Disentangled representations of style and content for visual art with generative adversarial networks

Chris Donahue
Department of Music
University of California, San Diego
cdonahue@ucsd.edu

Julian McAuley
Department of Computer Science
University of California, San Diego
jmcauley@cs.ucsd.edu

Abstract

We propose an approach to learning disentangled representations of style and content for generative modeling of visual art. By fixing the style portion of the latent representation, we can generate diverse images in a particular style. Furthermore, we can fix the content portion, examining a particular scene through the lens of a variety of styles. Our approach pairs generative adversarial networks with Siamese discriminators; samples from the real dataset consist of two works from the same artist, and samples from the generator consist of two images with common style code. Unlike recent style transfer approaches, our work can imagine both style and content without the need of images to characterize each.

1 Introduction

Researchers have long experimented with systems that generate visual art. Early systems employed evolutionary strategies [1, 14], and more recently, researchers have leveraged neural networks, trained to recognize content in images, to disentangle representations of *style* and *content* in art [7, 9]. These *style transfer* approaches generate new works by combining the style of one image with the content of another; they cannot imagine either from scratch. Researchers have also experimented with using generative adversarial networks (GANs) to generate art from scratch [6, 10], however these approaches learn entangled representations of style and content. Here we propose a system for learning disentangled representations of style and content in art using GANs. In practice, this means that we can fix the style portion of the representation, imagining new work in a particular style.[1]

2 Method

Generative adversarial networks [8] learn mappings from latent codes $\mathbf{z}$ in some low-dimensional space $\mathcal{Z}$ to points in the space of natural data $\mathcal{X}$. To train the generative model $G : \mathcal{Z} \mapsto \mathcal{X}$, GANs pit G against a discriminative model $D : \mathcal{X} \mapsto [0, 1]$ in an adversarial learning framework. The learning process consists of a minimax game between G and D (see [8] for details):

$$\min_G \max_D V(G, D) = \mathbb{E}_{\mathbf{x}\sim P_R}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim P_{\mathcal{Z}}}[\log(1 - D(G(\mathbf{z})))]. \tag{1}$$

Semantically Decomposed GANs (SD-GANs) [5] combine GANs with pairwise training to decompose the generator's latent space $\mathcal{Z}$ into $\mathcal{Z}_S$, the variation caused by identity (style in our case), and $\mathcal{Z}_O$, the variation caused by other factors. Each sample from the real data consists of a pair of distinct images with common style: $\mathbf{x}_s^1, \mathbf{x}_s^2 \sim P_R(\mathbf{x}|S = s)$. Each sample from the generator consists of $G(\mathbf{z}_S^1), G(\mathbf{z}_S^2) \sim P_G(\mathbf{z} \mid \mathcal{Z}_S = \mathbf{z}_S)$, a pair of images with common identity code $\mathbf{z}_S \in \mathcal{Z}_S$:

$$\mathcal{L}_{SDGAN}(G, D) = \mathbb{E}_{\mathbf{x}_s^1,\mathbf{x}_s^2\sim P_R}[\log D(\mathbf{x}_s^1, \mathbf{x}_s^2)] + \mathbb{E}_{\mathbf{z}_S^1,\mathbf{z}_S^2\sim P_{\mathcal{Z}}}[\log(1 - D(G(\mathbf{z}_S^1), G(\mathbf{z}_S^2)))]. \tag{2}$$

[1] Interactive web demo: `https://chrisdonahue.github.io/sdgan_art`

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

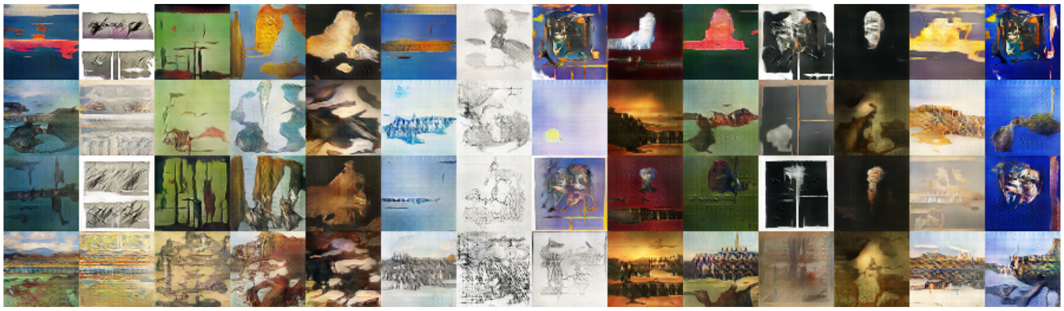

Figure 1: Generated samples from our method. Each of the four rows has a distinct style code and each of the fourteen columns has a distinct content code.

Elgammal et al. [6] propose creative adversarial networks, an extension of GANs with applications to visual art generation. Motivated by [11], they propose additional terms for the GAN loss function (Equation 1) that, they claim, encourage the generator to simultaneously generate realistic data while maximizing deviation from established styles. To this end, they additionally task the discriminator with classifying the style c of the input into one of C classes (reminiscent of [15]). Specifically, they minimize the cross entropy between softmax posterior $D(c \mid \mathbf{x})$ and real style labels:

$$\mathcal{L}_{class}(D) = \mathbb{E}_{\mathbf{x},\hat{c}\sim P_R}\left[-\frac{1}{C}\left(\log D(\hat{c} \mid \mathbf{x})) + \sum\nolimits_{c\neq\hat{c}} \log(1 - D(c \mid \mathbf{x}))\right)\right]. \tag{3}$$

Rather than *providing* the style to the generator as in [12], they train the generator to minimize cross entropy between $D(c \mid G(\mathbf{z}))$ and a uniform distribution:

$$\mathcal{L}_{uni}(G, D) = \mathbb{E}_{\mathbf{z}\sim P_{\mathcal{Z}}}\left[-\frac{1}{C}\sum\nolimits_{c=1}^{C} \log D(c \mid G(\mathbf{z}))\right]. \tag{4}$$

In this work, we combine our previous SD-GAN method [5] with the creative adversarial network approach of [4], yielding the adversarial value function:

$$\min_G \max_D V(G, D) = \mathcal{L}_{SDGAN}(G, D) - \mathcal{L}_{class}(D) + \mathcal{L}_{uni}(G, D) \tag{5}$$

To train our system, we follow the algorithm in [5]; to compute $\mathcal{L}_{class}$ and $\mathcal{L}_{uni}$ over minibatches consisting of pairs, we simply treat each pair as if it were two examples (effectively doubling the size of the minibatch for those terms).

3 Experiments

For our experiments, we use the training set of the *Painter by Numbers* Kaggle challenge.[2] This dataset contains works of 1584 artists across 135 categorical styles (e.g. renaissance, impressionism, cubism); we define *style* as an individual artist evoking a particular categorical style. By this definition, there are 3008 styles averaging 26 works in each, yielding one million style-matched pairs.

We train a DCGAN architecture [13] using the method described in Section 2 to generate 64x64 images. To compute the $\mathcal{L}_{SDGAN}$ portion of Equation 5 for a style-matched pair $(\mathbf{x}_s^1, \mathbf{x}_s^2)$, we concatenate $[D(\mathbf{x}_s^1); D(\mathbf{x}_s^2)]$ along the channel axis, performing one additional strided convolution before flattening and connecting to a sigmoid output (a Siamese approach [2, 3]). To compute the other portions, we flatten each $D(\mathbf{x}_s)$ and fully connect to a softmax layer with $C = 135$. We train using minibatches of 16 pairs for 150 epochs; all other training details are the same as in [5].

In Figure 1, we depict samples from the trained generator. Samples in each row share a common style vector and samples in each column share a content vector. The generator learns that content such as color palettes and contours can be represented in different styles. Furthermore, the model can imagine new content while keeping style fixed. While these results are preliminary, we would like to expand on this work by performing an ablation study for the terms of Equation 5. We also hope to train generative models capable of producing higher resolutions.

[2] `https://www.kaggle.com/c/painter-by-numbers/data`

References

[1] Ellie Baker and Margo I Seltzer. Evolving line drawings. 1993.

[2] Jane Bromley. Signature verification using a "siamese" time delay neural network. In *NIPS*, 1994.

[3] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *CVPR*, 2005.

[4] Simon Colton. Creativity versus the perception of creativity in computational systems. In *AAAI spring symposium: creative intelligent systems*, 2008.

[5] Chris Donahue, Akshay Balsubramani, Julian McAuley, and Zachary C Lipton. Semantically decomposing the latent spaces of generative adversarial networks. *arXiv:1705.07904*, 2017.

[6] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. In *International Conference on Computational Creativity*, 2017.

[7] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *CVPR*, 2016.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, 2014.

[9] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, 2016.

[10] Kenny Jones and Derrick Bonafilia. Gangogh: Creating art with gans. `https://medium.com/towards-data-science/gangogh-creating-art-with-gans-8d087d8f74a1`, 2017.

[11] Colin Martindale. *The clockwork muse: The predictability of artistic change*. Basic Books, 1990.

[12] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *ICML*, 2017.

[13] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *ICLR*, 2016.

[14] Jimmy Secretan, Nicholas Beato, David B D Ambrosio, Adelein Rodriguez, Adam Campbell, and Kenneth O Stanley. Picbreeder: evolving pictures collaboratively online. In *SIGCHI Conference on Human Factors in Computing Systems*, 2008.

[15] Jost Tobias Springenberg. Unsupervised and semi-supervised learning with categorical generative adversarial networks. In *ICLR*, 2015.