Generative Adversarial Network and Image Processing

Zach Brill, Dan Kershner, Annan Miao, Troy Schwab

December 2018

Contents

1 Introduction

Art is classically viewed as a uniquely human endeavor, for art requires a level of original intentionality that only humans possess (Haugeland 2000). For many generations, the idea of art being created by something other than a human was never even considered, as consciousness was so narrowly restricted to us. However, with more and more inquiry into the nature of consciousness alongside the explosion of artificial intelligence, it is due time we begin to explore what it means for something to be art, and who can even create it.

This paper will introduce an approach to creating images and even art pieces using a generative adversarial network (GAN), an algorithm consisting of convolutional neural net structures. The program will take in a dataset of images, such as the MNIST or CIFAR10 built into keras as training data, and train and model the GAN to get output images.

The manipulation of custom image sets was not implemented in time, due to unusual mode collapse, though the functionality with CIFAR10 and MNIST sets acts as a proof of concept for the general model.

This report is structured as follows - the related work presents the background, history and applications of GAN. The implementation describes the goal, the overall approach and detailed implementation taken in this research, marked with a chronological steps. Finally, the paper will conclude with the ethical and philosophical implications, and our conclusions after such a project.

2 Related Work

In statistical classification, including machine learning, generative and discriminative modeling are two main approaches. Let x be the input data and y be the corresponding output. A discriminative model tries to directly learn the conditional probability distribution $P(y|x)$, while a generative model tries to learn the joint probability distribution $P(x, y)$. Additionally, generative models can use the joint distribution $P(x, y)$ to generate likely (x, y) samples.

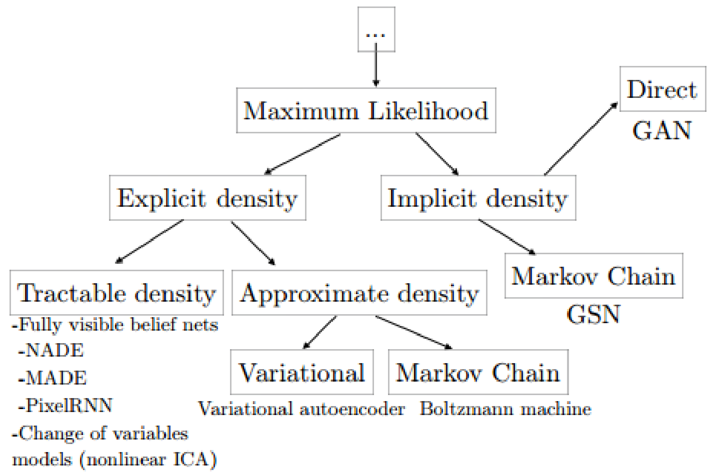

Figure 1: Different Generative Models

There are various approaches in generative modeling as shown in Figure 1. Fully visible belief networks (FVBNs) is one of the most popular approaches to generative modeling. It is a model that use the chain rule of probability to decompose a probability distribution over an n-dimensional vector x into a product of one-dimensional probability distributions. The most popular model in this family is PixelCNN. The main drawback with FVBNs is that samples must be generated one entry at a time: first x_1, then x_2, etc. These steps cannot be parallelized, so the rate of generating samples is slow.

Variational autoencoders (VAE) is another popular approach. It is a model that maximizes the lower bound on the log likelihood of the data. The main drawback of VAE is that the model is asymptotically consistent only if the distribution is perfect; otherwise, there would be a gap between the lower bound and actual density of the data. Also the quality of samples generated is relatively low.

GAN is also a popular approach of generative modeling. The idea of GAN started with Ian Goodfellow et al.'s landmark paper "Generative Adversarial Nets" (GANs), published at the Neural Information Processing Systems

(NIPS) conference in 2014. The idea is to propose a new framework for estimating generative models via an adversarial process, in which the program simultaneously trains two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability for D to make mistakes. This framework constructs a two-player minimax game.

GANs were designed to avoid many of the drawbacks associated with the above models. As opposed to Fully Visible Belief Networks, GANs use a latent code and thus can generate samples in parallel. Also unlike Variational Autoencoders, GANs are asymptotically consistent.

GANs have a wide range of applications. For instance, one can feed some text written with a particular handwriting as input to a generative model to generate more text in the same handwriting. Generative models can also be used in exploration in reinforcement learning where they can be used to generate artificial environments. Other applications include the conversion of sketches to images, image de-noising, conversion of low resolution images to high resolution, generation of art, and the conversion of satellite images to maps.

3 Implementation

GANs implicitly estimate a data distribution using the following minimax objective loss function:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$

Figure 2: minimax function for GAN

This loss function narratively describes a game between two opponents, a discriminator and a generator. The goal of discriminator is to judge whether that example was drawn from the true distribution $p_{data}(x)$, or from the generated fake distribution $G(z)$. Hence, proper optimization over the loss function in

Figure 2 results in an equilibrium state where the discriminator and generator are evenly matched and the discriminator cannot distinguish between real and fake generated samples, assigning a half probability to any sample it receives.

The goal of the research is to understand the theory and structure of GAN, implement a GAN from scratch and apply it to generate images and possibly art. The generator and discriminator are convolutional neural networks (because GANs are mainly used for image tasks). This network takes some random noise vector and outputs an image. When training the generative network, it learns which areas of the image to change so that the discriminator would have a harder time differentiating its generated images from the real ones. The generative network keeps producing images that are closer in appearance to the real images while the discriminative network is trying to determine the differences between real and fake images. The ultimate goal is to have a generative network that can produce images which are indistinguishable from the real ones.

The program uses *matplotlib* for plotting, *tqdm* to show a fancy progress bar for each epoch (iteration), and $tensorflow$ as the backend library for $Keras$, which is a high-level network to help develop and evaluate deep learning models. Both the generator and discriminator networks consist of a neural network with three hidden layers with the activation function of Leaky ReLU. The discriminator uses dropout layers to improve robustness on unseen images.

4 Ethics

The incorrect usage of this powerful network can create very extreme ethical dilemmas. Most noticeably, the open source project of Robbie Barrat, an artist who works with artificial intelligence. He developed a GAN model to produce abstract portraits around 2015, which was used by the French art collective called Obvious. This painting was then sold for nearly half a million dollars, excluding Barrat, and Obvious claiming they used his code as inspiration and modified it slightly. In Figure 3 we can see the painting which Obvious made, and in Figure 4, we see Barrat's paintings from his algorithm.

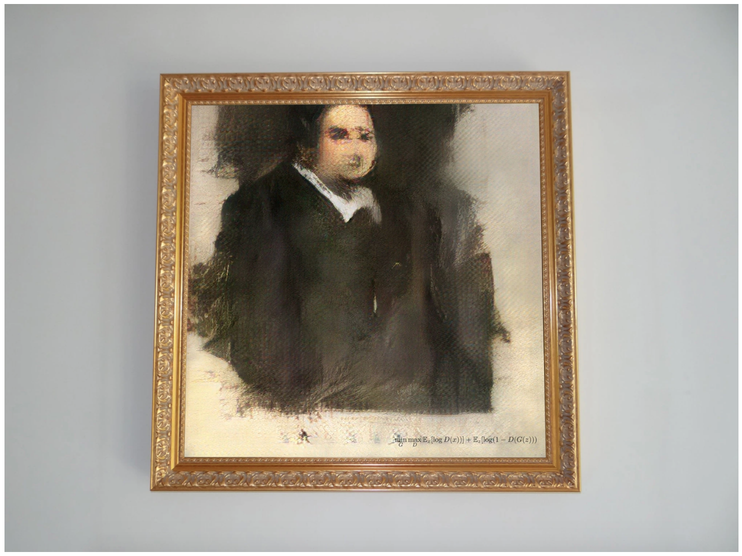

Figure 3: Obvious's Painting using Barrat's Code (Cohn 2018)

Figure 4: Some of Barrat's Portraits (Vincent 2018)

4.1 GAN Art Ownership

Extending from this controversy, we need to consider if the art belongs to anyone in particular? The creator of the algorithm? The creator of the specific code? Perhaps the art belongs to all the artists that were enveloped in the training data. Following this consideration, does the fact that the code is open source affect the ownership? Beyond that, is it even possible to ethically sell one of these works with these questions in mind? With these many questions, we will dilute the possibilities by setting a foundation of yes, the algorithm and code was stolen. The main question we attempt to now answer focuses on the rights of open source algorithm development, and the profiting of from use of that open source algorithm. In kantian ethics, one may be said to have acted morally if the rules used to do so may be considered a moral law. In this situation, the theft and profit of another's work, even though it was offered freely, is wrong, as it was only offered freely on platforms like github to aid other artists in their experimentation and research. Therefore, from the perspective of a kantian approach, it is clear that this is unethical.

Though, let's assume Barrat was never aware of this utilization of his software. He then is never hurt by the profits that Obvious made, and the software publicly is still maintained and aiding other artists with experimentation and research. His well-being therefore does not decrease at all, while Obvious, who receives fame and economic profits experience an increase of well-being. This act then would increase the overall well-being of all parties involved. Thus, from the standpoint of a utilitarian analysis, it is fair to say that no immoral act occurred and this was perfectly ethical. The main distinction in both of these analyses is that the utilitarian analysis equates happiness (usually offered as the metric in utilitarian analysis) to well-being and also assumes ignorance from Barrat. Therefore, it is fair to say that the kantian approach is more effective in analyzing this dilemma and that Obvious's utilization of Barrat's code was unethical.

5 Philosophy

With this model now creating gorgeous art works, we must consider philosophical consequences of AI generated art. Who is the artist in these cases? Can this be thought of as a creative endeavor? Does this generation of works function within the art world? Barrat relates his role often to the artist Sol Lewitt (Bailey 2018), who would write out instructions or rule sets for creating art that others would execute. Barrat simply programs the machine, offering it boundaries in how it operates and then provides the training set to see what it does. It can be thought of synonymous to a person who has only ever seen Monet paintings grabbing acrylics and copying the gradients of colors. In this case, the person simply learns from Monet, and is bound by the canvas as well as acrylics and knowledge of the color gradients. Thus, the role of Barrat can be thought of almost as a conductor, where the performing role belongs to the GAN. The GAN is not independent as it is simply a tool, identical in utility to markov chains in poetry, or random granular synthesis in music. It is up to the conductor in how to manipulate (or not) the output to some other form.

5.1 Intentionality

Framing the GAN as a tool removes the questioning of its creativity, but for the sake of discussion, we will consider the GAN as the artist for a brief moment. In this context, the GAN may be said to have creativity, but only in a short term perspective. This being a result of only a basic "grasp" of the color gradients. There may be no innovation over time that a human artist experiences that requires true experimentation and progression. Only information provided by the training set is knowledge that the GAN has and can have. It can never grow beyond the bounds of that set. This clearly demonstrates original versus derivative intentionality, as the intent of the colors chosen pixel by pixel is derived from the knowledge gained from the training set, though the model may never offer an original intent in a painting. Similar to Searle's Chinese Room Argument (Haugeland 2000), our GAN is restrained to the bounds of the room (being the training set) and the forms of the letters (being the pixels), but at no point in time, may the GAN alter those restrictions, or even understand the meaning of them. Returning to our prior conclusion that a GAN is simply a tool, how does this fit in the art world and coincide with art (and aesthetic) theory? W. E. Kennick has held less strict rules on that which is or is not art (Kennick 1964). He likens art to being something which is considered art, as such art is not art without the audience or the creator who would need to have the perception of it. Though with the interaction provided by the developer creating the GAN and making the images, would be enough to qualify an audience member, and as such the work as a normal piece of art despite the means by which it was made.

6 Conclusion

6.1 What Worked

Our project was a general success; our implementation worked properly and we found interesting results from the input image sets while learning how to properly utilize libraries to their potential to minimize our work. We also encountered some fun and interesting generated images when we trained the discriminator only on images of birds, for example. When we used this approach, our results were quite unique around 10 epochs into training. This is because the generator was attempting to produce images of birds, but in practice it was generating images that mostly had some indication of a blue sky and a winged object. To the untrained eye, if these were enlarged and clarified they could have been mistaken as abstract paintings of a sort. This was about as close as we came to generating something that we would consider rudimentary art.

6.2 What Didn't

The difficulty came when attempting to implement our own training image sets, due to the fact that unexpected mode collapse had occurred. We deduce that this is a fault in our image manipulation prior to analysis by the GAN, as our model functions properly on the built in image sets. We also encountered extreme PC slowdown once we began manipulating our own data set (a large image set of photos of flowers.) We were able to resize all of the images to be the same dimensions after which we converted each image to a numpy array as this enables easy training in Keras. However it was after we began to run our GAN on this data set that we encountered what is likely hardware problems. One member even had to reinstall their operating system due to some extreme overworking of their harddrive that persisted even after multiple restarts.

6.3 Parting Words

It was particularly interesting and ultimately the most rewarding to analyze the ethics and philosophy of our project, especially noting the events between Robbie Barrat and Obvious. The philosophy questions more broadly considered the ripple effect that computer science and artificial intelligence may cause in external fields. Though, with heavy application of our in-class readings derived from Haugeland's Mind Design II as well as articles from different sources online, we find ourselves excited to see computer science become implemented more often and to grander extent within external (particularly more creative humanities driven) fields.

References

[1] Haugeland, J. (2000). *Mind design II: Philosophy, Psychology, Artificial Intelligence.* Cambridge (Massachusetts): MIT.

[2] Cohn, G. (2018, October 25). AI Art at Christie's Sells for $432,500. Retrieved from https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html

[3] Vincent, J. (2018, October 23). How three French students used borrowed code to put the first AI portrait in Christie's. Retrieved from https://www.theverge.com/2018/10/23/18013190/ai-art-portrait-auction-christies-belamy-obvious-robbie-barrat-gans

[4] Karras, T, et al. (2018). *Progressive Growing of GANs for Improved Quality, Stability, and Variation.* Santa Clara (California): NVIDIA.

[5] Huang, H, et al. (2018). *An Introduction to Image Synthesis with Generative Adversarial Nets.*

[6] Bailey, J. (2018, April 5). AI Art Just Got Awesome. Retrieved from https://www.artnome.com/news/2018/3/29/ai-art-just-got-awesome

[7] Zuo, Y, et al. (2018). *Generative Adversarial Forests for Better Conditioned Adversarial Learning.*

[8] Goodfellow, I, et al. (2014). *Generative Adversarial Nets*

[9] Amos, B. (2016). *Image Completion with Deep Learning in TensorFlow.* Retrieved from http://bamos.github.io/2016/08/09/deep-completion/ml-heavy-generative-adversarial-net-gan-building-blocks

[10] Kennick, W.E. (1964). *Theories of Art and the Artworld: Comments.* Journal of Philosophy, Inc. Retrieved from https://www.jstor.org/stable/2022938