TERRAIN GAP FILLING FROM SINGLE-VIEW SATELLITE IMAGERY SUPPLEMENTARY DOCUMENT

Jianguo Pan, Huachao Yang*, Peichi Zhou*, Yuan Yang*, Chen Li*$

*The College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, China
$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 and fusion 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].

Convolutional Generative Adversarial Networks(CGANs), as a representative deep learning architecture, have been widely applied to the task of missing data reconstruction in DEM.Through adversarial learning, this model can not only generate height maps that resemble real data but also effectively restore missing regions. Numerous inpainting models based on generative adversarial networks have demonstrated outstanding performance in practical applications. For example, Nazeri et al. proposed EdgeConnect, which performs inpainting guided by image edges in the missing regions and has achieved remarkable results in pixel-level tasks [6]. 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 in an attempt to improve the effectiveness of the refinement models.For example, Dong employed 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, on the other hand, jointly trained on global mountainous SRTM data together 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 [9].

Although these methods have made certain progress in the integration of terrain information, several challenges remain, particularly regarding how to optimize the selection and integration of terrain 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 such information into deep learning models remain unresolved issues.To address these challenges, Li proposed a constrained terrain knowledge-based CGAN model (TKCGAN), which effectively transfers knowledge of terrain features into the training process, thereby enhancing the model' s ability to recover critical terrain characteristics [10]. Zhou proposed a multi-scale feature fusion CGAN approach, which, after performing preliminary inpainting, employs a multi-attention refinement network to further recover details in the missing regions, while introducing a channel-spatial pruning mechanism to enhance network performance [11].

In summary,inpainting plays a crucial role in addressing missing data in DEM, and deep learning methods represented by CGANs 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.

2. METHOD

2.1. Differentiable renderer based on mountain shadow priors

The shadow-constrained terrain refinement framework renders DEM through a differentiable renderer to obtain the shadow mask corresponding to the generated terrain, which is used to constrain model training.This chapter mainly introduces the process of generating shadow masks through 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 chapter adopts the illumination direction inversion algorithm proposed by Zhou et al. to obtain the zenith and azimuth angles of the light source [8].

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 ϵ

Ensure: Zenith angle θ, azimuth angle ϕ

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

For the reconstructed digital elevation model, the method simulates sunlight using white light and employs 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 (CLAHE) 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.

2.2. The terrain prediction module

we also introduces the style loss function [12],

$$L_{style} = \|Gram(T_{gt}) - Gram(T_{\text{pred}})\|_1 \quad (1)$$

Here, $Gram(\cdot)$ denotes the Gram matrix, which is used to capture the style features of the terrain. $\|\cdot\|_1$ represents the L1 norm, which measures the discrepancy between the style features of the generated terrain map and the real terrain map. The style loss function L_{style} achieves style matching by comparing the Gram matrices of the generated and real terrain maps. Essentially, the Gram matrix represents the style features of an image by capturing the correlations between different terrain structures. Minimizing the L1 loss between these two Gram matrices.

The overall loss function can then be expressed as:

$$L_{predict} = \lambda_{g_1} L_{G_1} + \lambda_m L_{mask} + \lambda_t L_{style} + \lambda_s L_{slope} \quad (2)$$

Here, λ_{g_1}, λ_m, λ_t, andλ_s represent the weighting coefficients for the generator loss, shadow mask loss, style loss, and slope loss, respectively. λ_{g_1} is used to control the overall image quality and the stability of generator training, λ_m emphasizes the distribution characteristics of the shadow mask, helping to improve the consistency of terrain reconstruction in illuminated regions, λ_t maintains the continuity and realism of image style, while λ_s suppresses noise in local terrain regions and enhances the spatial smoothness of the elevation map.

3. EXPERIMENTAL RESULTS AND ANALYSIS

The training and testing of the digital elevation prediction module and the digital elevation refinement module were conducted in a 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 imagery in the dataset was obtained from Google Maps at a zoom level of 12, and the corresponding digital elevation models (DEMs) were derived from Tangrams Heightmapper [1], maintaining the same zoom level. Both the satellite images and the DEMs were processed at a

[1]https://tangrams.github.io/heightmapper/

resolution of 256×256. The training set consists of 4,096 paired samples of satellite imagery, DEMs, and shadow masks, while the testing set contains 1,024 samples.

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

The digital elevation refinement module was also trained using the Adam optimizer, with an initial learning rate of 0.0002. The momentum parameters were set to $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The loss weighting coefficients were 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 was set to 4, and the model was trained for a total of 400 epochs.

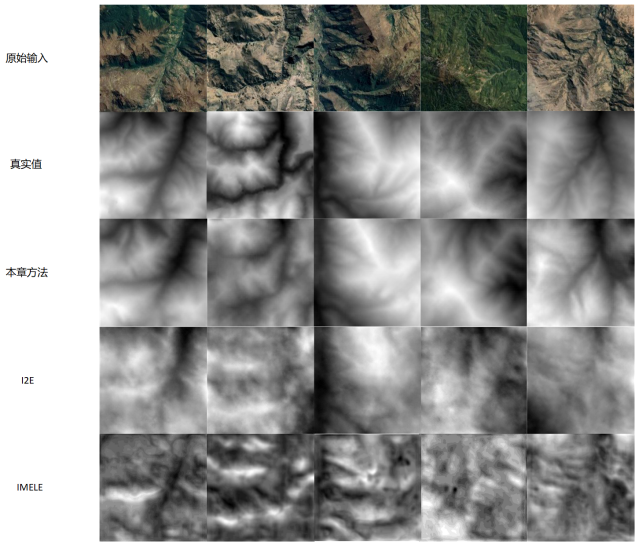

Fig. 1. Comparison results of terrain refinement elevation map

Fig. 1 presents the results of DEM refinement for the input satellite imagery. The results demonstrate that when large shadowed regions appear in the satellite images, the comparison methods exhibit inferior performance. For terrains at the same elevation, the I2E method produces inconsistent elevation predictions between the illuminated and shaded sides. This is because I2E, which relies on feature analysis in the satellite image pixel space, misinterprets dark shadows as terrain entities (e.g., valley depressions or mountain ridges). Its pixel-space–based refinement mechanism cannot decouple the inherent coupling between shadows under illumination conditions and the underlying terrain. The results generated by the IMELE method, on the other hand, contain substantial noise, particularly in the shadowed regions. This arises from the local receptive field characteristics of its convolution–deconvolution neural network architecture, which lacks long-range terrain continuity constraints. In the absence of physically grounded priors, repetitive local feature extraction amplifies the interference caused by shadow occlusion, leading to unresolved multi-solution ambiguities in shadowed areas and ultimately trapping the refinement results in a local optimum.

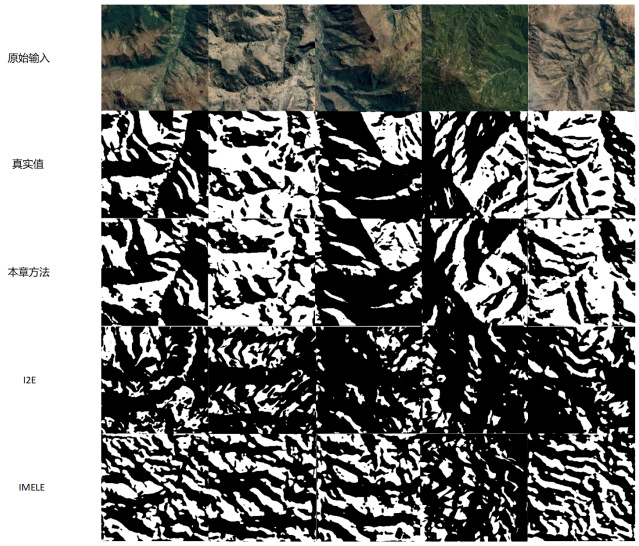

Fig. 2. Comparison results of terrain refinement shadow masks

Figure 2 shows the comparison results of shadow masks generated by different methods. In this experiment, differentiable rendering was employed to simulate illumination, producing terrain shadow masks consistent with the lighting conditions of the satellite imagery. It can be observed that the shadow distribution generated by our method is closest to the ground truth, demonstrating superior capability in learning both shadow constraints and terrain semantics. In contrast, the shadow masks produced by the I2E method exhibit 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 geomor-

phological structures in real terrain.

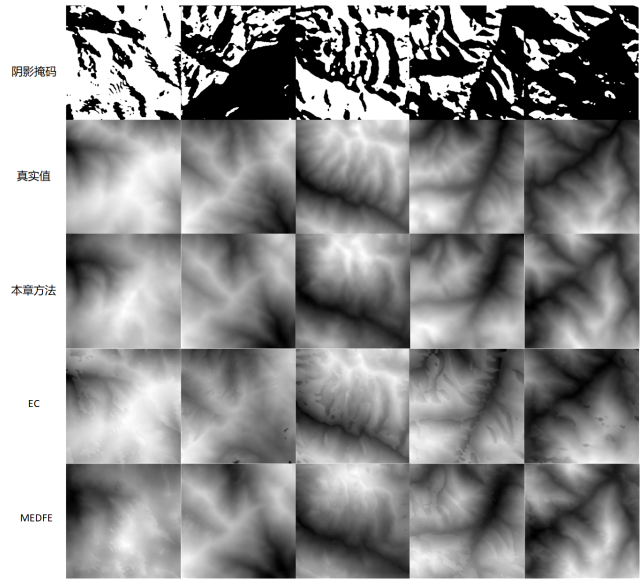

Fig. 3. Comparison of terrain vacancy filling effects

As illustrated in Fig. 3, our method demonstrates a clear advantage over the comparison 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 regions 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.

As shown in Table. 1, the proposed method consistently outperforms IMELE, I2E, and the ablation models across all four evaluation metrics, demonstrating superior performance in multiple dimensions. These results indicate that our approach produces visual details that more closely resemble real terrain data,and that the feature distributions of the generated outputs align well with those of the ground truth. The ablation study further reveals that although the model without the terrain refinement module achieves higher PSNR values compared to I2E and IMELE, it performs worse on LPIPS and FID. This suggests that while a single prediction module can improve reconstruction accuracy in illuminated regions, it lacks the ability to address the inherent ambiguity in shadowed areas, thereby reducing the overall quality of terrain refinement. By integrating both the prediction and refinement modules, our method demonstrates the critical role of the shadow-constrained framework in achieving high-fidelity terrain refinement.

Table 1. QUANTITATIVE EVALUATION OF TERRAIN refinement FRAMEWORK

Method	LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓
IMELE [13]	0.66	12.02	0.43	1.76
I2E [14]	0.55	13.78	0.64	1.45
w/o refinement	0.62	14.63	0.55	1.79
ours	**0.45**	**18.61**	**0.84**	**1.28**

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.

[12] L. Yue, B. Gao, and X. Zheng, "Generative DEM void filling with terrain feature-guided transfer learning assisted by remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.

[13] C.-J. Liu, V. A. Krylova, P. Kane, et al., "IM2ELEVATION: Building height estimation from single-view aerial imagery," *Remote Sens.*, vol. 12, no. 17, pp. 2719, 2020.

[14] E. Panagiotou, G. Chochlakis, L. Grammatikopoulos, et al., "Generating elevation surface from a single RGB remotely sensed image using deep learning," *Remote Sens.*, vol. 12, no. 12, pp. 2002, 2020.