DEEP GENERATIVE FRAMEWORK FOR INTERACTIVE 3D TERRAIN AUTHORING AND MANIPULATION

Shanthika Naik, Aryamaan Jain, Avinash Sharma, K S Rajan

International Institute of Information Technology, Hyderabad

ABSTRACT

Automated generation and (user) authoring of realistic virtual terrain is most sought for by the multimedia applications like VR models and gaming. The most common representation adopted for terrain is *Digital Elevation Model* (DEM). In this paper, we propose a novel realistic terrain authoring framework powered by a combination of VAE and generative conditional GAN model. Our framework is an example-based method that attempts to overcome the limitations of existing methods by learning a latent space from a real-world terrain dataset. This latent space allows us to generate multiple variants of terrain from a single input as well as interpolate between terrains while keeping the generated terrains close to real-world data distribution. We also developed an interactive tool that lets the user generate diverse terrains with minimal inputs. We perform a thorough qualitative and quantitative analysis and provide a comparison with other SOTA methods.

Index Terms— Terrain Authoring, GAN, VAE

1. INTRODUCTION

Terrain modelling aims to create a digital representation of the real-world topography and is useful in both scientific applications of land surface processes like flooding, soil erosion, as well as virtual terrain rendering in graphics and computer vision applications. It is also most sought for by the multimedia applications like Virtual Reality (VR) models and gaming. The real-world terrains undergo a range of natural transformations such as erosion, weathering, and landslides over the years, leading to the formation of complex topographies such as hills, mountain ranges, canyons, plateaus, and plains. This makes the terrain generation and authoring a challenging task. Existing terrain authoring and modelling techniques have addressed some of these and can be broadly categorised as *procedural modelling, simulation method*, and *example-based methods* (refer to [1] for a detailed survey).

Recent advancements in deep learning have enabled us to learn diverse terrain features for tasks like terrain amplification [2], modifications [3], etc. In the context of deep learning-based automated terrain authoring, the literature is very sparse. One of the most relevant example-based terrain authoring methods referred to in this work as *TSynthNet* [4], trains a conditional Generative Adversarial Network (cGAN) on a large set of real-world terrain data to generate realistic virtual terrains from hand-drawn user inputs. However, they provide limited user control and generate a single terrain for a given input. Additionally, their method allows the use of either drawing of level-sets or ridge/valley strokes (with altitude cues). However, these two representations are complementary and jointly provide richer information for terrain authoring tasks.

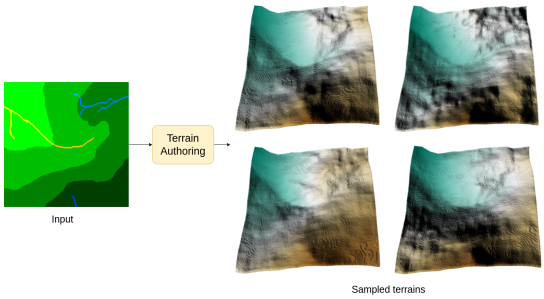

Fig. 1. Generation of multiple variants of terrain from a single input topographic map sketch.

In this paper, we propose a novel realistic terrain authoring framework powered by a combination of Variational Auto-encoder (VAE) [5], and conditional GAN model. Our framework attempt to overcome the limitations of TSynthNet by learning a latent space from a real-world terrain dataset. This latent space allows us to generate multiple variants of terrain from a single input (as depicted in Figure 1) as well as interpolate between terrains while keeping the generated terrains close to real-world data distribution. We also design a novel VAE loss function to exploit sparse topographic features like ridge/valley lines. Finally, we developed an interactive tool that lets the user model diverse terrains with minimal inputs. We perform a thorough qualitative and quantitative analysis and provide a comparison with TSynthNet to show the superiority of our method over SOTA methods.

2. METHODOLOGY

We propose a two-stage framework to learn the topographical structure of real-world terrains from existing datasets

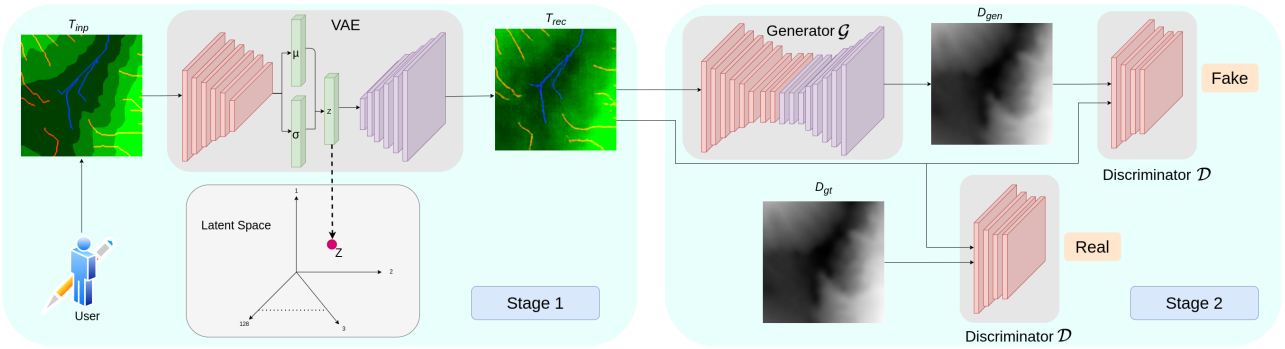

Fig. 2. The architecture of the proposed two-stage deep generative framework.

and generate plausible DEM from input sketches that can be thought of as *topographic maps* primarily consisting of DEM level-sets and ridge/valley lines representing underlying abstract topological features of the terrain. Learning such a latent space enables two key use-cases of the proposed method, namely, automated generation of multiple variants of terrain from a single user input sketch and interpolating between two terrains while keeping the generated virtual terrains closer to a real-world dataset consisting of realistic topographical features. Figure 2 provides an overview of the proposed two-stage framework.

2.1. Stage 1: Latent Space Learning

In the first stage, we aim to learn a generative latent space for topographic maps using a Variational Auto-encoder (VAE) model from a real-world terrain dataset. We extract ground truth topographic maps from real-world terrain data (see Section 3.1) and autoregress using VAE to learn the latent space. Let T_{inp} be our (hand-drawn) input sketches representing a rough topographic map. VAE learns to approximate a distribution $q(z)$ and learns the parameters μ and σ, from which the latent vector z is sampled using the re-parameterization trick as $z = \mu + \sigma * \epsilon$, where ϵ is sampled from a Standard Normal distribution. This sampled vector is fed to the decoder, which predicts T_{rec}, that is the reconstruction of original input T_{inp}.

We propose a novel auto-regressive reconstruction loss L_{recons} between VAE input T_{inp} and output T_{rec} by modifying the traditional Binary Cross Entropy (BCE) loss to emphasize the ridge/valley lines in the topographic map. We propose to give higher weightage to the loss on red and blue channels to give more importance to ridge and valley lines/strokes in the topographic map sketches. Additionally, a traditional KL divergence loss L_{KL} ensure that the probability distribution of latent vector z follows a Standard Normal distribution. Thus, the final VAE loss L_{VAE} is a combination of reconstruction L_{recons} and KL divergence loss L_{KL} .

$$L_{VAE} = L_{recons} + \gamma * L_{KL} \tag{1}$$

The γ parameter in Eq.1 is the weighting of the latent loss L_{KL} which is set to 0.65.

2.2. Stage 2: DEM Generation

The second stage consists of a conditional Generative Adversarial Network (cGAN) (Pix2pix [6]) model that generates plausible DEM output. This stage aims to generate the DEM given user topographic map sketch or, in our case, generated sketch from the previous stage.

The overall network is trained such that both generator G and discriminator D reach a Nash equilibrium by playing a two-player minimax game while optimizing the value functions $V(G, D)$[6]. We use L1 loss for reconstruction from the generator, i.e., $L1(\mathcal{G})$. So the final loss for cGAN training is given in Eq. 2.

$$L = \min_{\mathcal{G}} \max_{\mathcal{D}} \left[V(\mathcal{D}, \mathcal{G}) + L1(\mathcal{G})\right] \tag{2}$$

3. EXPERIMENT DETAILS

3.1. Dataset

We use a popular DEM dataset used by other relevant works in the literature, e.g., [7, 8] which is part of DEMs of mountain ranges named Pyrenees [9] and Tyrol [10], respectively. DEM patches with a resolution of 2m/pixel have been used as ground truth elevation maps. Original DEM tiles were split into 200x200 pixels. We randomly sample 3000 image patches for training and 878 image patches for testing. More details about the dataset can be referred from [8]. We prepare the training dataset by extracting the topographic map input sketches as RGB images from DEMs. Here the Green channel is dedicated to representing the elevation in the form of 4 level-sets while the Red(/Orange) and Blue channels are used to represent ridge and valley lines, respectively.

3.2. Implementation Details

Our VAE model is a 12 layer network with 6 layers in the encoder and 6 layers in the decoder. The latent space dimension

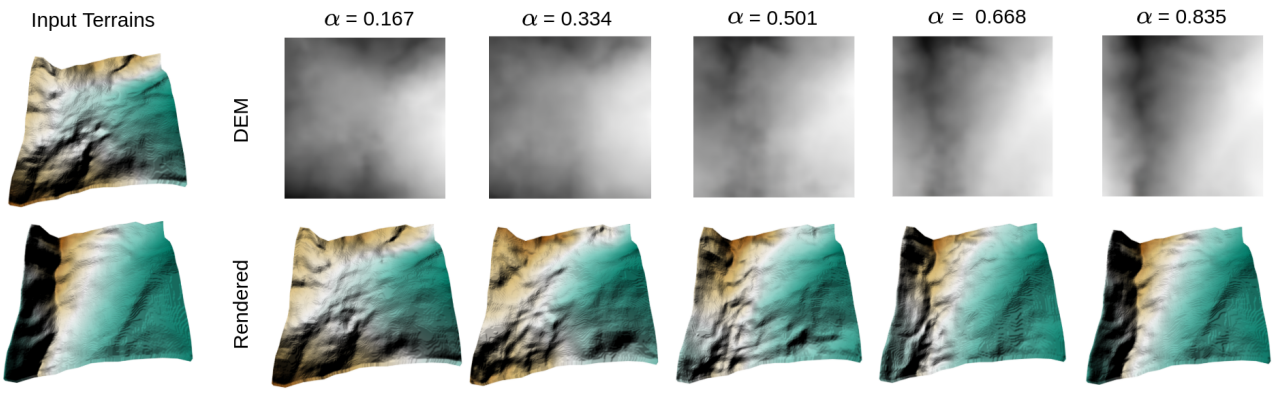

Fig. 3. Terrain interpolation results for varying α parameters.

is set to 128. All the layers consist of a 3x3 convolution with a stride of 2 and padding of 1, followed by Batch Normalization and using Leaky ReLU non-linearity. Adam optimizer was used to update the parameters with a learning rate of 0.001 and an exponential scheduler with gamma set to 0.95 while training the VAE.

Our conditional GAN generator is a U-net inspired Pix2pix architecture [6]. This model was trained using Adam optimizer with a learning rate of 0.0002, β_1 set to 0.5 and β_2 as 0.999 for both Generator and Discriminator. All our experiments were performed on a single Nvidia GTX 1080Ti.

3.3. Results

We provide quantitative evaluation results in Table 3.4. We can observe that we obtained superior performance with RMSE of 4.743 and PSNR of 34.189 from our model (VAE+cGAN) and beat the SOTA TSynthNet. We also demonstrate the results with the Baseline model (i.e., single-stage VAE based DEM generation) also performs inferior to our model, justifying need for a two-stage framework. Figure 4 shows the qualitative results where the first column shows the input and the VAE reconstructed topographic maps. The second column gives a comparison between the ground truth and generated DEMs. The last column shows the 3D rendering of these DEMs overlaid with the associated satellite image.

We also provide a qualitative comparison of our method with TSynthNet in Figure 5. The red circles depict the region where TSynthNet deviates from ground truth terrain while our method (green circles) stick closer to the ground truth. Additionally, our method also enables terrain interpolation and variant generation using the learnt VAE latent space.

Generating Terrain Variants: We utilise the latent space created by VAE to generate different samples from the same input. Different terrains generated from the latent space encoding of the same input topographic map are shown in Figure 1. We can observe the generated terrains have realistic but slightly different topographical features from that of input terrain. This provides the user with the flexibility to generate multiple terrain DEMs and use them for large scale generation of virtual terrain maps.

Terrain Interpolation: The latent space can also be used for automated fusion of topographic features across two terrains. Given two input DEMs we extract associated topographic map sketches and generate a new terrain by linear interpolation of the respective feature embedding (z_1 and z_2) in the VAE latent space. More specifically, we combine them using the formula $z = \alpha * z_1 + (1 - \alpha) * z_2$ while generating respective novel DEM using our framework. Figure 3 shows an example interpolation of two input terrains in the latent space for different values of α parameter.

3.4. User Study

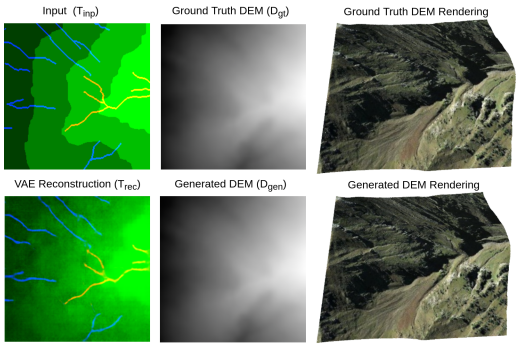

Fig. 4. Rendering of generated and ground truth terrains.

We performed a detailed user study involving 6 users. We presented users with a set of generated and ground truth terrain pairs overlaid with satellite images. The users were unable to decisively differentiate the generated and ground truth terrains and choose the real terrain only 50% of the time. To-

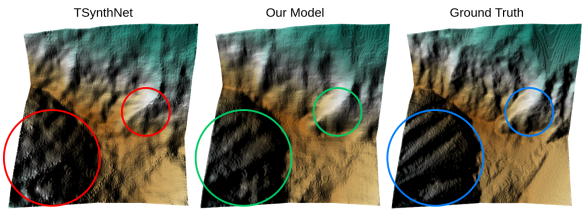

Fig. 5. Qualitative comparison with TSynthNet [4].

tal 83.3% of the users agreed that the terrains generated are very realistic, while 16.7% said that it is fairly realistic.

In the second experiment, we provide the user with a simple interface to draw input sketches. We provide the option to vary brush thickness so that the dense level-sets can be drawn with only a few strokes. The user interface and the DEM generated for a hand draw user input is shown in Figure 6. The input can also be interactively edited to get desired output. We asked the users several questions regarding the interface and the application. 33.3% users said that the input is very intuitive while 66.7% users agreed that it is fairly intuitive. 50% of the user strongly agree that the generated terrain follow the input sketches, while the remaining 50% fairly agree. All the users strongly agree that the system is fast and reactive. When asked to rate on a scale of 1 to 5, on how easy it was to express one's intent, 50% of the users gave a rating of 5, 16.7% gave a rating of 3, and 33.3% gave a rating of 2. We observed that the users were able to generate DEMs with ease after a couple of attempts.

Method	RMSE ↓	PSNR ↑
TSynthNet [4]	5.391	31.875
Baseline (VAE)	9.229	7.322
Our model (VAE+cGAN)	**4.743**	**34.189**

Table 1. Comparison with baseline and TSynthNet [4].

4. CONCLUSION

We proposed a novel realistic terrain authoring framework powered by a combination of VAE and conditional GAN model. Our framework learns a generative latent space from real world terrain dataset. This latent space allows us to generate multiple variants of terrain from a single input as well as interpolate between terrains, while keeping the generated terrains close to real world data distribution. While a preliminary interactive tool has been developed and used here, we further intend to provide user control to generate the terrain variants and interpolated terrains. The thorough qualitative and quantitative analysis and comparison with other SOTA methods support the superior outcome of our approach.

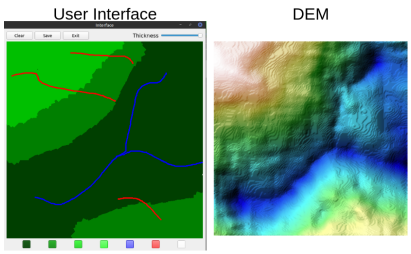

Fig. 6. The UI developed for user study.

5. REFERENCES

[1] E Galin, E Guérin, A Peytavie, G Cordonnier, Marie-Paule Cani, B Benes, and J Gain, "A review of digital terrain modeling," *Computer Graphics Forum*, vol. 38, no. 2, pp. 553–577, 2019.

[2] A Kubade, Dn Patel, A Sharma, and K. S. Rajan, "Afn: Attentional feedback network based 3d terrain super-resolution," in *Asian Conference on Computer Vision (ACCV)*, 2020.

[3] M Santini, S Grimaldi, F Nardi, A Petroselli, and M C Rulli, "Pre-processing algorithms and landslide modelling on remotely sensed dems," *Geomorphology*, vol. 113, no. 1-2, pp. 110–125, 2009.

[4] E Guérin, J Digne, E Galin, A Peytavie, C Wolf, B Benes, and B Martinez, "Interactive example-based terrain authoring with conditional generative adversarial networks," *ACM Transactions on Graphics (ToG)*, vol. 36, no. 6, 2017.

[5] D P Kingma and M Welling, "Auto-encoding variational bayes," 2014.

[6] P Isola, Jun-Yan Zhu, T Zhou, and A A Efros, "Image-to-image translation with conditional adversarial networks," *CoRR*, vol. abs/1611.07004, 2016.

[7] O Argudo, A Chica, and C Andujar, "Terrain super-resolution through aerial imagery and fully convolutional networks," in *Computer Graphics Forum*. Wiley Online Library, 2018, vol. 37, pp. 101–110.

[8] A Kubade, A Sharma, and K S Rajan, "Feedback neural network based super-resolution of dem for generating high fidelity features," in *IGARSS*, 2020, pp. 1671–1674.

[9] "Institut cartogràfic i geològic de catalunya (icc)," `http://www.icc.cat/vissir3`, Accessed: February 2, 2019.

[10] "Südtiroler bürgernetz geokatalog (sbg)," `http://geokatalog.buergernetz.bz.it/geokatalog`, Accessed: February 2, 2019.