DEEP LEARNING FOR 3D RECONSTRUCTION OF THE MARTIAN SURFACE USING MONOCULAR IMAGES: A FIRST GLANCE

Zeyu Chen, Bo Wu*, Wai Chung Liu

Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, - bo.wu@polyu.edu.hk, (ze-yu.chen, mo.liu)@connect.polyu.hk

Commission III, ICWG III/II

KEY WORDS: Mars, Surface Reconstruction, Convolutional Neural Networks, Monocular Images

ABSTRACT:

The paper presents our efforts on CNN-based 3D reconstruction of the Martian surface using monocular images. The Viking colorized global mosaic and Mar Express HRSC blended DEM are used as training data. An encoder-decoder network system is employed in the framework. The encoder section extracts features from the images, which includes convolution layers and reduction layers. The decoder section consists of deconvolution layers and is to integrate features and convert the images to desired DEMs. In addition, skip connection between encoder and decoder section is applied, which offers more low-level features for the decoder section to improve its performance. Monocular Context Camera (CTX) images are used to test and verify the performance of the proposed CNN-based approach. Experimental results show promising performances of the proposed approach. Features in images are well utilized, and topographical details in images are successfully recovered in the DEMs. In most cases, the geometric accuracies of the generated DEMs are comparable to those generated by the traditional technology of photogrammetry using stereo images. The preliminary results show that the proposed CNN-based approach has great potential for 3D reconstruction of the Martian surface.

1. INTRODUCTION

3D reconstruction of the planetary surface is important for planetary exploration missions and scientific research. 3D surface models, such as digital elevation models (DEMs), can be used in a variety of applications including landing site evaluation (De Rosa et al., 2012; Wu et al., 2014; 2020), geomorphological study, and geological analysis (Head et al., 2010). Commonly used methods for image-based 3D surface reconstruction include photogrammetry and shape-from-shading (SfS). Photogrammetry has been widely used in planetary mapping and surface reconstruction (Wu et al., 2014; Beyer et al., 2018), which requires stereo or multiple images with appropriate imaging geometry for 3D reconstruction. However, high-resolution stereo coverages of images on planetary surfaces are rare. SfS is a technique that can retrieve pixel-wise 3D information of the surface from a single image based on the photometric content of the image (Horn, 1990). SfS has been applied for lunar topographic mapping using orbital images and showed favorable performances (Grumpe et al., 2014; Wu et al., 2018; Liu et al., 2018; Liu and Wu, 2020). However, the use of the SfS technique on Mars is more challenging due to its thin atmosphere. The atmospheric attenuation modifies the measured radiance and adds ambiguities to SfS.

In recent years, convolutional neural networks (CNNs) based methods have been developed for 3D reconstruction. In these methods, the 3D information of the scene, and the corresponding images are utilized as training labels and samples, and multiple network architectures are developed. For example, Eigen et al. (2014) proposed a CNN architecture to predict depth maps from close-range images. They used a coarse prediction and a refined prediction to improve accuracy. Laina et al. (2016) proposed a fully convolutional network to estimate depth maps from images. They removed the fully connected layer and transposed the convolution layers by fast up-convolution blocks, which up-sampled features by combining multiple convoluted features. Cao et al. (2018) treated the regression problem as a classification and reduced the difficulty of solving regression tasks. Ma and Karaman (2018) not only used a single image to CNN to predict depth but also used a sparse depth corresponding to the image. Results with close-range images on Earth have shown promising performances of the CNN-based depth estimation.

3D reconstruction of the Martian surface from monocular images by using SfS or CNN techniques is very different from the situations on Earth or the Moon. For Mars images collected from the orbit, they are usually contaminated by various noises such as atmospheric and camera noises. SfS normally relies on physical models, in which each parameter has an explicit physical meaning. However, SfS for Martian surface reconstruction requires additional atmospheric parameters and can be computationally expensive. CNNs analyzes and considers these elements implicitly and process them together, which simplifies the physical modeling and speeds up the processing. Thus, CNN based method might be a feasible solution for 3D reconstruction of the Martian surface using monocular images.

This paper presents an endeavor for CNN based 3D reconstruction of the Martian surface. A CNN based approach that considers atmospheric influences and camera noises have been developed. The network contains two subnetworks. The first one removes noises, shadows, and albedo differences in the images. The second one focuses on predicting high-quality DEMs by concatenating results from the first subnetwork and sparse DEMs. Preliminary experimental analysis using images collected by the Context Camera (CTX) on Mars shows promising results of the proposed approach.

* Corresponding author. Email: bo.wu@polyu.edu.hk

2. CNN MODEL FOR MARTIAN SURFACE RECONSTRUCTION

2.1 Overview of the CNN Model

The CNN model for Martian surface reconstruction requires an image and an existing sparse DEM (either from photogrammetry or laser altimetry) corresponding to the image coverage as inputs. After processing it generates a DEM with the same resolution of the input image. The network eliminates noises and extracts topographic details automatically from the inputs.

2.1.1 Network Architecture

As shown in Figure 1, the framework of the proposed approach consists of two networks. The first network devotes to remove non-Lambertian reflecting components, and the second network aims to predict DEMs with the same resolution of the input image. The input of the first network is pre-processed Martian surface images, and its output is the estimated Lambertian reflectance of the surface. The second network concatenates the sparse DEMs and the output of the first network as the input and generates high-resolution DEMs. Both the two subnetworks have the same architecture, i.e., the encoder-decoder networks. The encoder extracts features from the inputs, and then the decoder converts images to other types of images by utilizing these features. The Inception-Resnet-V2 (Szegedy et al., 2017) is employed as the encoder part, and the architecture of the decoder part is composed of five deconvolution layers. Concatenation paths are added between the encoder and decoder section to integrate both low-level and high-level features (Ronneberger et al., 2015). The advantage of concatenation connections is to incorporate low-level information from previous blocks and high-level information passed to the deconvolution blocks to generate better results. The original network of Szegedy et al. (2017) chooses ReLU as the activation function, while this proposed approach selects the simplified SReLU function (Jin et al., 2015). To keep the zero-centered property, hyparameters are replaced by constants. The function is defined by the formula below:

$$\mathrm{SReLU}(x_i) = \begin{cases} 3 + 0.8(x_i\text{-}3), & x_i \geq 3 \\ x_i, & 3 > x_i > -3 \\ -3 + 0.8(x_i + 3), & x_i \leq -3 \end{cases} \quad (1)$$

where X_i = features of the i^{th} layer.

Also, in the original network, padding methods include “SAME” padding and “VALID” padding. The “VALID” padding is removed because of the added concatenation connections. The prerequisite to add concatenation connections is to keep the same sizes of features among encoder and decoder blocks.

After the last layer of the network, a scale recovery unit is added to recover the original elevations coordinated of predictions. Linear regression is applied to find the best scale and bias to fit for the coordinates of ground truth. The formula is defined below:

$$G = \alpha P + \beta \quad (2)$$

where G = ground truth
P = prediction
α = scale
β = bias

2.1.2 Loss Function

The standard loss function for the regression problem is $\mathcal{L}_2$ loss, which is also called the mean-squared-error. However, in our experimental analysis, we found that the Huber loss has better performance than the $\mathcal{L}_2$ loss. The Huber loss is defined by the formula below:

$$\mathrm{Huber}(\hat{y}, y) \begin{cases} \frac{1}{2}(\hat{y}\text{-}y)^2 & \text{for} |\hat{y}\text{-}y| \leq \delta \\ \delta|\hat{y}\text{-}y| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases} . \quad (3)$$

where δ is set to 0.1 in this research. When the loss is larger than 0.1, the Huber loss equals to $\mathcal{L}_1$ loss, which is also called the mean-absolute-difference. If the loss is less than 0.1, the Huber loss equals to $\mathcal{L}_2$ loss, which penalizes large differences more than smaller differences and has better convergence property than the $\mathcal{L}_1$ loss.

In our network, a regularization term is used to decrease the overfitting phenomenon. It shrinks large numbers of parameters to constrain their ability of expression and decreases weights of final results. If the results are determined by many parameters rather than a small number of them, the probability of the happening of overfitting decreases. The $\mathcal{L}_2$ regularization is defined by the formula below:

$$L_{reg} = \lambda \sum_i x_i^2 \quad (4)$$

where L_{reg} = regularization loss
λ = regularization term
x_i = the i^{th} parameter

Eigen and Fergus (2015) proposed a first-order matching term in their loss function to encourage the network not only to minimize pixel differences, but also similar local structures. Assuming the ground truth and the prediction are $\hat{y}$ and y , and $d = \hat{y}\text{-}y$, the first-order term loss is defined below:

$$L_{fo}(\hat{y}, y) = \frac{1}{n}\sum_i \left[\left(\nabla_p d_i\right)^2 + \left(\nabla_q d_i\right)^2 \right] \quad (5)$$

where L_{fo} = the first order form loss
$\nabla_p d_i = \partial d / \partial x$
$\nabla_q d_i = \partial d / \partial y$

2.2 Training of the CNN Model

The training dataset is from the Viking colorized global mosaic and Mar Express HRSC blended DEM. The illumination direction of all the data is uniformed with an azimuth of 270° and the elevation angle is kept as original. Paired training data are uniformed to the same spatial resolution of 400 m/pixel. The images are clipped into 224*224 sub-images and are regularized to follow the standard normal distribution. Sparse DEMs are built by sampling clipped DEMs. Each pixel of the sparse DEMs is chosen with an interval of 33 pixels along with columns and rows from clipped DEMs. Pixels without a value are set as zero.

The simulated images with the Lambertian reflectance model are generated by inverting the DEMs. The formula of the Lambertian reflectance model is listed below:

$$L = I \cdot \cos\alpha \quad (6)$$

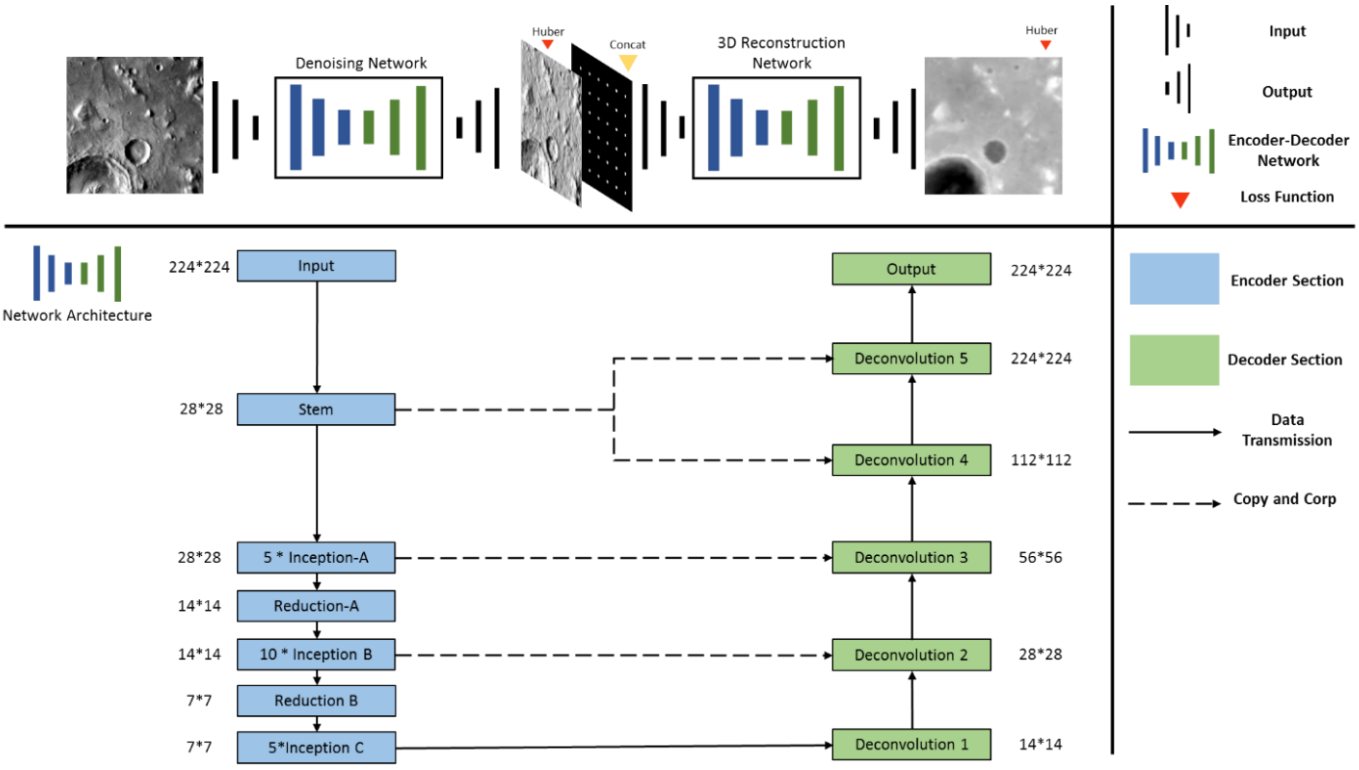

Figure 1. The network architecture of the proposed approach. Each coloured box represents a block in the network. Numbers beside blocks represents the output dimensions of the block.

where L = simulated image
I = illumination intensity
α = the angle between the illumination direction and the surface normal

The illumination direction is assigned with the azimuth of 270° and the elevation angle of 45°. The surface normal is calculated by the formula below:

$$S_n = g(x) \otimes g(y) \quad (7)$$

where S_n = surface normal
$g(x)$ = partial derivative along the horizontal direction
$g(y)$ = partial derivative along the vertical direction

The network is implemented on TensorFlow 1.14.0 platform and trained on two RTX 2080Ti graphic cards with 24GB of GPU memory (Abadi et al., 2016). Parameters are initialized by Xavier initialization (Glorot and Bengio, 2010). The Adam optimizer is used for optimization. The learning rate is 8e-5; Beta1 is 0.9; Beta2 is 0.999; Epsilon is1e-8. The batch size is assigned by 4, which balances the iteration efficiency and the network performance. The regularization term is set to be 0.0001.

3. EXPERIMENTAL EVALUATION USING CTX IMAGES

The performance of the trained network has been tested using monocular CTX images on Mars covering typical terrain types. Detailed information about the used CTX images is shown in Table 1. The Mars Orbiter Laser Altimeter (MOLA) DEM is used as the input sparse DEM. In order to provide a quantitative evaluation, reference DEMs are obtained from the USGS Mars dataset and used as ground truth for comparison. They were generated using photogrammetric techniques. Since those DEMs only have a spatial resolution of 20 m/pixel (about three times of the resolution of the CTX images), the input CTX images are down-sampled to 20 m for the subsequent comparison purpose. Other pre-processing procedures are the same as the training data.

Patches from the four CTX images are chosen to show the results (Figure 2). Most of them include at least one crater in the images. According to the profile comparison, shapes of these craters are well recovered. Main differences exist at the edges of craters, where we speculate that the ground-truth DEM often contains sharper and higher edges than the results from our CNN approach. Edges without apparent image contrast are less well-recovered, which is a common problem in 3D reconstruction methods using monocular images (Wu et al., 2018).

Shadows and spatial varying albedos do not affect the results obviously. Four CTX images were illuminated under different angles, and the surface albedo varies across different sections of images. Results reveal that such differences do not affect results significantly. The main reason is likely that the sparse MOLA DEMs helped the network to recognize and eliminate these errors, and the interpolation operation only happened around sparse DEM pixels.

As can be noticed in Figure 2, the DEMs generated by our CNN model contain checkboard artefacts. It is produced during the procedure of upsampling. When up-convolution operates, each time images are doubled, and such operation introduces and accumulates subtle noises on images. The problem can be solved by well-designed transposed convolution.

From Figure2, a few craters with a diameter of less than 10 pixels have not been successfully reconstructed, and small features that

have large contrast with the background are depicted better than in the reference DEM. Because of the selective responses, predictions generally are smoother than the reference DEM, which is a good character for images with severe noises. Table 2 presents the RMSE of the differences between the DEMs generated by our CNN approach and the reference DEMs. RMSE values of 3.7 m - 13.5 m shows the promising potential of our proposed approach.

Table 1. Information of the CTX images used for validation

Number	Image ID	Incidence Angle	Resampled Resolution	Center Latitude	Center Lontitude	Description
1	B17_016219_1978_XN_17N282W	42.28°	20 m/pixel	17.87°	77.29°	Candidate Mars 2020 Landing Site Northeast Syrtis Center
2	F21_043841_1654_XN_14S184W	59.27°	20 m/pixel	-14.72°	175.43°	Candidate Mars 2020 Landing Site McLaughlin Center
3	J03_045994_1986_XN_18N282W	47.59°	20 m/pixel	18.63°	77.44°	Candidate Mars 2020 Landing Site Jezero West
4	P17_007556_2012_XI_21N285W	42.18°	20 m/pixel	21.41°	74.45°	Candidate Mars 2020 Landing Site Nili Fossae Center

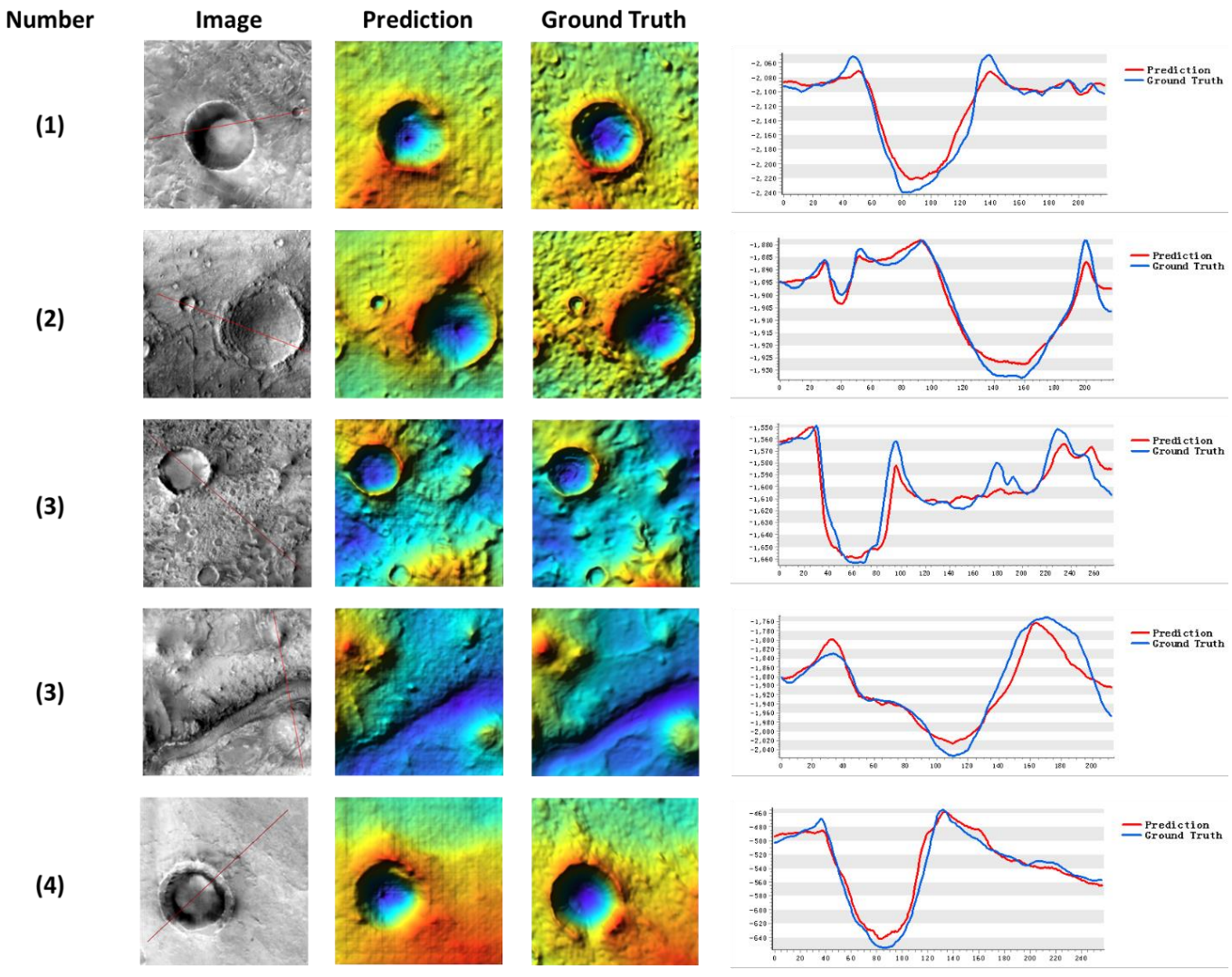

Figure 2. 3D reconstruction results and comparison. Numbers indicate patches selected from the CTX images listed in Table 1. Locations of profiles are depicted on images by red lines.

Table 2. Quantitative comparison between the DEMs generated by our CNN approach and the reference DEMs

Image ID	*RMSE (m)*
B17_016219_1978_XN_17N282W	13.5
F21_043841_1654_XN_14S184W	3.7
J03_045994_1986_XN_18N282W	12.5
P17_007556_2012_XI_21N285W	11.9

4. CONCLUSION

In this paper, we present an endeavor for 3D reconstruction of the Martian surface based on CNN using monocular images. The CNN model is trained by clipped samples from the Viking colorized global mosaic and Mar Express HRSC blended DEM. Image patches are input into the first subnetwork to generate noiseless results, and then sampled DEMs are concatenated with predictions of the first subnetwork as the input to the second subnetwork. Experimental analysis using typical CTX images show promising results, as compared with reference DEMs generated using the conventional photogrammetric technology. In addition, some favorable aspects of the proposed approach, such as feature-based interpolation and strong denoise ability are represented.

The work presented in this paper shows the great potential for CNN based method for 3D reconstruction of the Martian surface from monocular images, which is of significance for Martian topographic mapping and scientific research.

ACKNOWLEDGEMENTS

The work described in this paper was funded a grant from the Research Grants Council of Hong Kong (Research Impact Fund – Project No: R5043-19) and a grant from the National Natural Science Foundation of China (Project No: 41671426). The authors also would like to thank all those who worked on the archive of the datasets to make them publicly available.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: A System for Large-Scale Machine Learning. Presented at the 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), pp. 265–283.

Beyer, R.A., Alexandrov, O., McMichael, S., 2017. The Ames Stereo Pipeline: NASA's Open Source Software for Deriving and Processing Terrain Data. Journal of Geophysical Research: Planets 537–548.

Cao, Y., Wu, Z., Shen, C., 2018. Estimating Depth From Monocular Images as Classification Using Deep Fully Convolutional Residual Networks. IEEE Transactions on Circuits and Systems for Video Technology 28, 3174–3182.

De Rosa, D., Bussey, B., Cahill, J.T., Lutz, T., Crawford, I.A., Hackwill, T., van Gasselt, S., Neukum, G., Witte, L., McGovern, A., 2012. Characterisation of potential landing sites for the European Space Agency's Lunar Lander project. Planetary and Space Science 74, 224-246.

Eigen, D., Fergus, R., 2015. Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-scale Convolutional Architecture, in: 2015 IEEE International Conference on Computer Vision (ICCV). Presented at the 2015 IEEE International Conference on Computer Vision (ICCV), IEEE,

Eigen, D., Puhrsch, C., Fergus, R., 2014. Depth Map Prediction from a Single Image using a Multi-Scale Deep Network, in: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 27. Curran Associates, Inc., pp. 2366–2374.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Presented at the Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256.

Grumpe, A., Belkhir, F., Wöhler, C., 2014. Construction of lunar DEMs based on reflectance modelling. Advances in Space Research 53, 1735–1767.

Head, J.W., Fassett, C.I., Kadish, S.J., Smith, D.E., Zuber, M.T., Neumann, G.A., Mazarico, E., 2010. Global distribution of large lunar craters: Implications for resurfacing and impactor populations. Science 329, 1504-1507.

Horn, B.K.P., 1990. Height and gradient from shading. International Journal of Computer Vision 5(1), 37-75.

Jin, X., Xu, C., Feng, J., Wei, Y., Xiong, J., Yan, S., 2015. Deep Learning with S-shaped Rectified Linear Activation Units. arXiv:1512.07030 [cs].

Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N., 2016. Deeper Depth Prediction with Fully Convolutional Residual Networks, in: 2016 Fourth International Conference on 3D Vision (3DV). Presented at the 2016 Fourth International Conference on 3D Vision (3DV), pp. 239–248.

Liu, W.C., Wu, B., Wöhler, C., 2018. Effects of Illumination Differences on Photometric Stereo Shape-and-Albedo-from-Shading for Precision Lunar Surface Reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing 136, 58-72.

Liu, W.C., Wu, B., 2020. An integrated photogrammetric and photoclinometric approach for illumination-invariant pixel-resolution 3D mapping of the lunar surface. ISPRS Journal of Photogrammetry and Remote Sensing 159, 153-168.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 234–241.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence 4278–4284.

Wu, B., Li, F., Ye, L., Qiao, S., Huang, J., Wu, X., Zhang, H., 2014. Topographic Modeling and Analysis of the Landing Site of Chang'E-3 on the Moon, Earth and Planetary Science Letters, 405, pp. 257-273.

Wu, B., Hu, H., Guo, J., 2014. Integration of Chang'E-2 Imagery and LRO Laser Altimeter Data with a Combined Block Adjustment for Precision Lunar Topographic Modeling. Earth and Planetary Science Letters 391, 1–15.

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 Journal of Photogrammetry and Remote Sensing 140, 3–19.

Wu, B., Li, F., Hu, H., Zhao, Y., Wang, Y., Xiao, P., Li, Y., Liu, W. C., Chen, L., Ge, X., Yang, M., Xu, Y., Ye, Q., Wu, X., Zhang, H., 2020. Topographic and Geomorphological Mapping and Analysis of the Chang'E-4 Landing Site on the Far Side of the Moon. Photogrammetric Engineering & Remote Sensing 86(4), 247-258.