From satellite image to glacier scenery: A novel GlacierPix2Pix model for glacier visualization

Xinyuan Duan

Eastside Preparatory School, Kirkland, Washington, USA

xinyuand6@gmail.com

Abstract. Glacier visualization is critical for studying climate change as glaciers respond rapidly to global warming. It serves as an indicator of environmental changes and provides insights into the magnitude of climate change. Traditional on-site photography faces challenges such as safety hazards, high costs, and tight time constraints. Existing terrain generation models have limitations such as not being capable of using real satellite imagery to control generated terrain and achieving photorealistic image quality. To address these issues, this study introduces GlacierPix2Pix, a novel deep learning model that generates glacier scenery from satellite images. This study uses the Landscape HQ dataset containing 90K terrain images and the USGS Glacier Benchmark dataset containing Digital Elevation Models of 5 different glaciers over many years. This study curates a dataset of over 9K glacier images that can be used in related studies from the LHQ dataset. Glacier contours are extracted from the DEM images and glacier images. Based on the StyleGAN architecture, GlacierPix2Pix comprises a terrain generator, discriminator, and style encoder, and it supports both 2D and 3D methods. The model achieves a Fréchet Inception Distance (FID) of 31.90, which surpasses the performance of other similar datadriven methods. It can successfully produce photorealistic glacier images aligned with satellite-derived contours. This research contributes to a better understanding of the urgent issues surrounding climate change, especially its effect on glaciers. Potential applications of this study include demonstrating glacier recession, helping with mapping snow cover and mass balance, and allowing detailed changes in remote locations to be observed.

Keywords: glacier visualization, digital elevation model, GlacierPix2Pix, contour extraction, Fréchet Inception Distance.

1. Introduction

Glacier visualization holds paramount importance across various fields, such as climate, geography, tourism, and education. Most significantly, it demonstrates environmental changes, particularly those pertaining to climate change.

Glaciers indicate environmental changes and provide insights into the magnitude of climate change. To understand glacier changes over time, glaciologists utilize the method of Glacier Repeat Photography, which requires glacier images taken at the exact location at different times for comparison. However, receding glaciers often render it impossible to return to the same location. Additionally, the process of glacier photography itself poses many significant risks, with potential dangers such as falling ice and hidden crevasses beneath the snow. Glaciologists and mountaineers must take meticulous safety measures, such as utilizing crevasse probes for detection, rope ties for group security, and ice axes and crampons for stability on the ice, which are all costly equipment. Moreover, the narrow window for capturing high-quality glacier images through on-site photography, typically occurring from late August to early September, further adds to the challenges associated with this method. Computer Graphics researchers developed methods to solve similar visualization problems using implicit parametric methods to model 3D terrains [1-3]. However, the visualizations produced by these methods are difficult to manipulate and often fall short of achieving photorealistic quality. Additionally, neither of the aforementioned methods can offer visualization of glaciers from arbitrary viewpoints.

This research develops a novel deep learning model, GlacierPix2Pix, for visualizing glaciers using satellite imagery to overcome the limitations mentioned above. This study proposes a data-driven approach to reconstruct glaciers from given satellite images, which not only eliminates the risks and costs associated with on-site photography, but also achieves the synthesis of photorealistic glacier images from any arbitrary view, therefore increasing the immersive experience of the viewer. Specifically, this study aims to employ advanced terrain generation and deep learning techniques to generate images of different glaciers using contours extracted from digital elevation models (DEMs). In this study, we utilize the Landscape HQ dataset [4] and the USGS Benchmark Glaciers dataset [5].

This study proposes a deep learning model that takes contour images as input and generates images of glaciers that match the input contour with the ability to change the camera position and angle as output. For the model architecture, we employ a latent generative model to ensure consistency in the results. To achieve DEM contour's control of the result, we deploy a style encoding model to control the generative model. Existing 3D terrain generation works using implicit parametric method that is unable to perform conditional generation and achieve photorealistic image quality. Our approach utilizes extracted contours from satellite imagery as a condition for the generative model to achieve realistic generation of glacier visualizations. Diverging from traditional approaches such as on-site photography and existing terrain generation methods, GlacierPix2Pix offers a safer, more costeffective, and more efficient alternative to capturing glacier imagery.

2. Related Work

In the field of image and view extrapolation, Kaneva et al. [6] laid the foundation with their pioneering work on infinite image extrapolation. They employed a vast image database for classical 2D image retrieval, stitching, and rendering. Recent advancements have brought forth learningbased 2D image inpainting [7] and outpainting [8] methods, effectively filling in missing image regions and expanding the field of view by synthesizing coherent content. While earlier attempts explored single-view 3D view extrapolation [9], their limitations are evident, often confined to a narrow range of viewpoints.

In the field of video generation, both unconditional [10] and conditional methods [11] have been developed, operating from noise input to sequences conditioned on images or text prompts. However, extending these concepts to 3D requires multi-view training data, and existing approaches lack persistent 3D scene content at runtime. Recent efforts introduced additional 3D geometry inputs like point clouds [12] or voxel grids [13] to maintain global scene consistency, yet they fall short in achieving consistent content generation.

In the field of generative view synthesis, existing methods strive to produce new scene views from single [14] or multiple [15] image observations, often constrained by limitations in interpolation and extrapolation distances. Novel generative view synthesis methods, employing neural volumetric representations [16], have demonstrated impressive results in generating diverse objects, faces, or indoor environments. However, these methods face challenges in generating unbounded outdoor scenes due to a lack of multiview data for supervision. Recent perpetual view generation methods, such as InfiniteNature [17] and InfiniteNature-Zero [18], excel in generating unbounded flythrough videos of natural scenes but lack globally consistent 3D scene content. Concurrent works like InfiniCity [19] and SceneDreamer [20] leverage birds-eye-view representations, while SceneScape [21] constructs a mesh representation from text.

Our work builds upon recent terrain generation methods exemplified by PersistentNature [22], and style encoder methods exemplified by pixel2style2pixel [23]. Our research introduces unique advantages in four key aspects:

- We offer a unique, first-person side view perspective of glaciers that is different from the traditional topdown view in satellite imagery.
- We incorporate real satellite data as a condition for the generative model to control the generated terrains.
- Our work focuses on glacier visualization, an area largely unexplored by previous terrain modeling studies.
- We curated a dataset of over 90K glacier images to address the current lack of readily available glacier datasets.

3. Methods

In this section, we provide a detailed overview of the dataset construction process and the training pipeline used in our study. To achieve our goal of generating glacier scenery from satellite image, we utilize DEM images from the USGS Benchmark Glacier dataset [5] and natural terrain images from the Landscape HQ dataset [4] to enable our GlacierPix2Pix model to find the correspondence between satellite images and glacier images. We divided this section into several parts to cover each step in our pipeline in detail. These parts include extracting side view contours from DEM images, curating a glacier image dataset, unifying the representation of both types of data, and constructing the GlacierPix2Pix model.

3.1. DEM to Side View

Figure 1. Extracting contours from Digital Elevation Models (DEM) is necessary since our generative model takes in contour as a condition. The value at each pixel of a DEM image represents the height of the terrain at that location.

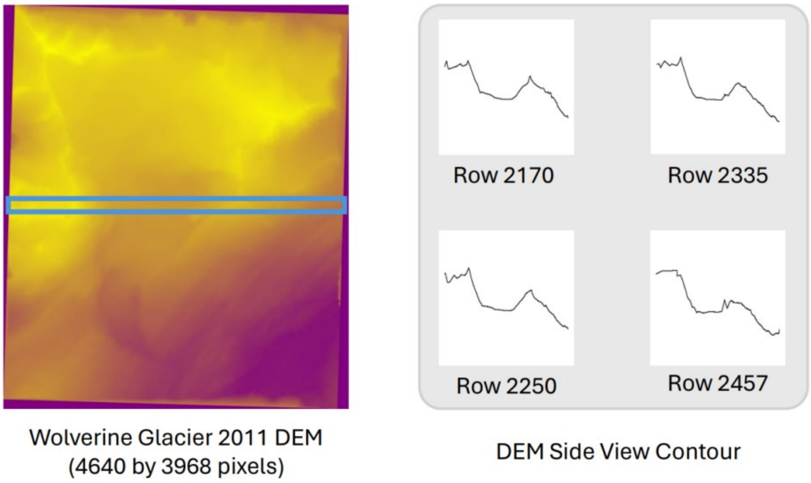

Figure 1. Example of contours extracted from a DEM image.

Therefore, to extract the contour along any chosen line on a DEM image:

- The x value of each point on the contour is the position along the chosen line.
- The y value of each point on the contour is the value at the position along the chosen line.

3.2. Glacier Scenery Dataset Curation

To support the data-driven nature of our model, we need a large number of glacier images. In this study, we use the Landscape HQ dataset [4] that contains 90,000 high-resolution natural landscape images collected from photo sites Unsplash and Flickr by the authors of this dataset. This dataset contains versions with different image resolutions. We chose to use the LHQ256 version, which contains images with 256 by 256 resolution.

Some LHQ images contain geometry that is not suitable for camera movements during training, such as a landscape pictured through a window or a closeup of trees. Therefore, we performed a filtering process on all 90K images to remove those unwanted images using a pretrained semantic segmentation model (DPT) [24], which resulted in 52,402 usable images.

Since the goal of this study is to achieve glacier visualization, we need to separate glacier images from other non-glacier images in this dataset. We explored two methods via the Contrastive Language-Image Pre-Training (CLIP) model [25]:

- CLIP Text-Image method
- CLIP Feature K-Nearest Neighbor method

In the CLIP Text-Image method, we input keywords for identifying glacier and non-glacier images into the CLIP model, such as "glacier", "snowy mountain", "green mountain", "lake", etc. Then, the CLIP model outputs the probability of the given image being each of the given keywords. Based on the probability, we decide whether each image is a glacier image using certain thresholds. In this case, we classify an image as a glacier image if its glacier probability is greater than 0.45 and its non-glacier probability is less than 0.15. In the CLIP Feature K-Nearest Neighbor method, we first use CLIP to obtain the image features; then, we use a K-Nearest Neighbor model to classify the image feature based on 10 images that we pre-labeled as glacier images. As shown in Figure 2, the results of both methods have unique images and overlap images. We decided to select the overlap images of both methods as our glacier dataset (9,697 glacier images in total).

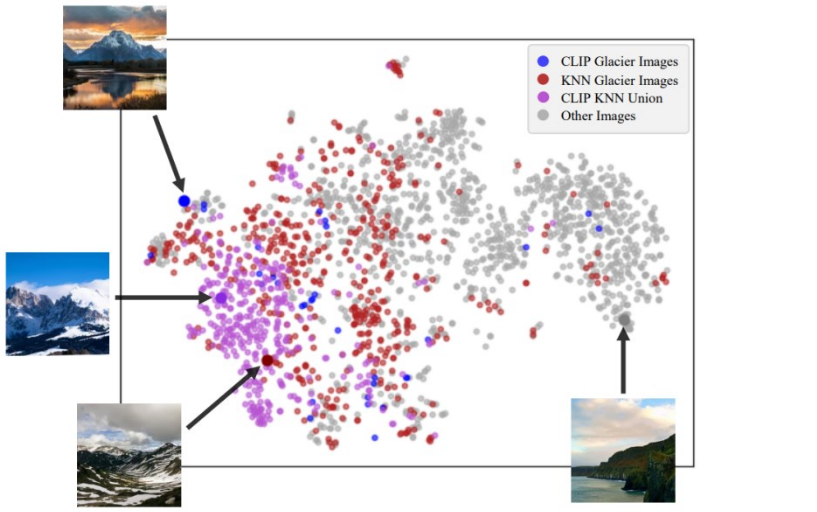

Figure 2. Both glacier classification methods have unique images and overlap images as represented above using TSNE. If two images are closer on the TSNE graph, their features are more similar.

3.3. Unify Representation

Terrain and DEM images are different data representations, thus they cannot be compared directly. Contour of terrain elevation is a common ground between them; therefore, contour extraction of glacier images is necessary to eliminate the domain gap. We experimented with 8 different edge filtering methods:

- Canny (with and without Gaussian Blur)
- Sobel X (with and without Gaussian Blur)
- Sobel Y (with and without Gaussian Blur)

- Sobel XY (with and without Gaussian Blur)

As shown in Figure 3, Canny with Gaussian Blur algorithm seems to be the best in this case due to the clear edges and the small amount of noise. We performed this algorithm on all glacier images to obtain their contours.

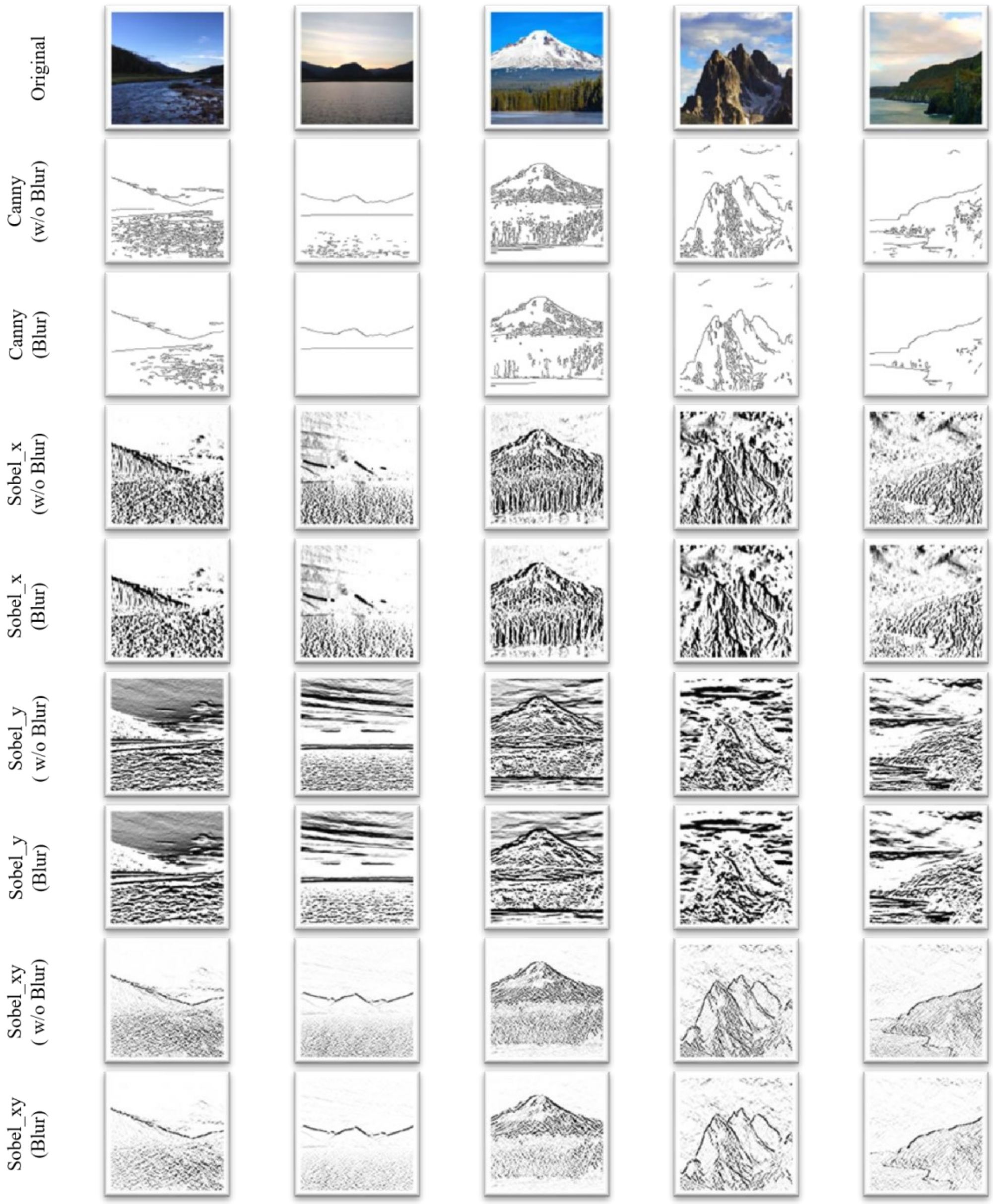

Figure 3. Results of 8 different contour extraction methods on 5 randomly selected terrain images.

3.4. GlacierPix2Pix Model

As shown in Figure 4, the GlacierPix2Pix model employs a StyleGAN [26] architecture comprising a generator and a discriminator. The StyleGAN architecture has two key advantages: it is able to generalize to most terrains since the latent space learns all the possible combinations of terrain in the input data, and it is able to achieve high-resolution image quality. The generator functions to produce 3D glacier terrains, while the discriminator evaluates the generated results against ground truth images, adjusting model weights accordingly. The first stage of training terminates when the discriminator can no longer discern between generated and training images.

Subsequently, we add a conditional style encoder [23] that matches the contours to the ground truth images. The encoder encodes the contours into style vectors, then those style vectors become parts of the input to the terrain layout generator. Through iterative training, the style encoder adjusts its weights to achieve optimal alignment between contour inputs and generated terrains. The second stage of training terminates when for each contour, the style encoder is able to generate the style vector that directly matches one glacier terrain generated by the StyleGAN.

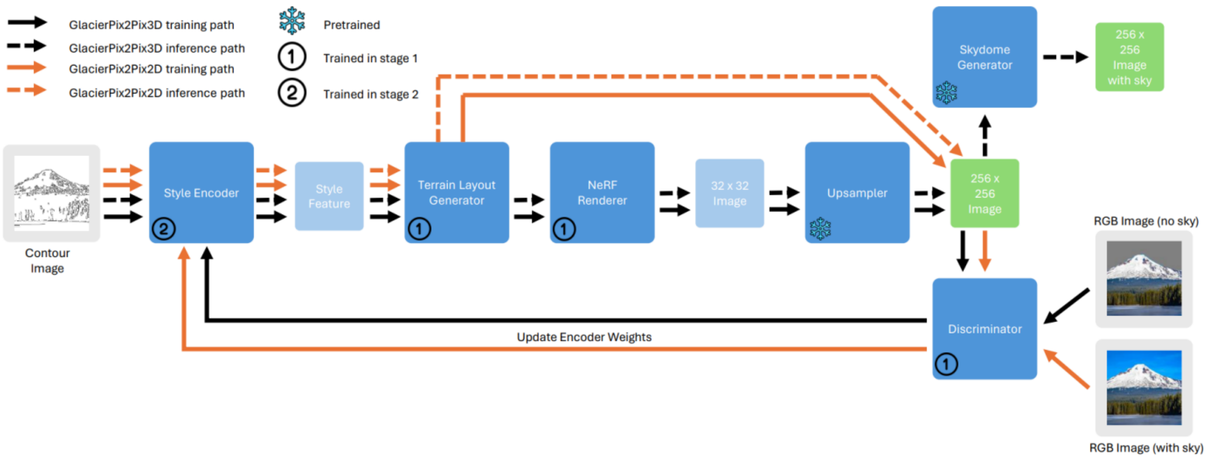

Figure 4. GlacierPix2Pix Model pipeline. Both 2D and 3D training and inference paths are shown.

4. Results

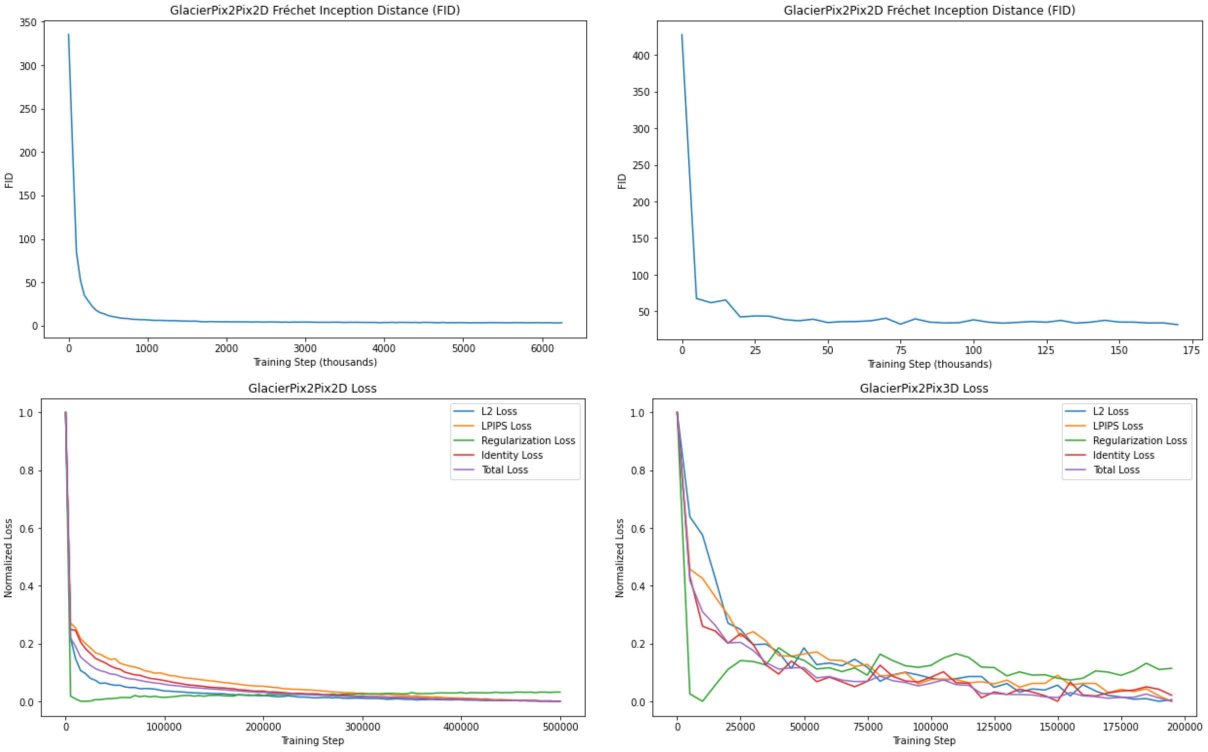

Figure 5. FID and loss metrics vs. training steps are shown above.

Figure 5. To measure the performance of the model, we use several metrics:

- Fréchet Inception Distance (FID): calculates the distance between feature vectors for images generated by GAN models
- Reconstruction Loss (L2): calculates the pixel difference between the generated images and ground truth images
- Perceptual Loss (LPIPS): calculates feature difference in the latent space between the generated and ground truth images
- Identity Loss: measures the cosine similarity between the features of the generated and ground truth images
- Regularization Loss: measures the distance between the feature vectors of the generated image and the average latent vector

Table 1. Performance compared to other data-driven methods

Model	FID
GSN [27]	45.48
EG3D [28]	32.08
PersistentNature [22]	32.02
GlacierPix2Pix	31.90

As shown in Table 1, the GlacierPix2Pix model achieves an FID value of 31.90, which surpasses the performance of other similar data-driven methods. As shown in Figure 6, the model can successfully produce photorealistic glacier images aligned with satellite-derived contours.

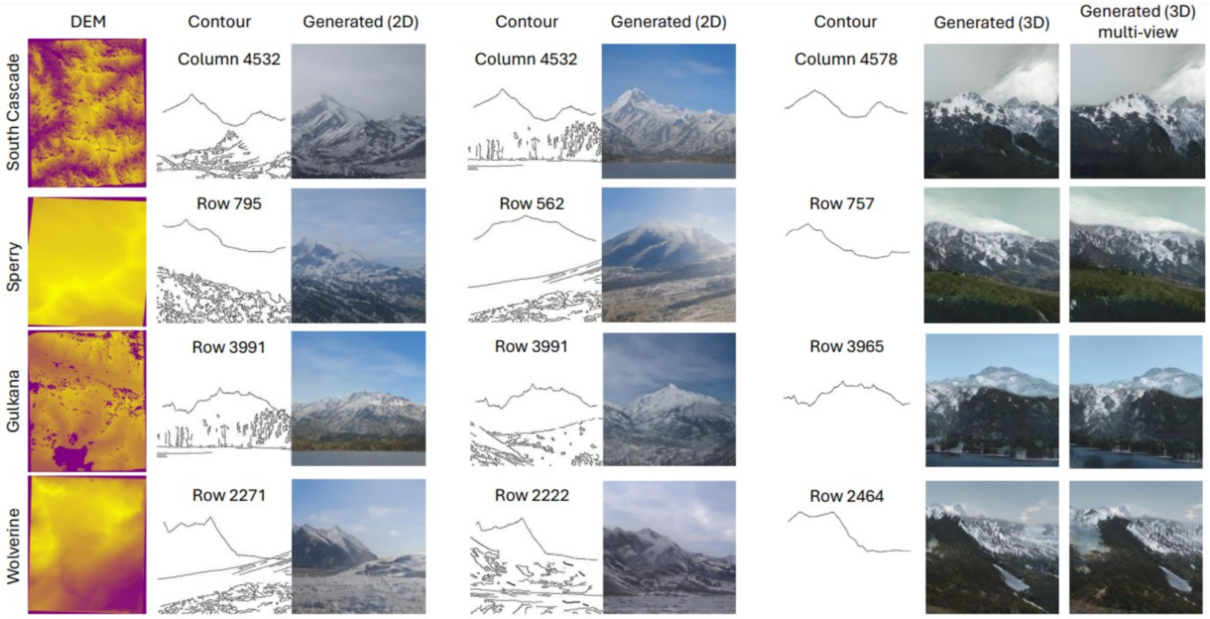

Figure 6. The GlacierPix2Pix model can successfully produce photorealistic glacier images aligned with satellite-derived contours. Shown above are examples of glacier visualizations of 4 different North American glaciers. On the left side are 2D pipeline results, and on the right side are 3D pipeline results. (see video for dynamic visualization)

5. Conclusion

This study proposes GlacierPix2Pix, a novel deep learning model that generates glacier scenery from satellite images. GlacierPix2Pix can successfully produce photorealistic glacier images aligned with satellite-derived contours and it achieves higher image quality than previous similar datadriven models.

It has the potential real-world applications of demonstrating glacier recession, helping with mapping snow cover and mass balance, and allowing detailed changes in remote locations to be observed.

This study also brings many potential benefits beyond the technical aspects. Climate change, a global concern, often faces challenges in conveying precise environmental changes to the public as concrete evidence. Our work addresses this by visually presenting the effects of climate change on different glaciers through photorealistic images by employing cutting-edge technology. This approach eliminates the need for on-site glacier photography, offering a cost-effective means of showcasing the exact changes in our environment. By bridging the gap between technological innovation and environmental awareness, this research contributes to a better understanding of the urgent issues surrounding climate change, especially its effect on glaciers.

6. Future Work

Potential future work includes:

1. Building an interactive website where users can upload their own contours, choose a style, and see the generated glaciers.
2. Extend this work to other terrain types, such as canyons, plateaus, volcanos, deserts, etc.
3. Refine GlacierPix2Pix to achieve even higher image quality.

References

[1] Zhiqin Chen. 2023. A Review of Deep Learning-Powered Mesh Reconstruction Methods. *arXiv preprint arXiv:2303.02879* (2023).

[2] Eric Galin, Eric Guérin, Adrien Peytavie, Guillaume Cordonnier, MariePaule Cani, Bedrich Benes, and James Gain. 2019. A review of digital terrain modeling. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 553–577.

[3] Johan WH Tangelder and Remco C Veltkamp. 2008. A survey of content based 3D shape retrieval methods. *Multimedia tools and applications* 39 (2008), 441–471.

[4] Ivan Skorokhodov, Grigorii Sotnikov, and Mohamed Elhoseiny. 2021. Aligning latent and image spaces to connect the unconnectable. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14144–14153.

[5] Christopher J McNeil, Caitlyn E Florentine, Valerie AL Bright, Mark J Fahey, Evan McCann Christopher F Larsen Evan, E Thoms, David E Shean Lisa A McKeon, Rod S March William Keller, Erin N Whorton, Shad O'Neel, et al. 2019. Geodetic data for USGS benchmark glaciers: Orthophotos, digital elevation models, glacier boundaries and surveyed positions. (2019).

[6] Biliana Kaneva, Josef Sivic, Antonio Torralba, Shai Avidan, and William T Freeman. 2010. Infinite images: Creating and exploring a large photorealistic virtual space. *Proc. IEEE* 98, 8 (2010), 1391–1407.

[7] James Hays and Alexei A Efros. 2007. Scene completion using millions of photographs. *ACM Transactions on Graphics (ToG)* 26, 3 (2007), 4–es.

[8] Richard Strong Bowen, Huiwen Chang, Charles Herrmann, Piotr Teterwak, Ce Liu, and Ramin Zabih. 2021. Oconet: Image extrapolation by object completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2307–2317.

[9] Ronghang Hu, Nikhila Ravi, Alexander C Berg, and Deepak Pathak. 2021. Worldsheet: Wrapping the world in a 3d sheet for view synthesis from a single image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12528–12537.

[10] Songwei Ge, Thomas Hayes, Harry Yang, Xi Yin, Guan Pang, David Jacobs, Jia-Bin Huang, and Devi Parikh. 2022. Long video generation with time-agnostic vqgan and time-sensitive transformer. In *European Conference on Computer Vision*. Springer, 102–118.

[11] Chelsea Finn, Ian Goodfellow, and Sergey Levine. 2016. Unsupervised learning for physical interaction through video prediction. *Advances in neural information processing systems* 29 (2016).

[12] Arun Mallya, Ting-Chun Wang, Karan Sapra, and Ming-Yu Liu. 2020. World-consistent video-to-video synthesis. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*. Springer, 359–378.

[13] Zekun Hao, Arun Mallya, Serge Belongie, and Ming-Yu Liu. 2021. Gancraft: Unsupervised 3d neural rendering of minecraft worlds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14072–14082.

[14] Wonbong Jang and Lourdes Agapito. 2021. Codenerf: Disentangled neural radiance fields for object categories. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12949–12958.

[15] John Flynn, Michael Broxton, Paul Debevec, Matthew DuVall, Graham Fyffe, Ryan Overbeck, Noah Snavely, and Richard Tucker. 2019. Deepview: View synthesis with learned gradient descent. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2367–2376.

[16] Jiatao Gu, Lingjie Liu, Peng Wang, and Christian Theobalt. 2021. Stylenerf: A style-based 3d-aware generator for high-resolution image synthesis. *arXiv preprint arXiv:2110.08985* (2021).

[17] Andrew Liu, Richard Tucker, Varun Jampani, Ameesh Makadia, Noah Snavely, and Angjoo Kanazawa. 2021. Infinite nature: Perpetual view generation of natural scenes from a single image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14458–14467.

[18] Zhengqi Li, Qianqian Wang, Noah Snavely, and Angjoo Kanazawa. 2022. Infinitenature-zero: Learning perpetual view generation of natural scenes from single images. In *European Conference on Computer Vision*. Springer, 515–534.

[19] Chieh Hubert Lin, Hsin-Ying Lee, Willi Menapace, Menglei Chai, Aliaksandr Siarohin, Ming-Hsuan Yang, and Sergey Tulyakov. 2023. Infinicity: Infinite-scale city synthesis. *arXiv preprint arXiv:2301.09637* (2023).

[20] Zhaoxi Chen, Guangcong Wang, and Ziwei Liu. 2023. Scenedreamer: Unbounded 3d scene generation from 2d image collections. *arXiv preprint arXiv:2302.01330* (2023).

[21] Rafail Fridman, Amit Abecasis, Yoni Kasten, and Tali Dekel. 2024. Scenescape: Text-driven consistent scene generation. *Advances in Neural Information Processing Systems* 36 (2024).

[22] Lucy Chai, Richard Tucker, Zhengqi Li, Phillip Isola, and Noah Snavely. 2023. Persistent Nature: A Generative Model of Unbounded 3D Worlds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20863–20874.

[23] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. 2021. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2287–2296.

[24] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. 2021. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*. 12179–12188.

[25] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[26] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4401–4410.

[27] Terrance DeVries, Miguel Angel Bautista, Nitish Srivastava, Graham W Taylor, and Joshua M Susskind. 2021. Unconstrained scene generation with locally conditioned radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14304–14313.

[28] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. 2022. Efficient

geometry-aware 3D generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16123–16133.