GENERATIVE MODELING FOR TOPOGRAPHY SUPER-RESOLUTION ON THE MOON

Matthew Repasky[1*], Erwan Mazarico[2], Stefano Bertone[2,3], Michael Barker[2], Terence Sabaka[2], and Yao Xie[1], [1]Georgia Institute of Technology, Atlanta, GA. [2]NASA Goddard Space Flight Center, Greenbelt, MD. [3]University of Maryland, College Park, MD. [*][mwrepasky@gatech.edu]

Abstract. *Generative Artificial Intelligence (AI) frameworks provide the means to sample from intractable prior probability distributions. Conditional generative models consider contextual information, such as degraded data or partial observations, to effectively sample from conditional distributions. We apply conditional generative AI to digital elevation models (DEMs) derived by Lunar Orbiter Laser Altimeter (LOLA) measurements at the Moon. Our Schrödinger bridge models are developed by optimizing vision transformers with respect to datasets of low-resolution DEMs and high-resolution optical imagery. Trained models are leveraged for lunar topography super-resolution, facilitating reconstruction of larger areas than those feasible using analytical methods.*

Introduction. A high-fidelity understanding of planetary topography is necessary for accurate modeling of surface conditions. This is of particular interest for areas of potential future human and robotic exploration, such as the candidate landing sites for the upcoming Artemis missions.[1] Altimetry measurements provided by LOLA have been used to develop topography models at moderate resolution,[2] e.g., down to 20 meters-per-pixel (mpp) near the lunar poles. However, higher-resolution topography is needed in many regions of interest.

Analytical methods, such as Shape-from-Shading (SfS),[3,4] incorporate contextual information in the form of high-resolution optical imagery, such as those provided by the Lunar Reconnaissance Orbiter Narrow-Angle Camera (LRO NAC). SfS takes an *a priori* low-resolution DEM as input alongside co-registered images at the target resolution, where each image is illuminated by the Sun from a different direction. While such methods provide statistical guarantees and interpretability of the output high-resolution DEM, they are computationally expensive and require human input (e.g., parameter fine-tuning). As such, it is cumbersome to apply to large areas.

We implement generative-AI-based super-resolution tools to develop accurate high-resolution DEMs on the Moon. In particular, we apply the Image-to-Image Schödinger Bridge (SB) method[5] to the conditional generation setting, which has enjoyed considerable success in super-resolution tasks. Our Image-to-Image SB transforms *a priori* samples (low-resolution DEMs) to *a posteriori* high-resolution DEMs while considering a set of optical images. Generative AI methods have the potential to scale more efficiently to larger inputs than analytical methods, and can generalize beyond the training dataset.

Data. We compiled a dataset of lunar topography patches at 20 mpp extracted uniformly from LOLA DEMs[2] of both lunar poles, over latitudes poleward of 80°. Each DEM is a 96 by 96 pixel (1.92 km) array, and we extracted 91,708 high-quality patches over the lunar surface. These data capture fine-grained features of lunar surface topography, representing a wide range of elevations and topographic features. For each DEM patch in the dataset, we also render 30 simulated LRO NAC "images". In each case, the DEM is illuminated from 30 Sun angles randomly scattered throughout the sky.

Schrödinger Bridges. Generative modeling frameworks aim to generate samples x_0 from the intractable, data-generating prior p_0. The Image-to-Image SB method[5] defines the generative process by a bridge between the target p_0 supported on $\mathcal{X}_0$ (the space of unobserved data, e.g., high-resolution topography) and a separate distribution p_1 supported on $\mathcal{X}_1$ (the space of observed data, e.g., low-resolution topography). This process is captured by the following Stochastic Differential Equations (SDEs) in the data x_t:

$$dx_t = f_t(x_t)dt + \sqrt{\beta_t}dw_t, \quad x_0 \sim \hat{\Psi}(\cdot, 0), \tag{1}$$

$$dx_t = f_t(x_t)dt + \sqrt{\beta_t}d\bar{w}_t, \quad x_1 \sim \Psi(\cdot, 1). \tag{2}$$

In (1) and (2), f_t is a linear drift in x_t, β_t is a variance schedule parameterizing the diffusion, and $dw_t/d\bar{w}_t$ represent Brownian motion in forward/reverse time. Time-varying potentials Ψ and $\hat{\Psi}$ are related to the boundary distributions via $\hat{\Psi}(\cdot, 0) = \delta_0(\cdot)$ and $\Psi(\cdot, 1) = p_1/\hat{\Psi}(\cdot, 1)$, where $\delta_0(\cdot)$ is the Dirac delta distribution at $x_0 \in \mathcal{X}_0$.

SBs are a generalization of Denoising Diffusion Probabilistic Models (DDPMs);[6] therefore, DDPM sampling and optimization techniques can be borrowed. Ancestral sampling is employed to iteratively sample $x_{t-1} \sim p(x_{t-1}|x_t)$, where a neural network $\epsilon(x_t, t; \theta)$ parameterized by θ is used to approximate the parameters of $p(x_{t-1}|x_t)$. Letting $f_t = 0$, the optimization objective is

$$L(\theta; x_0, x_1, t) = \left\| \epsilon(x_t, t; \theta) - \frac{x_t - x_0}{\sigma_t} \right\|, \tag{3}$$

where $\sigma_t^2 = \int_0^t \beta_\tau d\tau$, x_t is tractable at training time given x_1 and x_0. See the original work for details.[5] In practice, (3) is minimized via a stochastic gradient descent scheme using mini-batches of boundary pairs x_0 and x_1 with t sampled uniformly.

Finally, we extend SBs to consider additional conditional information. By default, SBs transform the conditional data $x_1 \sim p_1$ into the target data $x_0 \sim p_0$. Following other generative modeling works, and to consider additional context y beyond x_1, we allow the neural network to be a function of y as well: $\epsilon(x_t, t, y; \theta)$. Besides this change, the optimization and sampling scheme are the same as those described earlier in this section.

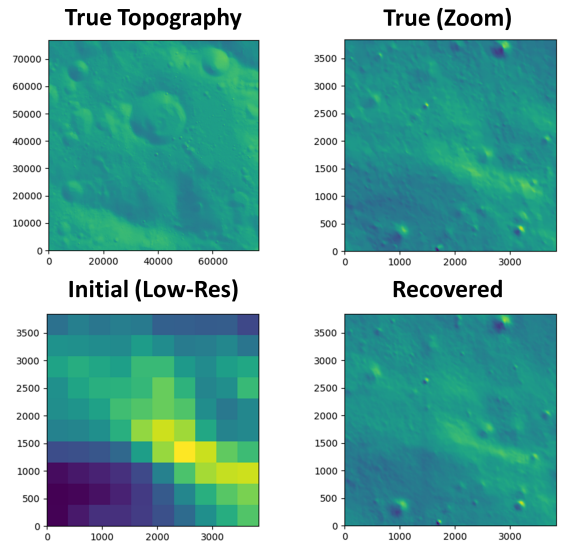

Figure 1. 16× topography super-resolution on a 76.8 km by 76.8 km region (3,840 pixels at 20 mpp) near the lunar South Pole (hillshade shown). Compute time is ∼2 minutes on 2 V100 GPUs. A zoomed 192-pixel region is shown, including bicubic interpolation and SB recovery.

Topography Schrödinger Bridge. In the lunar topography super-resolution settings, p_0 is the distribution of 96 × 96 pixel lunar topography maps at 20 mpp (i.e., $\mathcal{X}_0 \subset \mathbb{R}^{96\times96}$) while p_1 is the distribution of lower-resolution DEMs at 320 mpp (i.e., $\mathcal{X}_1 \subset \mathbb{R}^{6\times6}$). Additional context is provided by the rendered images $y \in \mathbb{R}^{96\times96}$. The goal is to learn a sampling scheme which transforms low-resolution DEMs x_1 into high-resolution DEMs x_0 given a set of images $\mathcal{Y} = \{y_i\}_{i=1}^{30}$.

Latent Space. To improve modeling efficiency, we develop a generative model operating in the latent space of a Variational Auto-Encoder (VAE).[7] VAEs learn a randomized, invertible, low-dimensional embedding of data using encoding/decoding neural networks. Latent generative models generate low-dimensional embeddings of data from the target distribution, which can be decoded using the VAE. We train a convolution-based VAE of approximately 20 million parameters on the high-resolution DEM data, following the approach of latent diffusion.[8] A SB is trained on data within the latent space $\mathcal{Z} \subset \mathbb{R}^{12\times12\times4}$.

Modeling & Optimization. The function $\epsilon(\cdot;\theta)$ is parameterized by a vision transformer[9,10] (ViT) of 6 layers, with 12 attention heads per layer, and patch size of 2 (latent) pixels. The input is the latent representation of the interpolated low-resolution DEM $z_1 \in \mathcal{Z}$ and the set of images $\mathcal{Y}$. Each image $y_i \in \mathbb{R}^{96\times96}$ is input to a shared convolutional encoder akin to the encoding arm of a U-Net,[11] composed of 4 layers with 2 residual blocks per layer. The embedded y_i are averaged before being input to a ViT. The ∼150 million parameters are optimized using AdamW[12] with learning rate 10^{-4}, no weight decay, batch size 256, and otherwise default PyTorch parameters. This was conducted for 1,000 epochs on 4 Nvidia V100 GPUs using 80% of the patches, which were augmented by flipping along each axis.

Results & Application. The latent SB achieves a root-mean-squared reconstruction error of 5.60 meters at 20 mpp over a holdout set of data (not included in training) of 500 patches, while the 95th percentile is 11.1 meters. The average slope error in the horizontal directions are 1.76° and 1.83°. The SB can be applied to large areas via tiling. The reconstruction of each pixel over a large area can be averaged over a sliding window of outputs from the SB; such a result is shown in Figure 1. Given that the sampling scheme is stochastic, a distribution of outputs can be generated for each location, and statistics of these outputs can provide pixel-level error estimation.

Summary. We have extended SBs to consider additional conditional information to model within the latent space of a VAE for super-resolution of lunar topography. This approach yields reasonable reconstruction for 16× super-resolution and can be extended to arbitrarily large areas via a sliding window scheme.

References.

[1] NASA-Press, "NASA identifies candidate regions for landing next americans on moon," *NASA*, 2022.

[2] M. K. Barker et al., "A new view of the lunar south pole from the lunar orbiter laser altimeter (LOLA)," *PSJ*, vol. 4, no. 9, p. 183, 2023. doi: 10.3847/PSJ/acf3e1.

[3] R. Zhang et al., "Shape-from-shading: a survey," *IEEE TPAMI*, vol. 21, no. 8, pp. 690–706, 1999. doi: 10.1109/34.784284.

[4] I. Fernandes and K. Mosegaard, "High-resolution topography from planetary images and laser altimetry," *Planet. and Sp. Sci.*, vol. 218, p. 105514, 2022. doi: 10.1016/j.pss.2022.105514.

[5] G.-H. Liu et al., "I2SB: image-to-image schrödinger bridge," ICML, 2023. doi: 10.48550/arXiv.2302.05872.

[6] J. Ho et al., "Denoising diffusion probabilistic models," NeurIPS, 2020. doi: 10.48550/arXiv.2006.11239.

[7] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013. doi: 10.48550/arXiv.1312.6114.

[8] R. Rombach et al., "High-resolution image synthesis with latent diffusion models," CVPR, 2022. doi: 10.48550/arXiv.2112.10752.

[9] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," ICLR, 2020. doi: 10.48550/arXiv.2010.11929.

[10] W. Peebles et al., "Scalable diffusion models with transformers," ICCV, 2023. doi: 10.48550/arXiv.2212.09748.

[11] O. Ronneberger et al., "U-net: Convolutional networks for biomedical image segmentation," MICCAI, Springer, 2015. doi: 10.1007/978-3-319-24574-4_28.

[12] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," ICLR, 2018. doi: 10.48550/arXiv.1711.05101.