Generative Adversarial Networks for Genre-Based Art Generation

Laura Miron
Computer Science Masters Program, Stanford University

Introduction / Related Work

- Now that CNNs have achieved human-comparable accuracy at image classification tasks, image *generation* is the next big frontier in visual machine learning
- Along with PixelCNN, the most promising results so far have come from generative adversarial networks, or GANs
- Since GANs were introduced in 2014 by Goodfellow *et al.,* many variants have been proposed that, for certain tasks, stabilize training or improve results
- However, the majority of analysis has been done on the generation of natural images
- Generating artwork has many potential applications in animation and video game world generation, but presents a greater challenge than natural image generation for several reasons
    1. Scarcity of labeled datasets of artwork
    2. Widely varied styles of artwork
    3. Abstract / representational nature of some works
    4. Difficulty of distinguishing foreground and background
- Here, I focus on three architectures of GAN: VanillaGAN, DCGAN, and conditional DCGAN, and analyze their ability to generate artwork selected from (or conditioned on) six specific genres: abstract, cityscape, flower, landscape, portrait, and religious art

Data Collection and Processing

Original | Linear Interpolation | Area Interpolation

Figure 1. Effects of different downsamplng interpolation on input images

- 55,734 images taken from Kaggle[2] dataset
- Of 20+ genres in full dataset, 6 were selected on basis of containing sufficiently many examples, and representing easily distinguishable genre from others selected: 15978 abstract, 7031 cityscape, 2147 flower, 19296 landscape, 21571 portrait, and 9020 religious art
- Images reshaped without crop to square and downsize to 64x64 px using cv2's area interpolation
    - Much better results with area interpolation downsizing than linear interpolation

Inceptions Scores and Architecture Comparison

Inception Scores	
VanillaGAN	0.732
DCGAN	2.32
cDCGAN	1.11

Table 1. inception Scores

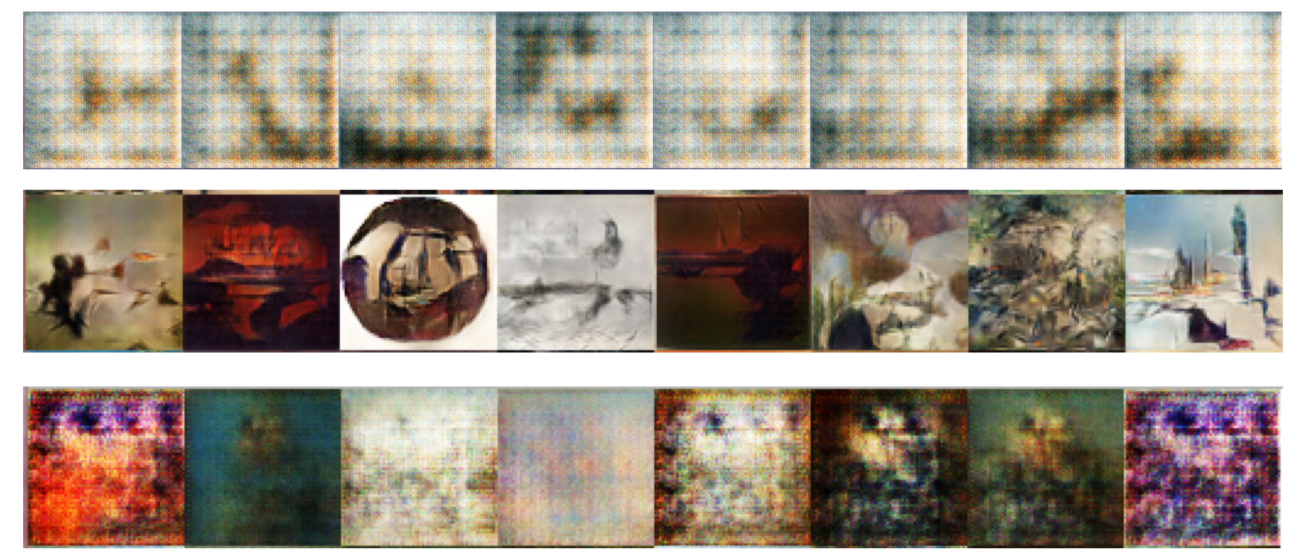

Figure 2. Results after 30 epochs of VanillaGAN (top), DCGAN (middle), conditional DCGAN (bottom)

- Since DCGAN produced significantly better results than VanillaGAN and conditional DCGAN, even after parameter turning, my remaining report focuses on DCGAN results and methods

DCGAN Samples

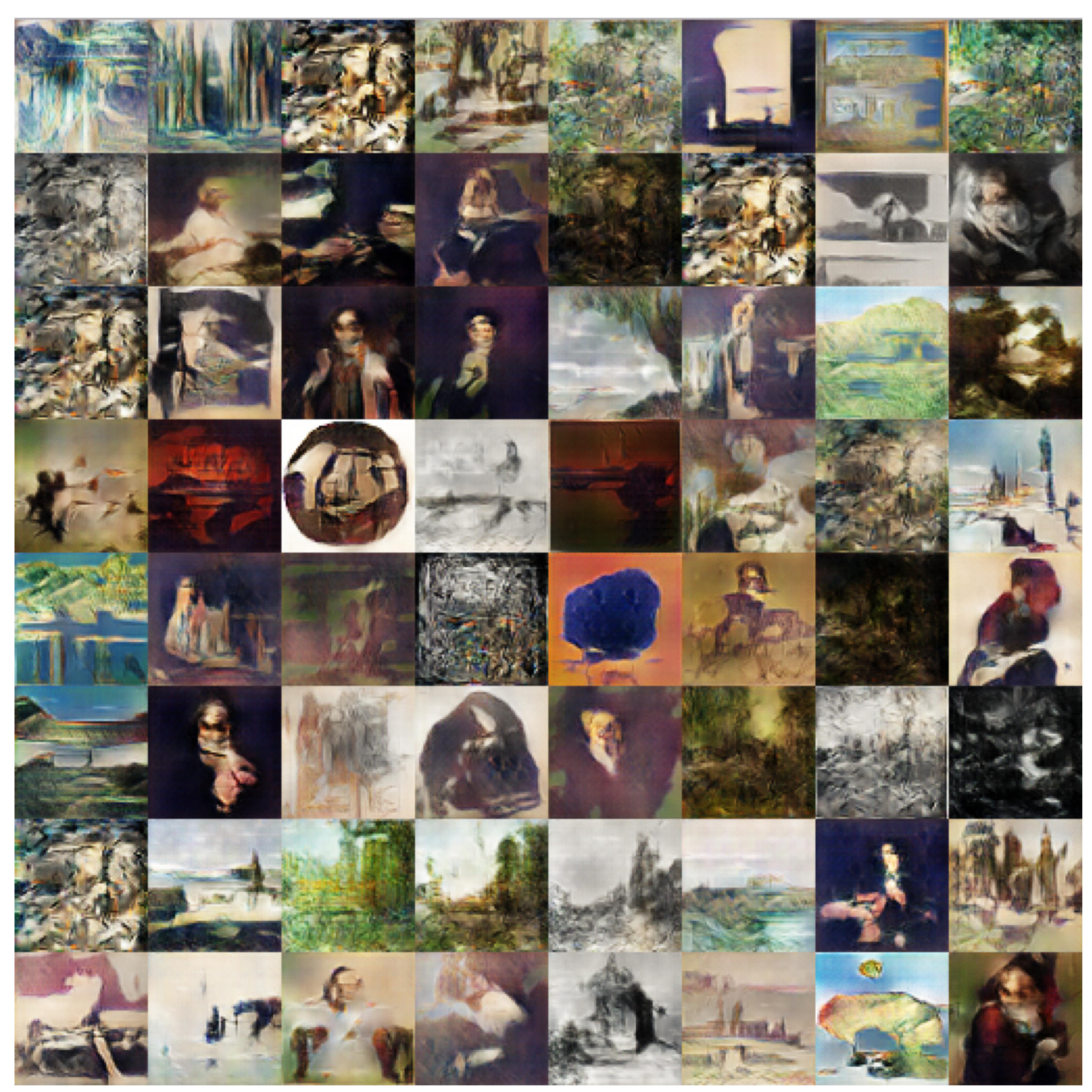

Figure 3. Results after running DCGAN on full dataset of 6 genres for 50 epochs

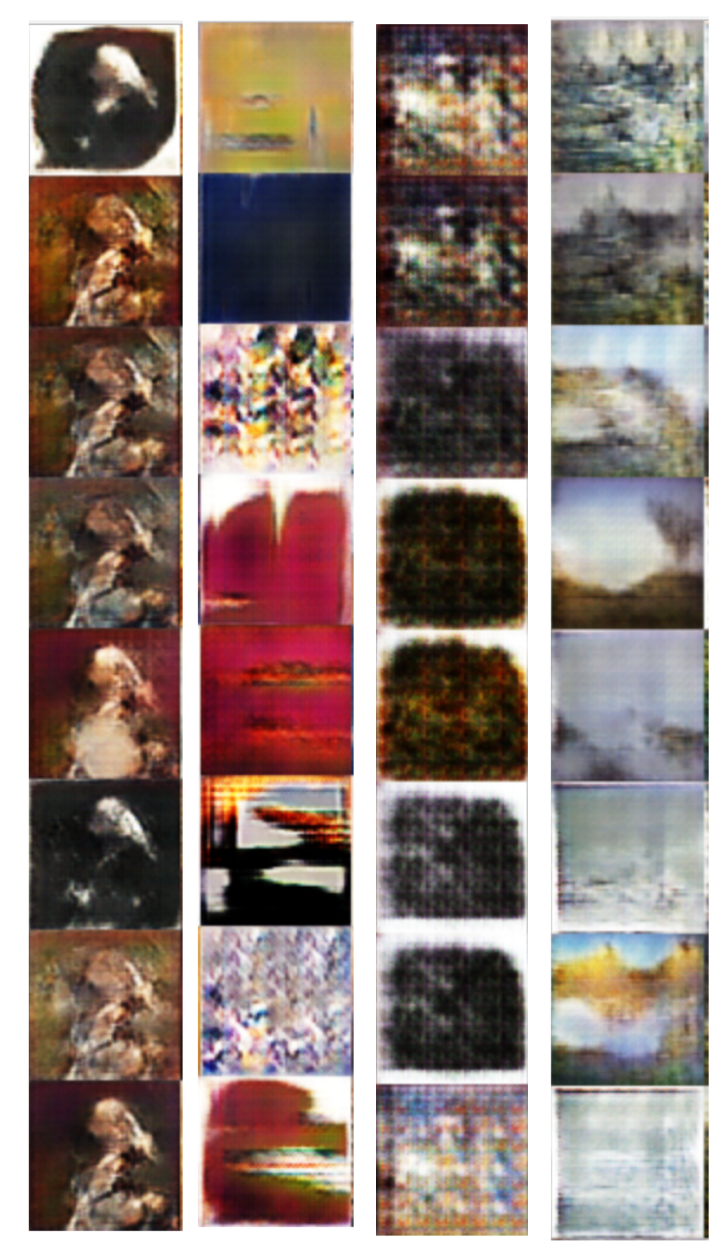

Figure 4. Results after running DCGAN for 30 epochs on single-genre data. Left to right: portraits, abstract, religious art, landscapes

Methods

- VanillaGAN architecture uses combination of convolutional and fully-connected layers
    - Generator uses bilinear interpolation for upsampling, discriminator uses average pooling for downsampling
- DCGAN and conditional DCGAN share same architecture (below), using only convolutional layers in discriminator and deconvolutional layers in generator for upsampling and downsampling
    - Conditional DCGAN injects label data as additional channels at each layer of generator and discriminator

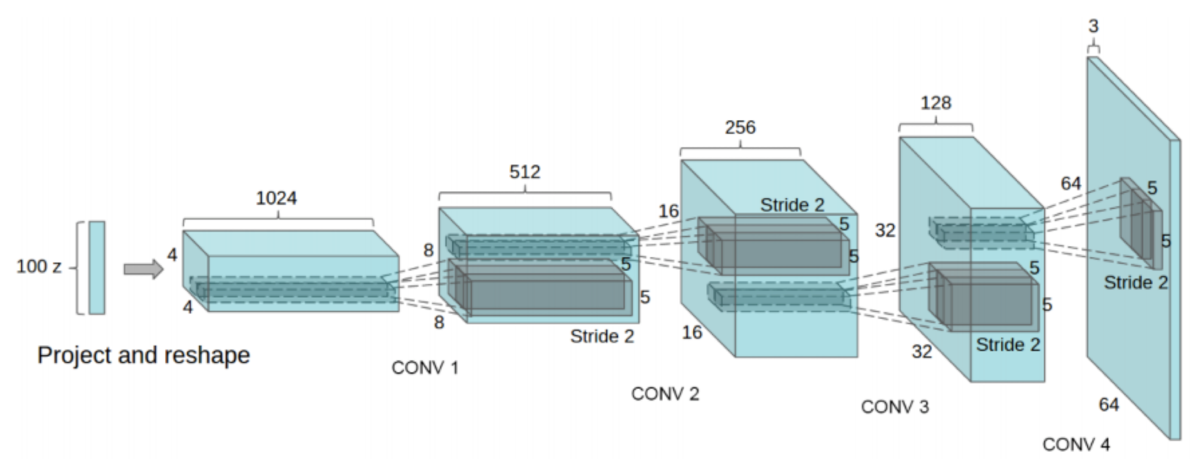

Figure 5. Architecture of DCGAN generator

- DCGAN was trained on both the full dataset of 6 genres, and on individual genres

Conclusions

- DCGAN outperforms VanillaGAN in both human assessment of result image quality and inception score
- Method of training image downsampling has significant effect on quality of generated images
- My experiments with conditional DCGAN did not converge, many possible explanations:
    - Imbalanced numbers of examples in classes/genres
    - Landscape and cityscape not sufficient distinct
    - More parameter tuning needed, different learning rate than non-conditional architecture
- Performance on individual genres varies by genre
    - Small number of examples most likely cause of religious art non-convergence