

# TERRAIN GAP FILLING FROM SINGLE-VIEW SATELLITE IMAGERY BASED ON PRIOR GEOGRAPHICAL KNOWLEDGE CONSTRAINTS

## SUPPLEMENTARY DOCUMENT

*Jianguo Pan\*, Huachao Yang\*, Zihao Lian\*, Peichi Zhou\*, Yuan Yang\*, Chen Li<sup>§</sup>*

\*The College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, China

<sup>§</sup>School of Computer Science and Engineering, Tianjin University of Technology, China

### 1. RELATED WORK

To address the problem of missing data in digital elevation models (DEMs), conventional approaches include manual reconstruction, integration with other data sources, as well as various interpolation techniques. However, these methods face significant limitations. Manual reconstruction is often costly in terms of time and labor, the integration of DEM from different sources frequently results in quality inconsistencies, and interpolation in complex terrain regions often fails to achieve satisfactory results [1, 2, 3]. Therefore, deep learning techniques, owing to their powerful data learning and pattern recognition capabilities, have gradually become the mainstream approach for addressing missing data in DEM. Unlike traditional methods, deep learning approaches are able not only to learn local terrain features but also to capture global information and contextual characteristics. This enables the model to extract latent information from consistent terrain patterns, thereby facilitating the effective reconstruction of complete DEM [4, 5].

Numerous inpainting models based on generative adversarial networks have demonstrated outstanding performance in practical applications. For example, Nazeri et al. proposed EdgeConnect [6], which performed inpainting guided by image edges in the missing regions and achieved remarkable results in pixel-level tasks. However, inpainting methods based on pixel-level loss functions neglect the semantic constraints inherent in terrain, thereby violating the principle of terrain continuity. As a result, the completed terrain often exhibits artifacts such as distorted shadow distributions and disrupted runoff patterns, which are inconsistent with real-world physical rules [7, 8].

To address these issues, many studies have leveraged the CGAN framework in combination with specific terrain information to improve refinement effectiveness. For example, Dong employed the automatically extracted projected shadow maps together with known solar directions as shadow-based supervisory signals, in conjunction

with the conventional supervision derived directly from DEM, thereby successfully enhancing the quality of the refinement results [7]. Qiu et al. [9], on the other hand, jointly trained on global mountainous SRTM data along with elevation-related terrain features such as relief, effectively capturing more fine-grained topographic information and thereby enhancing the model's capability to complete missing regions.

Although these methods have made certain progress in integrating terrain information, several challenges remain, particularly regarding how to optimize the selection and integration of such information to further enhance the quality of the reconstructed DEM. For example, identifying which specific types of terrain information can effectively improve refinement performance and how to accurately incorporate them into deep learning models remain open issues. To address these challenges, Li et al. [10] proposed a constrained terrain knowledge-based CGAN model (TKCGAN), which effectively incorporates terrain feature knowledge into the training process, thereby enhancing the model's ability to recover critical terrain characteristics. Zhou et al. [11] proposed a multi-scale feature fusion CGAN approach, which, after preliminary inpainting, employs a multi-attention refinement network to further recover details in the missing regions and introduces a channel-spatial pruning mechanism to enhance network performance.

In summary, inpainting plays a crucial role in addressing missing data in DEM, and deep learning methods have achieved remarkable results in terrain inpainting tasks. However, key challenges remain in further optimizing the integration and feature learning of terrain information, as well as in enhancing the model's generalization ability and refinement performance. First, existing approaches often underutilize critical terrain features, such as slope, aspect, and curvature, limiting the recovery of fine-scale details. Second, integrating multi-source terrain information (e.g., elevation, texture, satellite image) remains challenging, leading to potential information loss or redundancy. Third, current models

exhibit limited generalization to unseen or complex terrains, reducing the reliability of predictions. Addressing these issues requires more effective strategies for terrain feature selection, multi-source integration, and enhancement of model generalization and refinement capabilities.

## 2. METHOD

### 2.1. Differentiable renderer based on mountain shadow priors

The shadow-constrained terrain gap-filling framework renders the DEM using a differentiable renderer to obtain a shadow mask corresponding to the generated terrain, which is then employed to constrain model training. This section presents the process of generating shadow masks with a differentiable renderer.

In the shadow mask generation task, to ensure that the generated shadow masks are consistent with the real shadow masks in terms of illumination, it is first necessary to determine the illumination direction corresponding to each satellite image. Then, the generated terrain is illuminated under the same lighting conditions to maintain consistent lighting effects. This study adopts the illumination direction inversion algorithm proposed by Zhou et al. [8] to estimate the zenith and azimuth angles of the light source.

The algorithm first preprocesses the satellite image to extract an approximate shadow distribution, as described in Algorithm 1. In each iteration, a differentiable rendering algorithm is used to render the digital elevation model, from which shadow distribution information is obtained. Then, the zenith and azimuth angles are treated as optimization targets, and the L1 distance error between the target shadow distribution and the rendering result is minimized using the Adam optimization algorithm. Finally, when the error falls below a set threshold, the accurate lighting conditions are obtained, completing the illumination inversion process.

---

#### Algorithm 1 Inverse Lighting Estimation Algorithm

---

**Require:** Satellite image  $M$ , digital elevation model  $T$ , threshold  $\epsilon$

**Ensure:** Zenith angle  $\theta$ , azimuth angle  $\phi$

- 1: Extract approximate shadow distribution:  $S_{gt} = \text{ExtractShadow}(M)$
  - 2: Initialize lighting parameters  $(\theta, \phi)$
  - 3: **while**  $L > \epsilon$  **do**
  - 4:     Render shadow:  $S_{pred} = \text{DiffRender}(T, \theta, \phi)$
  - 5:     Optimize:  $(\theta, \phi) \leftarrow \text{AdamOpt}(L, \theta, \phi)$
  - 6: **end while**
  - 7: **return**  $(\theta, \phi)$
- 

For the reconstructed terrain, the method simulates

sunlight with white light and applies white as the terrain texture. The zenith and azimuth angles obtained from illumination inversion are used as the lighting direction for rendering. Subsequently, the method uses a differentiable image processing algorithm to convert the rendered result from the RGB color space to the YUV color space to better capture the luminance and chrominance information of the rendered image. Then, a differentiable contrast-limited adaptive histogram equalization algorithm is applied to the YUV image to enhance local contrast. Finally, the processed image is converted back to the RGB color space to obtain the final shadow mask.

## 3. EXPERIMENTAL RESULTS AND ANALYSIS

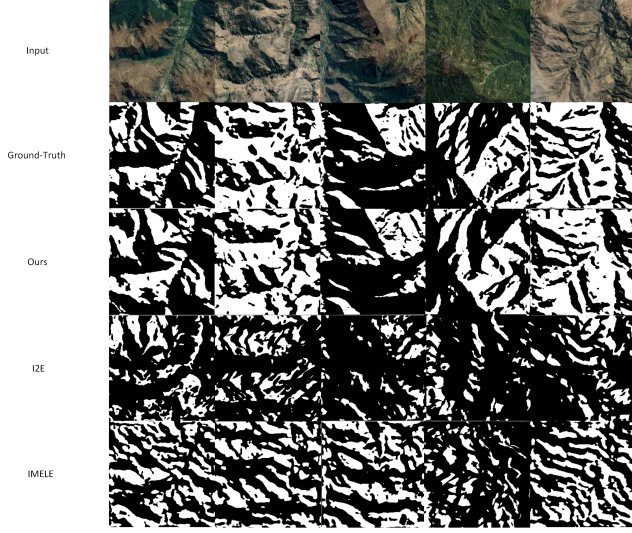
The training and testing of the terrain prediction module and the terrain refinement module are conducted in the Python 3.8.19 and PyTorch 1.13.0 environment, deployed on a desktop computer equipped with an NVIDIA GeForce RTX 3090 Ti 24GB GPU and an Intel Core i5-13600K 3.50 GHz CPU. The satellite image in the dataset is obtained from Google Maps at a zoom level of 12, and the corresponding DEMs are derived from Tangrams Heightmapper<sup>1</sup>, maintaining the same zoom level. Both the satellite images and the DEMs are processed at a resolution of  $256 \times 256$ . The training set consists of 4,096 paired samples of satellite imagery, DEMs, and shadow masks, while the testing set contains 1024 samples.

The terrain prediction module is optimized using the Adam optimizer with an initial learning rate of 0.0003 under a linear learning rate scheduling strategy. The momentum parameters are set to  $\beta_1 = 0.5$  and  $\beta_2 = 0.999$ . The loss weighting coefficients are assigned as  $\lambda_{g_1} = 1.1$ ,  $\lambda_m = 2.0$ ,  $\lambda_t = 0.5$ , and  $\lambda_s = 10^{-4}$ . The batch size is set to 4, and the model is trained for a total of 300 epochs.

The terrain refinement module is also trained using the Adam optimizer, with an initial learning rate of 0.0002. The momentum parameters are set to  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ . The loss weighting coefficients are assigned as  $\lambda_{g_2} = 1.0$ ,  $\lambda_m = 2.0$ ,  $\lambda_i = 1.5$ ,  $\lambda_s = 10^{-4}$ , and  $\lambda_k = 0.8$ . The batch size is set to 4, and the model is trained for a total of 400 epochs.

Fig. 1 shows the comparison results of shadow masks generated by different methods. In this experiment, differentiable rendering is employed to simulate illumination, producing terrain shadow masks consistent with the lighting conditions of the satellite image. It can be observed that the shadow distribution generated by our method is closest to the ground truth, demonstrating

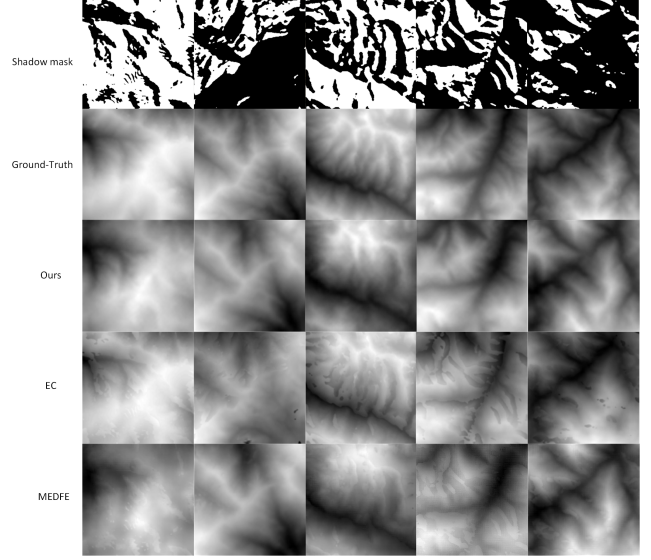
<sup>1</sup><https://tangrams.github.io/heightmapper/>



**Fig. 1.** Comparison results of terrain refinement shadow masks. The top row shows the input satellite imagery, while the second row depicts the ground-truth shadow masks. Rows 3–5 present the shadow masks generated by our method, I2E, and IMELE, respectively. It can be observed that our approach produces shadow distributions most consistent with the ground truth, effectively capturing terrain semantics, whereas I2E shows significant errors and IMELE generates fragmented shadows.

superior capability in learning both shadow constraints and terrain semantics. In contrast, the shadow mask produced by the I2E method exhibits significant errors, exposing the limitations of relying solely on pixel-space analysis. This indicates that the model fails to correctly distinguish between the optical properties of surface materials and the geometric characteristics of actual terrain undulations, resulting in mismatches between shadow distribution and terrain generation logic. The IMELE method, on the other hand, generates fragmented shadows, highlighting the limitations of conventional convolutional architectures. Although convolution-deconvolution networks are able to capture local texture features, their restricted receptive fields hinder the ability to model the systematic relationships between shadow distributions and geomorphological structures in real terrain.

As illustrated in Fig. 2, our method demonstrates a clear advantage over the baseline models in the task of filling missing regions within shadowed areas of complex terrain. The EC model generates irregular artifacts along the boundaries of shadow masks, which stem from structural limitations inherent in its two-stage generation mechanism. Specifically, the edge generation stage relies on ridgelines and valley lines as prior guidance; however, the spatial misalignment between shadow re-



**Fig. 2.** Comparison of terrain gap filling effects. The first row shows the shadow masks, followed by the ground-truth DEMs. Subsequent rows display the reconstructed terrain using our method, EC, and MEDFE, respectively. Our method better preserves terrain continuity within shadowed regions and avoids the irregular artifacts and elevation distortions seen in EC and MEDFE.

gions and terrain feature lines introduces deviations in edge prediction. Furthermore, the subsequent refinement stage employs a pixel-wise reconstruction loss without explicit enforcement of terrain continuity, leading to non-physical oscillations in local elevation values. The MEDFE model, although stable in large continuous shadow regions, produces significant distortions in complex mask areas where illuminated and shadowed regions are interwoven. On one hand, the channel-attention mechanism of its feature equalization module excessively amplifies high-frequency texture features, which conflicts with the low-frequency continuity characteristics of terrain data, resulting in grid-like artifacts within the restored regions. On the other hand, the spatial equalization strategy, which depends on local neighborhood statistics, fails to capture large-scale terrain gradient constraints, thereby causing discontinuous elevation jumps in the reconstructed results.

#### 4. REFERENCES

- [1] F. Hallo, G. Falorni, and R. L. Bras, “Characterization and quantification of data voids in the shuttle radar topography mission data,” *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 177–181, 2005.
- [2] S. J. Boulton and M. Stokes, “Which DEM is best for analyzing fluvial landscape development in mountainous terrains?,” *Geomorphology*, vol. 310, pp. 168–187, 2018.
- [3] H. I. Reuter, A. Nelson, and A. Jarvis, “An evaluation of void-filling interpolation methods for SRTM data,” *Int. J. Geogr. Inf. Sci.*, vol. 21, no. 9, pp. 983–1008, 2007.
- [4] G. Dong, F. Chen, and P. Ren, “Filling SRTM void data via conditional adversarial networks,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2018, pp. 7441–7443.
- [5] W. Li and C. Hsu, “Automated terrain feature identification from remote sensing imagery: A deep learning approach,” *Int. J. Geogr. Inf. Sci.*, vol. 34, no. 4, 2020.
- [6] K. Nazeri, E. Ng, T. Joseph, et al., “EdgeConnect: Structure guided image inpainting using edge prediction,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2019, pp. 3265–3274.
- [7] G. Dong, W. Huang, W. A. P. Smith, and P. Ren, “A shadow-constrained conditional generative adversarial net for SRTM data restoration,” *Remote Sens. Environ.*, vol. 237, pp. 111602, 2020.
- [8] P. Zhou, D. Lu, C. Li, et al., “Unsupervised textured terrain generation via differentiable rendering,” in *Proc. ACM Int. Conf. Multimedia (ACM MM)*, 2022, pp. 2654–2662.
- [9] Z. Qiu, L. Yue, and X. Liu, “Void-filling of digital elevation models with a terrain texture learning model based on generative adversarial networks,” *Remote Sens.*, vol. 11, no. 23, pp. 2829, 2019.
- [10] S. Li, G. Hu, X. Cheng, et al., “Integrating topographic knowledge into deep learning for the void-filling of digital elevation models,” *Remote Sens. Environ.*, vol. 269, pp. 112818, 2022.
- [11] G. Zhou, B. Song, P. Liang, et al., “Voids filling of DEM with multi-attention generative adversarial network model,” *Remote Sens.*, vol. 14, no. 5, pp. 1206, 2022.