Article:
Dong, Guoshuai, Huang, Weimin, Smith, William Alfred Peter orcid.org/0000-0002-6047-0413 et al. (1 more author) (2020) Filling Voids in Elevation Models Using a Shadow-Constrained Convolutional Neural Network. IEEE Geoscience and Remote Sensing Letters. pp. 592-596. ISSN 1545-598X

https://doi.org/10.1109/LGRS.2019.2926530

Filling Voids in Elevation Models using a Shadow Constrained Convolutional Neural Network

Guoshuai Dong, Weimin Huang, *Senior Member*, *IEEE*, William A. P. Smith, *Senior Member*, *IEEE*, and Peng Ren, *Senior Member*, *IEEE*

***Abstract*—We explore the use of convolutional neural networks (CNNs) for filling voids in digital elevation models (DEM). We propose a baseline approach using a fully convolutional network to predict complete from incomplete DEMs which is trained in a supervised fashion. We then extend this to a shadow constrained CNN (SCCNN) by introducing additional loss functions that encourage the restored DEM to adhere to geometric constraints implied by cast shadows. At training time, we use automatically extracted cast shadow maps and known sun directions to compute the shadow-based supervisory signal in addition to the direct DEM supervision. At test time, our network directly predicts restored DEMs from an incomplete DEM. One key advantage of our SCCNN model is that it is characterized by both CNN data inference and geometric shadow cues. It thus avoids the data restoration which may violate shadowing conditions. Both our baseline CNN and SCCNN outperform the inverse distance weighting (IWD) based interpolation method, with the shadow supervision enabling SCCNN to obtain the best performance.**

***Index Terms*—Convolutional neural network, shadow geometry constraint, shadow map**

I. Introduction

In 2000, the National Aeronautics and Space Administration (NASA) collected radar data covering more than 80% of the global land surface through the Shuttle Radar Topography Mission (SRTM) [1]. The SRTM data was used to build digital elevation models (DEMs). DEMs play an important role in various fields such as geological mapping [2] and natural disaster monitoring [3]. Therefore, there is a high requirement on the accuracy and completeness of DEM data. However, there are a large number of voids (areas with unknown elevation) in the SRTM data, especially in mountainous areas. This is because it is difficult for radar to image steep terrain. These void regions account for 0.3% of the total surveyed area [4], [5] but are concentrated in mountainous regions. It is therefore important to develop void-filling strategies for these areas. Most existing void-filling schemes are based on interpolation. Reuter *et al.* [6] introduced terrain restoration methods including the filling and feathering approach, the IWD based interpolation method, etc. However, these methods require the information from auxiliary DEM sources to improve the accuracy of restored results. Milan *et al.* [7] used an auxiliary DEM to fill mountainous missing data. However, the limitation of this method is that it cannot be extended to the DEM data with different resolutions and auxiliary DEM data are not always available. Hogan *et al.* [8] improved interpolation results according to geometric constraints provided by shadows. However, shadow geometric constraints obtained from shadows provide relatively sparse information and the method is accompanied by a nonconvex optimization problem which is difficult to compute and may fall into a local optimum. Ling *et al.* [9] employed satellite images to obtain the topographic information of valleys for interpolation. However, it is almost not scalable to other mountainous situations except valleys for the method.

Most existing interpolation based methods do not consider the global information contained in the non-void regions nor are they capable of correcting errors outside of the void regions. These limitations can possibly be addressed by exploiting deep learning models, e.g. convolutional neural networks (CNNs) [10], which are capable of learning the space of plausible DEMs. Hence, for our baseline method we train a void filling CNN in a supervised fashion using complete DEMs as the target output. However, such an approach is dependent on the quantity and quality of complete DEMs that can be provided. The training samples themselves may contain errors or may be produced by a separate void filling procedure and so may lead to restorations that violate geometric constraints. One such geometric constraint arises from observing cast shadows in terrain imagery. We make a preliminary attempt to introduce additional cast shadow supervision. This shadow constrained convolutional neural network (SCCNN) encourages the restored DEMs to adhere to shadow geometry constraints, potentially leading to improvements even in the non-void regions. Compared with existing interpolation based methods, our SCCNN model learns a powerful representation of DEMs and can thus comprehensively characterize the relationship between missing data and valid data. It does not need auxiliary DEMs and avoids the problem of ambiguous resolutions from different DEMs. Furthermore, in contrast to our baseline CNN, our SCCNN encourages the restored data to satisfy shadowing rules and thus further improves accuracy and robustness. The deep nets presented in [11] are among the first for deep learning based DEM restoration. Specifically, they consist of multiple neural nets, i.e., generators and discriminators. In contrast, our method only exploits one neural net and is structurally less sophisticated. Furthermore,

This work was supported in part by National Key R&D Program of China (2017YFC1405600) and the Fundamental Research Funds for the Central Universities under Project 18CX05014A. *(Corresponding author: Peng Ren).*

G. Dong and P. Ren are with the College of Information and Control Engineering, China University of Petroleum (East China), Qingdao 266580, China (e-mail: dongguoshuaiupc@163.com; pengren@upc.edu.cn).

W. Huang is with Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL, Canada. (e-mail: weimin@mun.ca).

W. A. P. Smith is with Department of Computer Science, University of York, York, YO10 5GH, UK. (e-mail: william.smith@york.ac.uk).

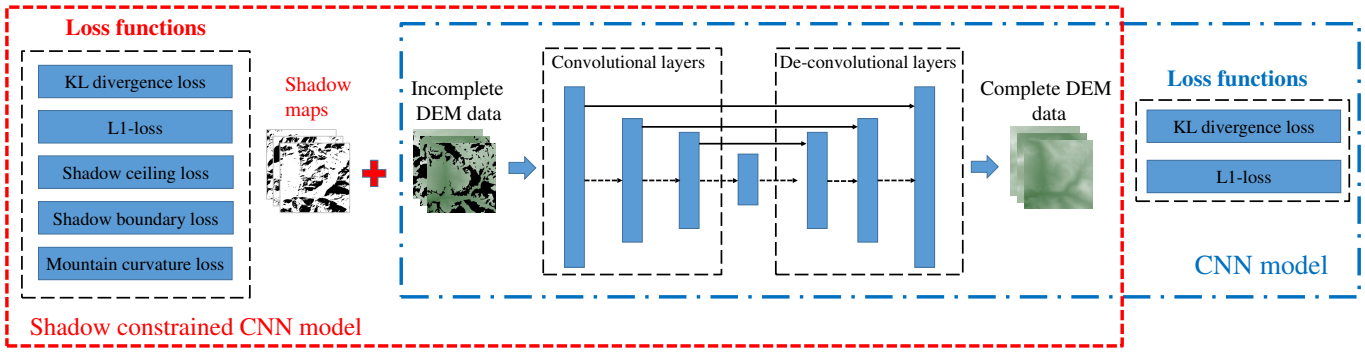

Fig. 1: DEM void filling with a CNN (blue rectangle) and shadow constrained CNN (SCCNN) red rectangle.

our method characterizes the shadow cues which are not considered in [11].

II. Void filling with a CNN

We propose to train a fully convolutional network (i.e., image-to-image with no fully connected layers) that takes an incomplete DEM as input (with elevation in void regions set to zero) and outputs a complete DEM of the same resolution. We use the U-Net architecture [12] that comprises a contractive convolutional encoder and a deconvolutional decoder with skip connections between encoder and decoder layers of the same spatial resolution. These skip connections are crucial for transferring high frequency detail from input to output. The U-Net architecture has proven very powerful on a wide range of image-to-image tasks.

At a DEM location $x \in X$, where $X \subset \mathbb{R}^2$ is the set of pixel locations in the DEM, we denote the altitude predicted by the CNN as $\hat{H}(x)$ and the corresponding altitude in the complete training DEM by $H(x)$. We train the CNN using two widely used loss functions. First the ℓ_1 norm:

$$L_{\ell_1} = \sum_x |\hat{H}(x) - H(x)|, \tag{1}$$

and second the Kullback-Liebler (KL) divergence:

$$L_{KL} = -\sum_x H(x) \log \frac{\hat{H}(x)}{H(x)}. \tag{2}$$

See Fig. 1 (blue) for an illustration of this approach.

Although this CNN model is straightforward, it has a significant advantage over purely interpolation-based approaches. By training on large datasets, it is able to learn general characteristics of elevation data and fill voids in a way that is consistent with data it has previously seen. We show this in our evaluation. However, convolution and pooling layers in the CNN are local operations and so it cannot learn (spatially) long range dependencies. Moreover, the completed DEM may not be consistent with other cues. In particular, we now show how to exploit geometric constraints provided by cast shadows.

III. Shadow Geometry

Any region with large altitude variations (i.e., mountainous areas) contains locations where the sun is occluded when not directly overhead. These *cast shadow* regions provide informative geometric cues that can aid DEM void filling. The basic geometry of shadowing is illustrated in Fig. 2.

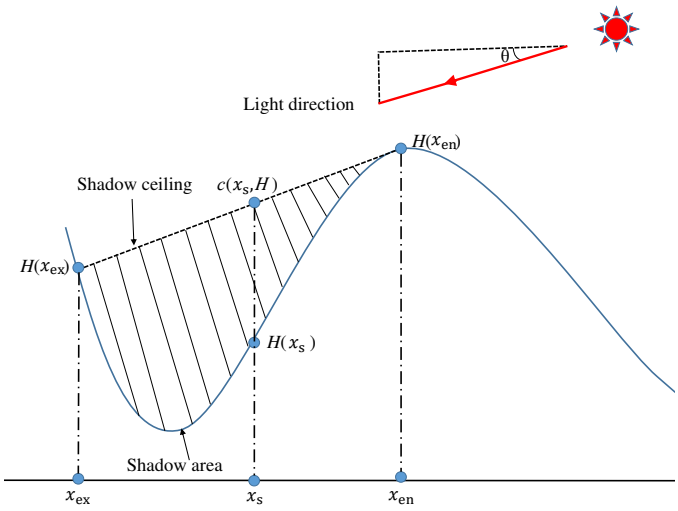

Fig. 2: Shadow geometry in a 2D slice through a DEM parallel to the light direction.

We treat the sun as a point source and denote by θ the angle between the light direction and the ground plane, i.e., $\theta = \arccos(s_3)$ where $\mathbf{s} \in \mathbb{R}^3$ is the unit length sun direction. Consider a 2D slice through the DEM that is parallel to both the light direction and the up vector (as in Fig. 2). We define pairs of locations $(x_{\text{en}}, x_{\text{ex}}) \in X_{\text{sbound}} \subset X \times X$ as the shadow entrance and exit points respectively with X_{sbound} containing all pairs of shadow boundary locations. The *shadow ceiling* is the line connecting the shadow entrance and exit points. It is clear that the shadow ceiling and light direction are parallel, which satisfies:

$$\frac{H(x_{\text{en}}) - H(x_{\text{ex}})}{||x_{\text{en}} - x_{\text{ex}}||} = \tan(\theta). \tag{3}$$

The region below the shadow ceiling lies in cast shadow and we denote this shadow area as the set of locations X_{s}. For a point in the shadow area $x_{\text{s}} \in X_{\text{s}}$, the elevation of the shadow ceiling at this location, denoted $c(x_{\text{s}}, H)$, is given by:

$$c(x_{\text{s}}, H) = \frac{H(x_{\text{ex}})\|x_{\text{s}} - x_{\text{en}}\| + H(x_{\text{en}})\|x_{\text{s}} - x_{\text{ex}}\|}{\|x_{\text{s}} - x_{\text{en}}\| + \|x_{\text{s}} - x_{\text{ex}}\|}. \tag{4}$$

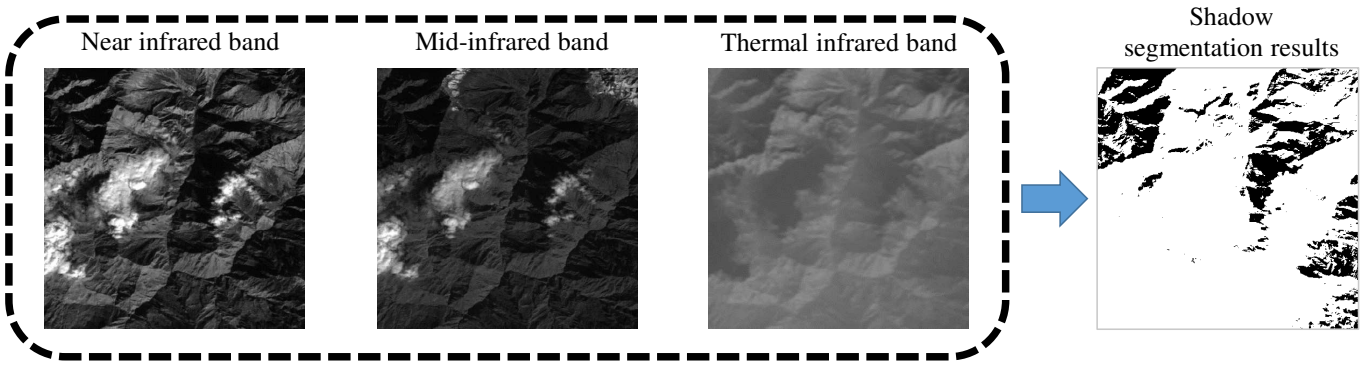

Fig. 3: Shadow segmentation.

For any location x_{s} within a shadow area, the altitude at that point, $H(x_{\mathrm{s}})$, must be lower than the ceiling elevation $c(x_{\mathrm{s}}, H)$. This observation provides the first shadow constraint (C1): $\forall x_{\mathrm{s}} \in X_{\mathrm{s}}, H(x_{\mathrm{s}}) < c(x_{\mathrm{s}}, H)$.

In addition, the terrain must be convex along the light source direction at a shadow entrance point. Specifically, the second directional derivative of H along direction $\bar{s} = \mathbf{Ps}/\|\mathbf{Ps}\|$ must be negative where $\mathbf{P} = [\mathbf{e}_x^T\, \mathbf{e}_y^T]^T \in \mathbb{R}^{2\times 3}$ and $\mathbf{Ps}$ is the orthogonal projection of $\mathbf{s}$ onto the ground plane. Using a finite difference approximation of the second derivative we obtain:

$$H''_{\bar{s}}(x) \approx H(x+\bar{s}) + H(x-\bar{s}) - 2H(\bar{s}). \tag{5}$$

The convexity constraint results in the second shadow constraint (C2): $\forall (x_{\mathrm{en}}, x_{\mathrm{ex}}) \in X_{\mathrm{sbound}}, H''_{\bar{s}}(x_{\mathrm{en}}) < 0$.

IV. Cast Shadow Supervision

We now show how to reframe the shadow constraints from the previous section as differentiable loss functions for use within a machine learning scheme.

A. Shadow Segmentation

We begin by explaining how we automatically detect shadow areas in terrain imagery. We employ the multi-band thresholding technique [8] to perform shadow segmentation. Specifically, we use three bands (near infrared, mid-infrared and thermal infrared bands) of multispectral satellite images from Landsat-5. Let $I_k(x)$ denote the normalized pixel intensity at x in the k-th band. The segmentation indicator $F(x)$ is formulated as:

$$F(x) = \prod_{k=1}^{K} (1 - I_k(x))^{\sigma_k}, \tag{6}$$

where σ_k is an empirical parameter. A shadow threshold η is applied to the segmentation indicator $F(x)$, resulting in the shadow region set X_{s}:

$$x = \begin{cases} \in X_{\mathrm{s}}, & \text{if } \ F(x) > \eta; \\ \notin X_{\mathrm{s}}, & \text{otherwise.} \end{cases} \tag{7}$$

Computing this for every pixel leads to a binary shadow map of the same size as the image, as shown in Fig. 3. The segmentation error is less than 10% which is empirically validated to be acceptable for shadow characterization [8].

B. Cast shadow loss functions

We now represent the shadow geometric constraints in the form of loss functions. Specifically, we design functions that take on a large value when a shadow constraint is violated and are otherwise zero. Hence, the first shadow constraint (C1) leads to a big loss in the case that a restored altitude in a shadow area is higher than the corresponding shadow ceiling elevation. We use an indicator function to characterize (C1) as follows:

$$\varepsilon[\hat{H}(x_{\mathrm{s}}) - c(x_{\mathrm{s}}, \hat{H})] = \begin{cases} 1, & \text{if } \ \hat{H}(x_{\mathrm{s}}) > c(x_{\mathrm{s}}, \hat{H}); \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

We use the indicator function (8) to enhance the disagreement penalty between the restored DEM and the true DEM, and obtain the shadow ceiling loss function L_{c} with respect to (C1) as follows:

$$L_{\mathrm{c}} = \sum_{x_{\mathrm{s}} \in X_{\mathrm{s}}} |\hat{H}(x_{\mathrm{s}}) - H(x_{\mathrm{s}})| \cdot \varepsilon[\hat{H}(x_{\mathrm{s}}) - c(x_{\mathrm{s}}, \hat{H})]. \tag{9}$$

According to the second shadow constraint (C2), we impose a large loss if the restored shadow entrance point is located at a valley rather than a peak. Following (5), we define a convexity characterization function $t(\hat{H}, H)$ as:

$$t(\hat{H}, H) = \frac{\hat{H}(x_{\mathrm{en}}+1) + \hat{H}(x_{\mathrm{en}}-1)}{2} - H(x_{\mathrm{en}}), \tag{10}$$

and another indicator function as follows:

$$\varepsilon(t(\hat{H}, H)) = \begin{cases} 1, & \text{if } \ t(\hat{H}, H) > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

This results in a value of 1 if the restoration violates (C2) and 0 otherwise. We use this in the shadow entrance curvature loss function L_{v} with respect to (C2) as follows:

$$\begin{aligned} L_{\mathrm{v}} = \sum_{x_{\mathrm{en}} \in X_{\mathrm{en}}} \varepsilon(t(\hat{H})) \cdot [|\hat{H}(x_{\mathrm{en}}+1) - H(x_{\mathrm{en}}+1)| + \\ |\hat{H}(x_{\mathrm{en}}-1) - H(x_{\mathrm{en}}-1)|]. \end{aligned} \tag{12}$$

The shadow entrance and exit points determine the shadow ceiling in (C1). The shadow entrance points are a dominant factor in (C2). Therefore, both the shadow entrance and exit points are significant for restoring DEM. We thus enhance disagreement at the shadow entrance and exit and define the shadow boundary loss function L_{b} as follows:

$$L_{\mathrm{b}} = \sum_{x_{\mathrm{en}} \in X_{\mathrm{en}}} |\hat{H}(x_{\mathrm{en}}) - H(x_{\mathrm{en}})| + \sum_{x_{\mathrm{ex}} \in X_{\mathrm{ex}}} |\hat{H}(x_{\mathrm{ex}}) - H(x_{\mathrm{ex}})|, \tag{13}$$

V. Shadow Constrained CNN

Our baseline CNN may produce results violating shadow formation mechanisms. To avoid this problem, we now incorporate geometric shadow constraints, i.e., the loss functions (9), (12) and (13), into a CNN to achieve shadow guided training. Such a scheme is referred as shadow constrained convolutional neural network (SCCNN), which is illustrated in Fig. 1 (red).

We retain the KL divergence and ℓ_1 losses to obtain the overall loss function for our SCCNN as follows:

$$L = \alpha|\hat{H} - H| + \beta D_{\mathrm{KL}}(\hat{H}|H) + \gamma_c L_{\mathrm{c}} + \gamma_v L_{\mathrm{v}} + \gamma_b L_{\mathrm{b}}. \tag{14}$$

The parameters α, β, γ_c, γ_v and γ_b balance the effects of different terms in the overall loss function. The SCCNN model takes incomplete DEM data as inputs and complete DEM data as targets with shadow maps as guiding knowledge. It employs the same U-net deep structure as the original CNN. The SCCNN model learns the transition between valid DEM data and void DEM data with disagreement losses enhanced by geometric shadow constraints, which enable the SCCNN to encode certain knowledge of shadow cues. Therefore, unlike the original CNN which is only driven by training data, our SCCNN not only learns from valid data but also follows geometric rules. It thus potentially has more effective restoration performance than a straightforward CNN.

VI. Empirical Validation

We use rectangular mountainous areas of western China, i.e., from $29°N85°E$ to $28°N86°E$, as the investigated region. We obtain remote sensing data of the investigated region from the SRTM version 2[1], which contains plenty of voids especially in mountainous areas. We use the data from the SRTM version 2 as incomplete DEM data. We use the corresponding data from the SRTM version 4, which does not contain voids [13], as ground truth DEM data. Shadow maps are segmented from the satellite images of Landsat-5 for the same region.

In our experiment, we empirically compare the void filling results obtained from the IWD based interpolation method [6] , the baseline CNN and the proposed SCCNN. We use 36 non-overlapping scenes to evaluate different methods. For the two learning models, i.e., CNN and SCCNN, cross validations are performed by using 33 scenes and 3 scenes for training and testing, respectively.

A. Qualitative evaluations

Fig. 4 illustrates the void filling results obtained from the IWD based interpolation method, the CNN and the SCCNN.

The first row in Fig. 4 displays the incomplete DEM data. The second, third and fourth rows illustrate the void filling results by using different methods. The bottom row shows the ground truth DEM data. The regions inside red boxes illustrate detailed contrastive restoration results obtained from different methods. We observe that the results from the SCCNN model agree best with the ground truth DEM data among the three methods.

[1]https://dds.cr.usgs.gov/srtm/version2_1/SRTM3/

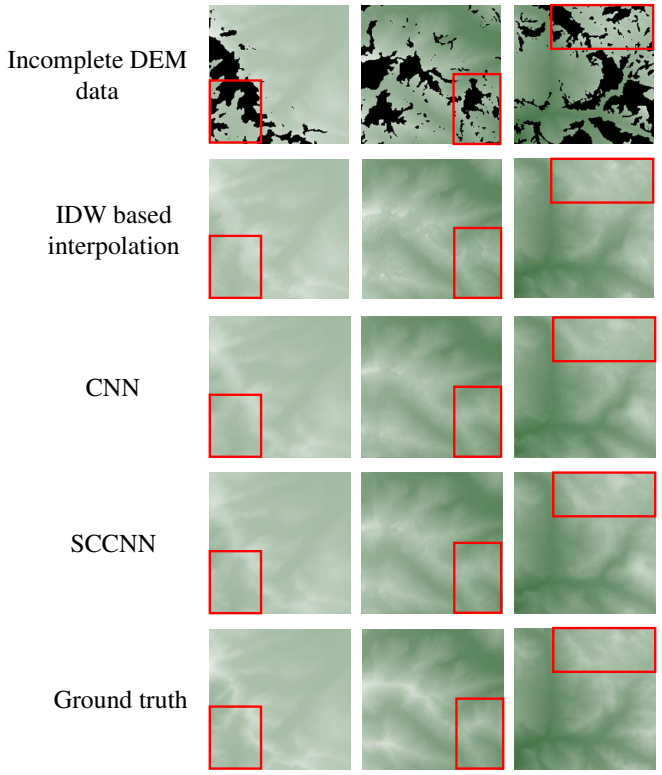

Fig. 4: Void filling results from different methods.

In order to further illustrate the performance difference qualitatively, we examine the sectional views of a restored mountain curve along $85°34'E$ at around $28°31'N$. The corresponding sectional views of the restored mountain curve are shown in Fig. 5. The top row shows the full sectional view and the bottom row displays the zoomed in restored curves with respect to one shadow area. It is clear that the SCCNN outperforms both the IWD based interpolation method and the CNN methods.

B. Quantitative evaluations

TABLE I: Peak signal to noise ratio (PSNR).

Images	Interpolation	CNN	SCCNN
$28°31'N85°34'E$	38.48dB	42.41dB	43.00dB
$28°23'N85°09'E$	37.18dB	37.19dB	38.58dB
$28°21'N85°29'E$	37.53dB	37.88dB	38.25dB

TABLE II: Root mean square error (RMSE).

Images	Interpolation	CNN	SCCNN
$28°31'N85°34'E$	133.75m	84.26m	80.16m
$28°23'N85°09'E$	155.31m	151.88m	130.45m
$28°21'N85°29'E$	149.33m	144.89m	138.39m

We use both peak signal to noise ratio (PSNR) and root mean square error (RMSE) for quantitatively evaluating the restoration accuracy. A larger PSNR value reflects better accuracy. On the other hand, a smaller RMSE reflects better

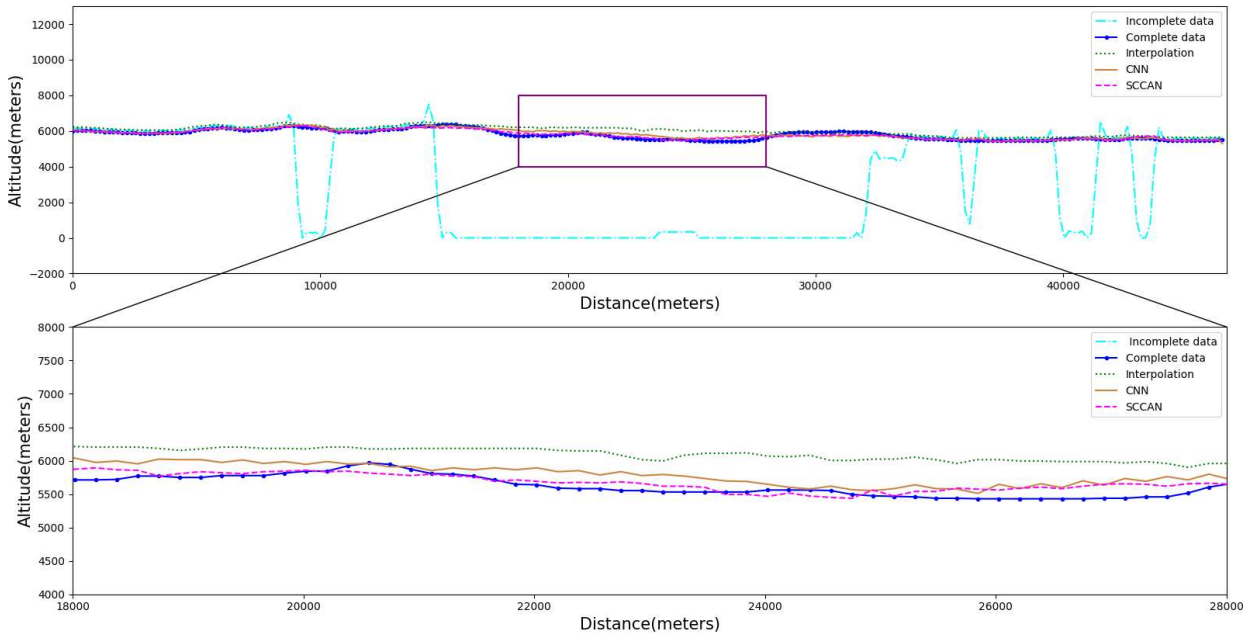

Fig. 5: A cross section comparing incomplete, ground truth and three restoration results.

accuracy. Table I shows the PSNR values of the IWD based interpolation method, the baseline CNN and the SCCNN.

We observe from Table I that both the CNN and the SCCNN significantly outperform the IWD based interpolation method. The key effective factor is that the two deep learning methods characterize and learn the varying heuristics of mountains from the training data, and in contrast the IWD based interpolation method does not explore the training data but just employs test data for restoration. Additionally, benefiting from incorporating the geometric shadow constraints into training the model, the SCCNN outperforms the baseline CNN.

Table II shows the RMSE values of different methods. Similar to those in Table I, both CNN and SCCNN exhibit much better RMSE than the IWD based interpolation method, and our SCCNN obtains the best RMSE among the three methods.

VII. Conclusions

We have presented a shadow constrained convolutional neural network (SCCNN) for filling the mountainous voids of a digital elevation map (DEM) and thus obtained the restored DEM. Compared with straightforward deep learning models such as convolutional neural networks (CNN), the proposed SCCNN model is characterized by geometric shadow constraints. Unlike the pure data driven strategy conducted via the straightforward CNN, the geometric shadow constraints endow our SCCNN with certain knowledge of shadow cues. The geometric shadow constraints incorporated into the SCCNN are in favor of restoring DEMs following the shadow cues. Therefore, the SCCNN potentially avoids the restoration which violates the geological shadowing rules. Empirical comparisons confirm that the SCCNN outperforms the IWD based interpolation method and the CNN based methods. In the future, we will investigate how to incorporate the shadow cues into more comprehensive deep learning methods such as the generative model in [11].

References

[1] P. Rizzoli, M. Martone, C. Gonzalez, C. Wecklich, D. B. Tridon, B. Bräutigam, M. Bachmann, D. Schulze, T. Fritz, M. Huber *et al.*, "Generation and performance assessment of the global tandem-x digital elevation model," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 132, pp. 119–139, 2017.

[2] K. Yang, L. C. Smith, V. W. Chu, C. J. Gleason, and M. Li, "A caution on the use of surface digital elevation models to simulate supraglacial hydrology of the greenland ice sheet," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 11, pp. 5212–5224, 2015.

[3] K. Jafarzadegan and V. Merwade, "A dem-based approach for large-scale floodplain mapping in ungauged watersheds," *Journal of Hydrology*, vol. 550, pp. 650 – 662, 2017.

[4] C. Hirt, "Artefact detection in global digital elevation models (dems): The maximum slope approach and its application for complete screening of the srtm v4.1 and merit dems," *Remote Sensing of Environment*, vol. 207, pp. 27 – 41, 2018.

[5] T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth *et al.*, "The shuttle radar topography mission," *Reviews of geophysics*, vol. 45, no. 2, 2007.

[6] H. I. Reuter, A. Nelson, and A. Jarvis, "An evaluation of void-filling interpolation methods for srtm data," *International Journal of Geographical Information Science*, vol. 21, no. 9, pp. 983–1008, 2007.

[7] D. J. Milan, G. L. Heritage, A. R. Large, and I. C. Fuller, "Filtering spatial error from dems: Implications for morphological change estimation," *Geomorphology*, vol. 125, no. 1, pp. 160–171, 2011.

[8] J. Hogan and W. A. P. Smith, "Refinement of digital elevation models from shadowing cues," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1181–1188.

[9] F. Ling, Q. Zhang, and C. Wang, "Filling voids of srtm with landsat sensor imagery in rugged terrain," *International Journal of Remote Sensing*, vol. 28, no. 2, pp. 465–471, 2007.

[10] J. Zhu, L. Fang, and P. Ghamisi, "Deformable convolutional neural networks for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 8, pp. 1254–1258, 2018.

[11] K. Gavriil, G. Muntingh, and O. J. Barrowclough, "Void filling of digital elevation models with deep generative models," *arXiv preprint arXiv:1811.12693*, 2018.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[13] A. Jarvis, H. I. Reuter, A. Nelson, and E. Guevara, "Hole-filled SRTM for the globe Version 4." CGIAR, 2008.