Anything You Can Do AI GAN Do Better: Generative Art

Keanu Spies
Stanford Univeristy
450 Serra Mall, Stanford, CA
keanuspies@stanford.edu

Abstract

Recently work has be done to apply Generative Adversarial Networks (GANs) to create AI generated artwork [2]. This combined with the recent work from Reed et al. that uses textual descriptions to create inputs, allows for better recognizable artwork from GANs [5]. In this work we hypothesize a fully connected automatic generation of artwork by two phases of generation: textual and imagery. With this we hope to create an architecture that is entirely automated and provides more believable artworks than that of standalone GANs.

1. Introduction

As we develop machines capable of conquering all tasks humans might, an interesting field lies still mostly unconquered - creativity. One such sub-field is the automatic generation of artwork. This project aims to further the field of automatic art generation to produce more believable and aesthetically pleasing artworks.

By using a Generative Adversarial Networks (GANs) I will train a Discriminator (D) and a Generator (G) set of neural networks which I will train concurrently to produce unseen and novel works of art.

While work has been done on creating artwork using GANs the results are not yet perfect. In this project I will be attempting to develop higher quality visual results with better conceptual meaning and more human-believable results.

2. Problem Statement

2.1. Dataset

I will use the WikiArt data set to train the Discriminator of the GAN. I will do this by separating out the data set into genres (abstract, animal paintings, city-scape, figurative, flower painting, genre painting, landscape, marina, mythological-painting, nude-painting, portrait, religious-painting, still-life, and symbolic painting) as per Tan *et al*. [4] .

Separation will allow for better performance, as well as smaller data-sub-sets, allowing for a quicker train time and therefore faster iteration process.

2.2. Expected Results

By collecting multiple methods of creating better images and adding in RNN-generated text the results should be able to outperform previous work done on GANs using the metric mentioned below in order to create visually appealing and closer to human-level artworks.

2.3. Evaluation Metrics

For art generation the best metric is that of the expert human eye, since we - as humans - developed the concept of aesthetic beauty and artwork quality. While the best, it is the least accessible metric, and requires alot of time and co-operation. Therefore some other metrics need to be used to approximate the experts eye.

I will be using the following metrics side-by-side to measure the outcomes of my GAN as opposed to that of its predecessors:

1. Inception Score: The inception score is the most commonly adopted metric for GAN performance, it uses conditional and marginal distributions over generated data or real data and is able to evaluate diversity and fidelity of samples. There is however evidence that shows that it might favor over fitting [1] the original data-set.

2. K-Nearest Neighbors (KNN): By using KNN we can determing if the output images are overfitting the original dataset. By using this as a good measure of over fit alongside the Inception Score [1], I will, hopefully be able to, balance overfit and good performance.

3. Technical Approach

I will also be experimenting on how input training images of larger pixel values (128x128) will allow for better performance. Jones and Bonafillia [2] found that quality increased drastically with larger image output, but requires significantly more compute.

This is not to mention that with such a broad latent space, many of the artworks fail to converge to meaningful outputs. This can be tackled by separating input artworks of specific genres [4]. This creates the issue of trying to find the balance between genre specific artwork and smaller image data-sets.

I will also be introducing a RNN based textual description generator with which we can use as conditional input to the GAN to attempt to produce better visual outputs.

4. Intermediate/Preliminary Results

4.1. Baseline Architecture

As a baseline I implemented the Deep Convolutional GAN (DC-GAN) as first introduced by Radford and Metz in 2016 [3]. This architecture uses multiple layers of convolutions within the Discriminator and Generator in order to create a more stable and much quicker learning turn-around.

Seeing as this architecture was used in almost all research papers on Art Generation it seems like the one I am attempting to surpass in this project's generation.

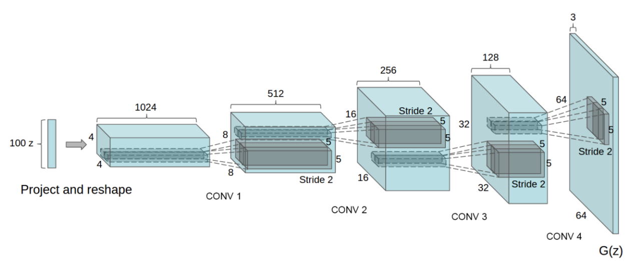

Figure 1. GAN Generator Architecture

The network also includes random noise within the iterative process as entailed in Radford and Metz [3] as well as used Leaky-ReLu activations within the Discriminator to promote stability of the network on top of the regular GAN architecture. I also included Batch-Normalization layers after every convolutional layer.

This, I hope, creates the best opponent in the field with which I can attempt to outperform.

4.2. Baseline Results

In attempt to evaluate performance of the network I initially trained the network on one of the smaller genres found in the WikiArt Data-set: **flower-paintings**, which contains 1800 paintings. Using 250 epochs and a batchsize of 64 the networks was able to train in 6 hours and produce the following results shown in figure 2.

I then attempted to use the larger dataset of **abstract-art** images which consist a NVIDIA Telsa GPU it was able to only complete 50 iterations to produce the results shown in figure 3.

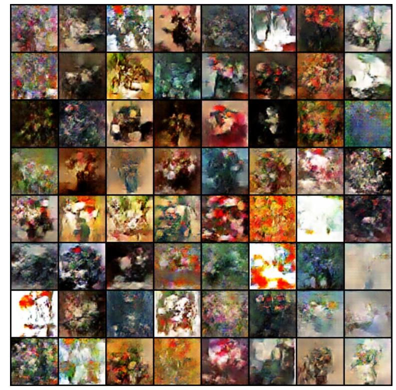

Figure 2. DC-GAN Generated Flowers after 250 epochs

Figure 3. DC-GAN Generated Abstract Art after 50 epochs

As a comparrison I will include the 50th epoch of the **flower-painting** trail to showcase how the size of the image set influenced the epoch-wise training speed of the network in figure 4.

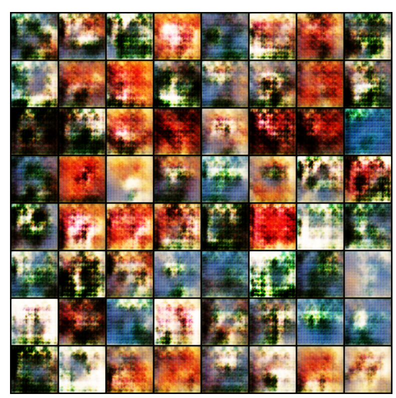

Figure 4. DC-GAN Generated Flower Art after 50 epochs

4.3. Evaluation

The DC-GAN was able to produce great artworks for a small subset of the images to the human eye which upon inspection include facets of many of the images in the data set but does not appear to be anyone of the images themselves. I have not yet produced the evaluation models (Inception score and KNN) but hope to in the near future.

4.4. Next Steps

Going into the final project I hope to move to implement my own network by tweaking the DC-GAN as well as applying a conditional initialization layer which will take textual input in order to create better results while still retaining the automatic aspect of the art generation.

I will also produce the evaluation metrics in order to determine if this change does well across the board further than the evaluation of my own eye.

References

[1] A. Borji. Pros and cons of gan evaluation measures, 2018. arXiv 1802.03446v1.

[2] K. Jones and D. Bonafilia. Gangogh: Creating art with gans, 2017. https://towardsdatascience.com/gangogh-creating-art-with-gans-8d087d8f74a1.

[3] A. Radford and L. Metz. Unsupervised representation learning with deep convolutional generative adversarial networks. 2016. arXiv 1511.06434v2.

[4] W. R. Tan, C. S. Chan, H. E. A. , and K. Tanaka. Artgan: Artwork synthesis with conditional categorical gans. 2017. arXiv 702.03410v2.

[5] J. Zhi. Pixelbrush art generation from text with gans. *CS231n*, 2017. http://cs231n.stanford.edu/reports/2017/pdfs/322.pdf.