a Research Fellow with Hong Kong Polytechnic University, Hong Kong, in 2006, and a Visiting Scholar with the University of California, Santa Barbara, CA, USA, between 2008 and 2009. He is currently a

Professor with the College of Surveying and Geoinformatics, Tongji University. His research interests include remote sensing, geographic information system, uncertainty and spatial data quality, and image processing for high-resolution and hyperspectral images.



Jürgen Oberst received the Ph.D. degree in planetary geodesy from the University of Texas at Austin, Austin, TX, USA, in 1989.

He is currently a Faculty Member and the Chair of Planetary Geodesy, Technical University Berlin, Institute of Geodesy and Geoinformation Sciences, Bonn, Germany. In his research, he works on various aspects of planetary geodesy and geophysics, using data from laser altimeters and cameras. His particular interest is on planet Mars. He has been a long-standing Co-Investigator in the team of the high-resolution

stereo camera on Mars Express.