ART AND MACHINE LEARNING
CMU 2019 SPRING
PROJECT 2

A Happy Family!

Josh Moavenzadeh
Zach Saffran
James Gualtieri
Umang Bhatt
Griffin Tang

Description

GANs have been declared one of the most interesting ideas proposed in the last 10 years in machine learning by top artificial intelligence researchers for their wide application to tasks relating to the generation of images, speech, text, and music, among other domains. They can mimic the distribution of high dimensional data to create original content within a target domain, however, they are limited by many pitfalls, including lack of interpretability, convergence problems, and the necessity of a robust dataset. Art datasets, regardless of whether the pieces share medias, styles, or content, lack concrete structure, which makes applying GANs to the creative domain particularly challenging. We experimented with multiple network architectures and objective functions, using a variety of art datasets with different media types, styles, and subject matter, in order to gain insight into the challenges and best configurations for applying generative models to the creation of original art pieces.

Concept

We aimed to use GANs to understand and visualize the machine translation of the human form. Through this, we wanted to test the robustness of different GAN techniques and architectures to produce abstract and experimental artificial paintings containing human features to achieve a unique visual interpretation of the human figure.

Technique

GANs include two components - a discriminator that aims to distinguish real from fake images, and a generator that creates fake content with the intention of tricking the discriminator into thinking it is real. The generator uses random noise as a seed for creating images, while the discriminator analyzes images and outputs a score for how real or fake it perceives the image to be.

We adapted a Deep Convolutional GAN architecture with binary cross entropy loss on both the generator and discriminator as our baseline. Many modern GANs build on top of the DCGAN architecture, introducing new objective functions and training techniques to avoid the common pitfalls of training an adversarial network. We experimented with many of these modifications to see their impact on the output images, and combined the successful approaches in a unique way to create our final result.

Initially, we chose datasets of modern art while we experimented with the process of using GANs to produce new images and identify a baseline. These initial experiments were performed on watercolor images, oil paintings, and minimalist art pieces. These results were abstract and messy which we attribute to the variability in the datasets, so we switched over to a renaissance period art dataset of nude portraits, inspired by some successful pieces from Robbie Barrat. This allowed us to experiment with different techniques to see what resulted in more human-like forms. After experimentation, we reapplied the best techniques with the original watercolor dataset to produce our final results.

Process:

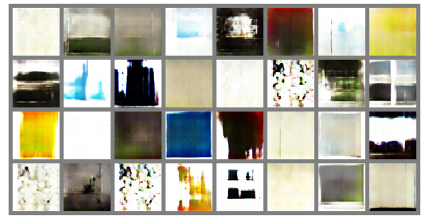
Our baseline model was a DCGAN with binary cross entropy loss trained on oil paintings of faces as well as minimalist art pieces. The results were very abstract. While they shared the colors and some of the style of the target pieces, it is hard to identify any structure to the pieces. We also noticed while training that the loss values were inconsistent and appeared somewhat arbitrary, since it did not follow the traditional downward trend and lower loss did not correlate with better results. This led us to experiment with other loss functions in order to get results that were more stable and interpretable.

We tried MSE loss on the same model, which is the approach proposed by LS-GAN (least-squares GAN), this time on a dataset of nude paintings which we reused throughout most of our experimentation. This resulted in marginally better results, but still suffered from very similar issues. We concluded that losses simply on the binary output of the discriminator are difficult to work with, so we researched alternatives.

The next step we took was applying feature matching, which is the idea of using a difference metric between real and fake image embeddings as a replacement for the generator loss. This required modifying the discriminator model to include an explicit feature embedding layer. The intuition around this approach is giving the generator the objective of producing natural looking images rather than strictly aiming to fool the discriminator. One downside was that the generator loss became small very quickly, since with a poor discriminator that produces poor embeddings, it became easy for the generator to match these embeddings. Instead, we decided to use feature matching embedding loss as an additional penalty on top of binary cross entropy loss for the generator. We found that this feature matching penalty appeared to help bring out more human figures in the produced images.

Some other attempts included Wasserstein and Energy-Based GAN, both of which struggled to produce any realistic images.

Baseline results on watercolor people

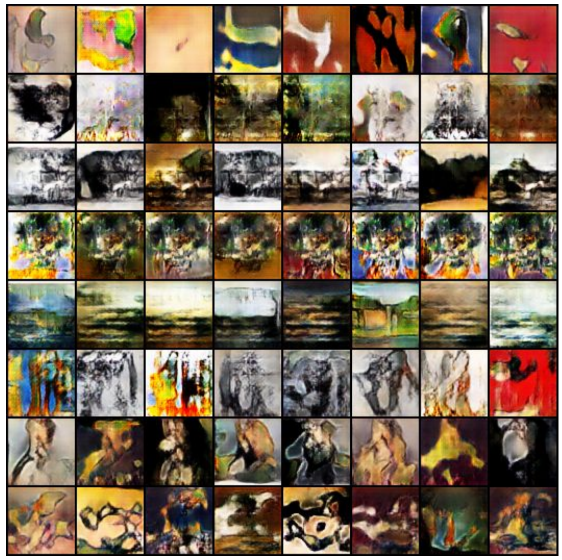


Baseline results on minimalist images

Nude paintings without feature matching

Nude paintings with feature matching

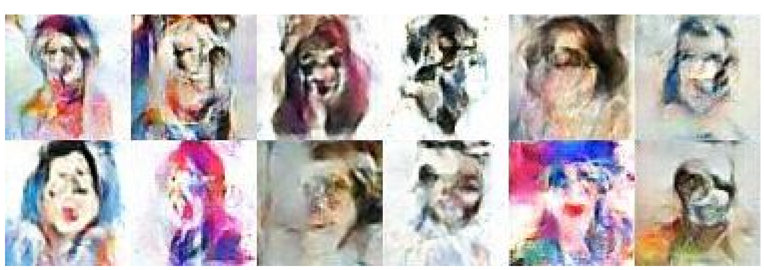
After these experiments, we began to explore the concept of conditional and classifier GANs, which leverage more data from multiple class labels to provide more useful information to both the discriminator and generator. The generator is fed a class label in addition to the random noise, and the discriminator learns to distinguish between the different classes of images. We took the ideas from the AC-GAN (auxiliary classifier GAN) and added a classification layer along with the validity scoring layer in the discriminator, so rather than just predicting real or fake, the discriminator would also try to predict what class the image belongs to. We hoped this would strengthen the embeddings that the discriminator learns, while providing information to the generator about how well it produces images of the target class. Our loss was a combination of the feature matching penalty on embeddings, binary cross entropy on validity scores, and softmax cross entropy on classification results. Based on the results, we can see that the generator is able to distinguish between classes reasonably well. The flowers have an identifiable silhouette, while the landscapes have an identifiable horizon. Within classes, however, we can see that the produced images all appear mostly similar, which is a common convergence issue in GANs known as mode collapse. It would have been interesting to have the time and resources to continue exploring these techniques, but due to the enormous amount of time it takes to train given increased model complexity and a substantially larger dataset, we were unable to explore this direction fully.



Classifier GAN - by row top to bottom the target classes were:
abstract, animal-painting, cityscape, flower-painting, landscape, nude-painting, portrait, still-life

Result

Our final image collage is the result of applying the feature matching penalty on binary cross entropy loss with our modified architecture of the DCGAN on the watercolor faces dataset. We were attracted to this model's results because it yielded exactly what we wanted our concept to show us: the machine interpretation of the human form. Even though each image is abstract and has its own style, the algorithm was still able to pull out explicit features of the face, such as roundness of the face, hair, lips, nose, and eyes. Some of the styles are vibrant and fun while others are quite grotesque. The images share many of the qualities of watercolor pieces including the fluid texture and blended colors.



Final result: feature matching with watercolor faces

Reflection

Overall, the results of this project showed the group that GANs are incredibly finicky. Generally we found that tracking the loss values gave very little insight into the performance of the models, and trying to use more interpretable loss functions did not always yield better results, as we saw with Wasserstein and Energy-Based loss functions. In order to produce non-abstract art, a consistent dataset of similar images is imperative. The distribution of art datasets is inherently noisey, so producing original but still concrete art was very difficult.

Reference

https://arxiv.org/abs/1712.01026
https://arxiv.org/abs/1609.03126
https://towardsdatascience.com/gan-ways-to-improve-gan-performance-acf37f9f59b

Code

https://github.com/saffranzachary/aaml_project2

Results

https://github.com/saffranzachary/aaml_project2/tree/master/Results