

WorldGAN: 3D Environment Generation Using Deep Adversarial Networks

Tim Whitaker
Colorado State University

December 6, 2019

1 Abstract

We introduce a method and application for generating 3D environments with generative adversarial networks. The outputs from our networks are able to be imported into popular three dimensional modeling software like Unreal, Unity and Blender. We've created a simple and easy to use web based interface to allow a user to control the output of the generated environment by painting with predefined brushes that represent terrain varieties: mountains, hills, terraces, plains, and rivers.

2 Introduction

This project explores the usage of conditional generative adversarial networks to synthesize 3 dimensional environments.

The motivations for this project are to make it easier for game, world or environment designers to create realistic environments procedurally while giving the designer more control over the results.

We aim to create a simple and easy to use GUI interface for a user to paint a 2 dimensional semantically significant image map. We then pass the user created image map through a generative adversarial network trained on elevation data gathered from the Earth in order to generate a realistic terrain heightmap that approximately represents the user defined map. The resulting heightmap produced from the neural network can then be uploaded into any of the popular 3 dimensional rendering softwares in-

cluding Unity, Unreal, Blender or World Machine.

3 Background

This project explores the new paradigm of 3 dimensional procedurally generated content via machine learning.

3.1 Procedural Content Generation

Procedural content generation is an algorithmic and stochastic process used in art, games, music and film. As a paradigm for content generation, it is used to increase replayability, variability, flexibility and reduce production cost, effort storage space [4].

3.2 Generative Adversarial Networks

Neural networks have become a pervasive part of computer science. It allows for a powerful model of learning patterns from complex datasets. Generative adversarial networks work by synthesizing novel data according to patterns it learns in the training set. We evaluate two models of conditional generative adversarial networks to achieve image to image translation.

We are going to evaluate two implementations of a specialized form of generative adversarial networks designed to perform conditional image to image translation. The two models we are using are called Pix2Pix and CycleGAN.

Both Pix2Pix and CycleGAN are based on the deep

convolutional generative adversarial network architecture proposed by Radford et al. in 2015 [2].

3.2.1 Pix2Pix

3.2.2 CycleGAN

3.3 Related Work

Nvidia GauGAN.

4 Methodology

In this section we go over all of the pieces of work that went in to this project.

4.1 User Interface Application

The user interface is a web based javascript application that allows a user to paint a semantically significant image using predefined brushes. We include tools to change the brush size and export or clear the image. The canvas size is 512 x 512 pixels, and these images are resized automatically when passed to our neural network.

The backend is powered by a Flask API. We set up an endpoint to be able to evaluate a drawn image on our trained models.

4.2 The Dataset

Generative adversarial networks need a good amount of training data to output good results and a key objective of this project is to generate realistic results. This leads to an interesting challenge of putting together a realistic yet sizable dataset. In order to achieve this, we use terrain data of the earth as opposed to fictional or manufactured terrain data. In the next two sections we go into the methodology behind building the dataset.

4.2.1 Training Data

I used an interactive open source map tool that visualizes elevation data gathered from Mapzen's global elevation service [3].

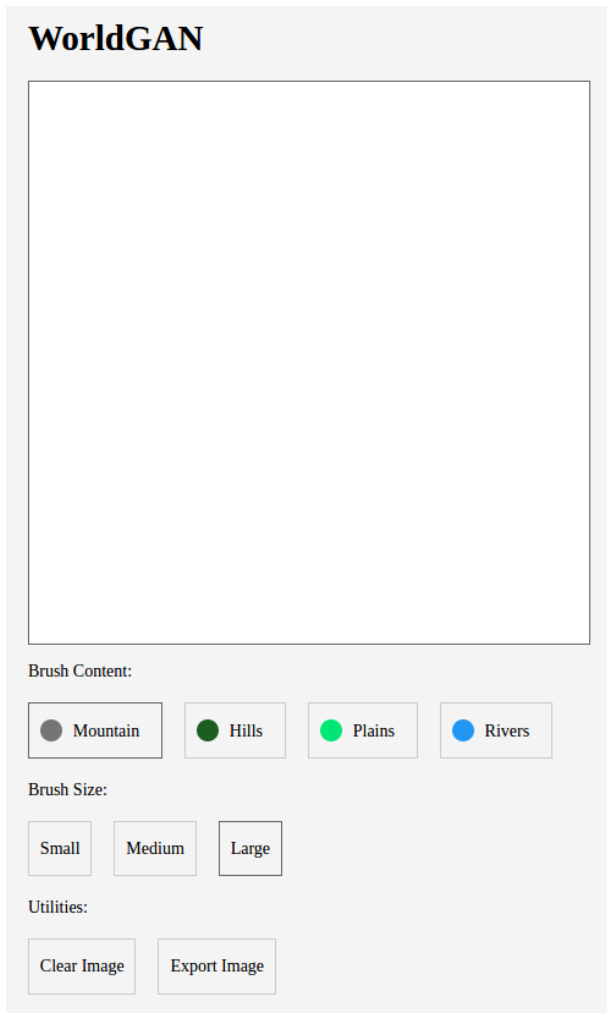


Figure 1: The user interface application

I then built a program that controls a headless browser using an open source library developed by Google called puppeteer. This script loads a random coordinate, checks if the coordinate is over land and if so, then we save an image of the elevation data at that coordinate. We resize the saved image to be 512 x 512 pixels.

We then started manually building the usermap training set using the WorldGAN user interface application. For each image in the heightmap training

set, We paint a usermap to represent that heightmap, using the semantically significant brushes that most closely match the terrain (mountains, hills, plains, rivers, and terraces). This process is a subjective one and all usermaps were painted by myself. This does bias the model to produce images according to how I read the heightmaps.

After this process we are left with 259 paired usermaps and heightmaps. We group all of these together into a batch to be trained on.

Creating the usermaps by hand is quite time consuming, and I wanted to experiment with models trained on a large dataset. We accomplish this by augmenting our first batch with simple transformations in order to artificially increase the number of images we can train on. Image augmentations are a well researched and effective approach to improving neural network performance [1].

We performed 5 operations on each usermap and heightmap. We mirror horizontally, flip vertically, rotate 90 degrees, rotate 180 degrees, and rotate 270 degrees. All of the resulting images are gathered into a second batch and we end up with a second batch of 1548 paired images (a 5x increase).

Pix2pix requires the training data from both the heightmaps and usermaps need to be paired. We stack each pair of training images horizontally and save this new image into our training batch. The resulting images are 1024 x 512 pixels.

4.2.2 Test Data

The test data is a collection of 5 usermaps I drew that are designed to be representative of a variety of environments a user might draw. These 5 environments include a mountainous region, a mountainous region fading into a hilly region, a hilly region fading into a flat region, a region with a main river and many sub rivers running through mountains, and a combined region containing mountains, hills, plains and a river.

4.3 Training The Models

We trained two models on this initial dataset, Pix2Pix and CycleGAN. GTX 1080ti GPU. i7 8700k.

processor.

4.3.1 Pix2Pix

1 hour training time. 4 hours training time.

4.3.2 CycleGAN

4 hours training time. 24 hours training time.

4.4 Visualizations Using 3D Software

Visualizations of results with 3d software.

5 Results

5.1 Visualizations Using 3D Software

6 Conclusion

6.1 Future Work

References

- [1] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning, 2017.
- [2] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2015.
- [3] Peter Richardson. Heightmapper, September 2016.
- [4] Adam Summerville, Sam Snodgrass, Matthew Guzdial, Christoffer Holmger, Amy K. Hoover, Aaron Isaksen, Andy Nealen, and Julian Togelius. Procedural content generation via machine learning (pcgml), 2017.