Generating Design Patterns for apparel using GANs

Anand Koshy
Oregon State University
Corvallis, OR - 97330
koshya@oregonstate.edu

Prachi Rahurkar
Oregon State University
Corvallis, OR - 97330
rahurkap@oregonstate.edu

Abstract

Can an algorithm create original and compelling fashion designs to serve as an inspirational assistant? To help answer this question, we propose a novel use of SA-GAN on clothing design pattern datasets (floral and Indian geometric designs) to generate new appealing patterns. This approach builds on the work of Ian Goodfellow's SA-GANs. As per our qualitative analysis, about 69% of our images are thought to be created by human designers rather than by a computer while also being considered original, and our proposed approach scores high in both novelty and likability. The datasets and the code can be found on this GitHub link.

1. Introduction

Imagine that you could be your own fashion designer, and be able to seamlessly transform your current outfit into a completely new one, by simply changing the design patterns on the clothing. In just minutes you could design hundreds of different shirts, dresses, or even pants, allowing you to easily discover what your liking lies in.

Artificial Intelligence (AI) research has been making huge progress in the machine's capability of human level understanding across the spectrum of perception, reasoning and planning (He et al., 2017; Andreas et al., 2016; Silver et al., 2016). Another key yet still relatively understudied direction is creativity where the goal is for machines to generate original items with realistic, aesthetic and/or thoughtful attributes, usually in artistic contexts. We can indeed imagine AI to serve as inspiration for humans in the creative process and also to act as a sort of creative assistant able to help with more mundane tasks, especially in the digital domain. Previous work has explored writing pop songs (Briot et al., 2017), imitating the styles of great painters (Gatys et al., 2016; Dumoulin et al., 2017) or doodling sketches (Ha and Eck, 2018) for instance. However, it is not clear how creative such attempts can be considered since most of them mainly tend to mimic training samples without expressing much originality. Creativity is a subjective notion that is hard to define and evaluate, and even harder for an artificial system to optimize for. Colin Martindale put down a psychology-based theory that explains human creativity in art (Martindale, 1990) by connecting creativity or acceptability of an art piece to novelty with "the principle of least effort". As originality increases, people like the work more and more until it becomes too novel and too far from standards to be understood.

Generative Adversarial Networks (Goodfellow et al., 2014; Radford et al., 2016) show a great capability to generate realistic images from scratch without requiring any existing sample to start the generation from. They can be applied to generate artistic content, but their intrinsic creativity is limited because of their training process that encourages the generation of items close to the training data distribution; hence they show limited originality and overall creativity. Similar conservative behavior can be seen in recent deep learning models for music generation where the systems are also mostly trained to reproduce pattern from training samples, like Bach chorales (Hadjeres and Pachet, 2017). Creative Adversarial Networks (CANs, Elgammal et al. (2017)) have then been proposed to adapt GANs to generate creative content (paintings) by encouraging the model to deviate from existing painting styles. Technically, CAN is a Deep Convolutional GAN (DCGAN) model (Radford et al., 2016) associated with an entropy loss that encourages novelty against known art styles. The specific application domain of CANs allows for very abstract generations to be acceptable but, as a result, does reward originality a lot without judging much how such enhanced creativity can be mixed with realism and standards.

DCGAN [2], a GAN formulation combined with convolutional networks, has been shown to be an effective model to produce realistic images. Moreover, it allows for an end-to-end embedding of textual descriptions to condition the image generation. The task of generating design patterns for clothing, presents two significant challenges which are difficult to address with the standard DCGAN. First, it directly targets the pixel values and provides no mechanism to enforce structural coherence with respect to the input. Second, it tends to average out the pixels [3], thus resulting in various artifacts, e.g. blurry boundaries.

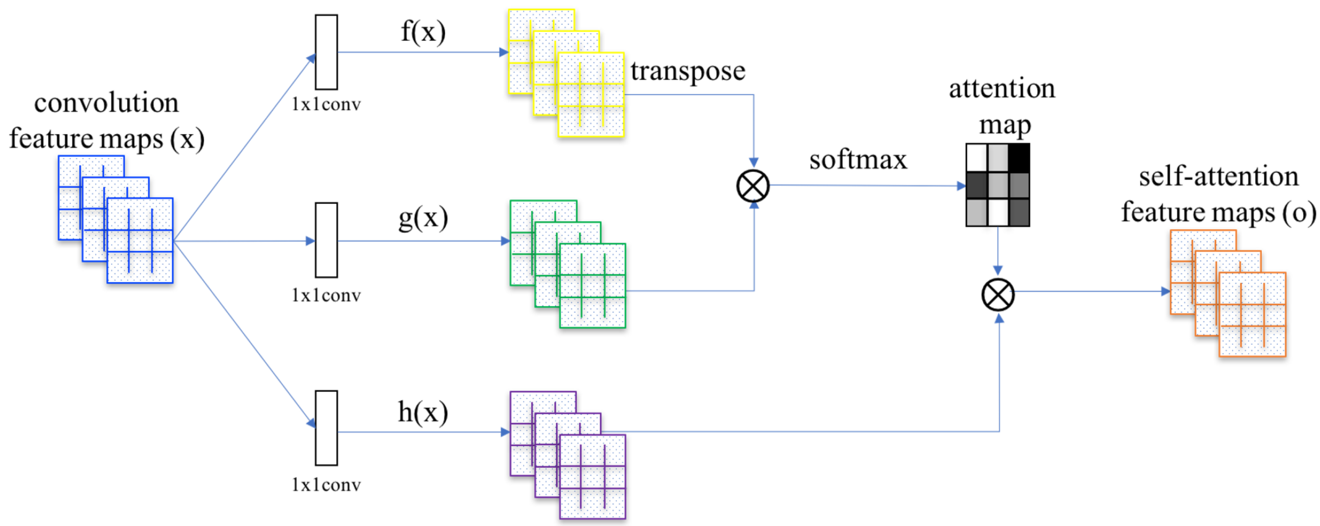

Figure 1: The Self-Attention mechanism

2. Related Work

Generative Adversarial Networks. GANs have achieved great success in various image generation tasks, including image-to-image translation [4, 5, 6, 7], image super-resolution [8, 9] and text-to-image synthesis [10, 11, 12]. Despite this success, the training of GANs is known to be unstable and sensitive to the choices of hyper-parameters. Several works have attempted to stabilize the GAN training dynamics and improve the sample diversity by designing new network architectures [13, 2, 12], modifying the learning objectives and dynamics [14, 15, 17, 18, 16], adding regularization methods [19, 20] and introducing heuristic tricks [21, 3]. Recently, Miyato et al. [20] proposed limiting the spectral norm of the weight matrices in the discriminator in order to constrain the Lipschitz constant of the discriminator function. Combined with the projection-based discriminator [22], the spectrally normalized model greatly improves class-conditional image generation on ImageNet.

Attention Models. Recently, attention mechanisms have become an integral part of models that must capture global dependencies [23, 24, 25, 26]. In particular, self-attention [27, 28], also called intra-attention, calculates the response at a position in a sequence by attending to all positions within the same sequence. Vaswani et al. [29] demonstrated that machine translation models could achieve state-of-the-art results by solely using a self-attention model. Parmar et al. [30] proposed an Image Transformer model to add self-attention into an autoregressive model for image generation. Wang et al. [31] formalized self-attention as a non-local operation to model the spatial-temporal dependencies in video sequences. In spite of this progress, self-attention has not yet been explored in the context of GANs. (AttnGAN [32] uses attention over word embeddings within an input sequence, but not self-attention over internal model states). SAGAN learns to efficiently find global, long-range dependencies within internal representations of images.

FashionGAN. Given an original wearer's input photo and different textural descriptions, this model generates new outfits onto the wearer while preserving the pose and body shape of the wearer. The authors of this paper propose fashion synthesis with structural coherence. The complex generative process is decomposed into two conditional stages. In the first stage, a plausible semantic segmentation map that obeys the pose of the wearer as a latent spatial arrangement is generated. In the second stage, a generative model with a newly proposed compositional mapping layer is used to render the final image with precise regions and textures conditioned on this map.

3. Self- Attention Generative Adversarial Networks

Most GAN-based models [13, 2, 3] for image generation are built using convolutional layers. Convolution processes the information in a local neighborhood, thus using convolutional layers alone is computationally inefficient for modeling long-range dependencies in images. Introducing self-attention to the GAN framework enables both the generator and the discriminator to efficiently model relationships between widely separated spatial regions.

The image features from the previous hidden layer are first transformed into two feature spaces to calculate the attention. Then the output of the attention layer is computed, as shown in Figure 1. The authors further multiply the output of the attention layer (o_i) by a scale

Figure 2: Floral Design Patterns Dataset

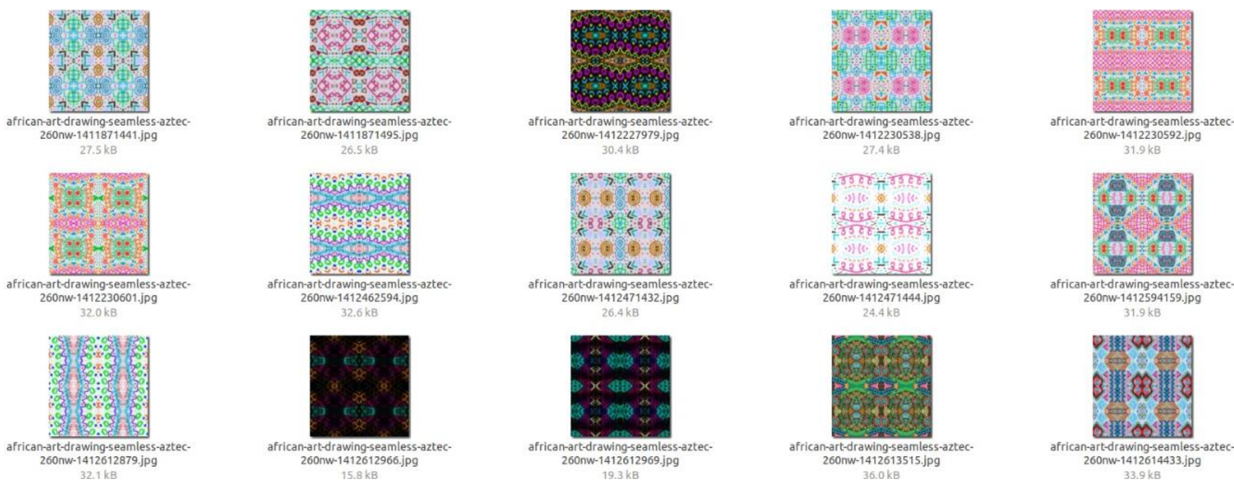

Figure 3: Indian Geometric Patterns Dataset

parameter and add back the input feature map (x_i). Therefore, the final output is given by $y_i = \gamma.o_i + x_i$, where γ is initialized as 0. [1]

This allows the network to first rely on the cues in the local neighborhood – since this is easier – and then gradually learn to assign more weight to the non-local evidence. The intuition for why this is done is straightforward: the easy task should be learnt first, and then progressively the complexity should be increased. In SA-GAN the proposed attention module has been applied to both generator and discriminator, which are trained in an alternating fashion by minimizing the hinge version of the adversarial loss [33, 20, 34].

4. Experiments

We present two unique and interesting datasets (floral & geometric) and use them to generate new designs using SA-GAN.

4.1 Datasets

Datasets for both Indian geometric patterns and floral design patterns were scraped from Shutterstock which is an image hosting site. The site has a rich collection for both geometric and floral images. We wrote a scraper in Python to extract images from Shutterstock for search queries of "Indian Geometric Patterns" & "Floral Design Patterns". Around 40000 images were scraped for each dataset, geometric and floral; but, the dataset was very noisy with images of different sizes and many images had text embedded in them.

Cleaning the dataset was a challenge since there were nearly 80000 images in total, but fortunately, Shutterstock names the images with the topic that it corresponds to. eg abstract-background-tribal-ikat-multicolored-260nw-247679563.jpg. Using these topics, we cleaned the data with a combination of both manual efforts and automated scripts.

Only images of the size 260 x 260 were used and the remaining images were filtered out by the above process. Some more manual filtering was required and eventually, the floral design pattern dataset (refer Figure 2) was created with about 2200 images; and the Indian geometric pattern dataset (refer Figure 3) was created with 3500 images; with both being publicly available on GitHub.

4.2 Implementation

We train the SA-GAN model on the above two datasets: floral and geometric. The SAGAN model we trained were designed to generate 64 x 64 images. By default, spectral normalization [20] is used for the layers in both generator and discriminator. Similar to [22], SAGAN uses conditional batch normalization in the generator and projection in the discriminator. For all models, we use the Adam optimizer [35] with $\beta1 = 0$ and $\beta2 = 0.9$ for training. By default, the learning rate for the discriminator is 0.0004 and the learning rate for the generator is 0.0001. Two layers of self-attention were used in both the generator and the discriminator.

Due to lack of computational power and resources, our training stopped at the 361600th step, which took almost two days to train. SA-GAN has achieved state-of-the-art performance on the ImageNet dataset with 1 million steps. After obtaining the new generated designs by SA-GAN we overlaid it over different types of apparels. We implemented this using a separate Python script.

5. Results

After running the SA-GAN on the floral dataset for 361600 steps, the result that we obtained is shown in Figure 5. The result of SA-GAN obtained on the geometric dataset after 115000 steps is shown in figure 4. The evaluation was focused on the floral dataset since the floral model was trained for many more steps.

Rating	Percent
1	31
2	14
3	55

Table 1: Qualitative Analysis

Qualitative evaluation of the results obtained was performed using human feedback. Forty-eight people were asked to score the images between 1 - 3, where -

1 - Does not resemble a flower,
2 - Slightly resembles a flower, and,
3 - Exactly resembles a flower.

Additional evaluation of FID was also computed; however, due to the low number of training steps, the values were relatively low - with the floral dataset achieving 157 after the 361600th step, and, the geometric dataset achieving 240 after the 115000th step.

Figure 4: Generated geometric design patterns

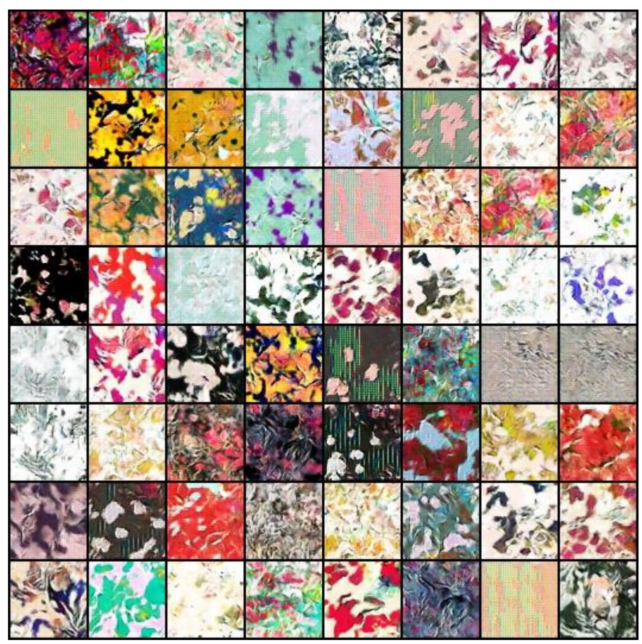

Figure 5: Generated floral design patterns

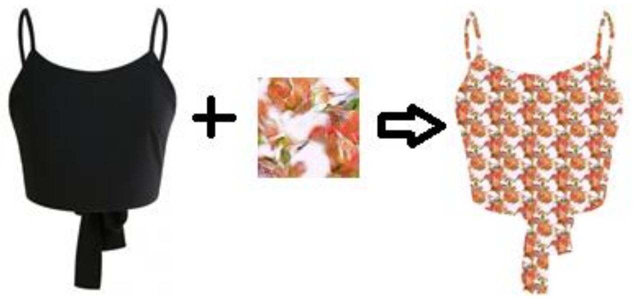

Figure 6: Applying the GAN-generated patterns on a sample blouse (above) and, on a sample dress (below)

6. Contributions

We present the following contributions through our project work:

1. We formalize a new task: Generating intricate design patterns specially for apparel.
2. We present two large-scale datasets – floral and geometric design patterns which are publicly made available.

7. Conclusion and Future Work

We still aim to obtain generated images of higher resolution. We also hope to increase the training data more, for much better results. We also wish to generate via GAN, a combined output of the shape of the apparel, along with the new design patterns on it.

References

[1] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena. Self-Attention Generative Adversarial Networks

[2] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.

[3] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. Improved techniques for training GANs.

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In CVPR, 2017.

[5] M. Liu and O. Tuzel. Coupled generative adversarial networks. In NIPS, 2016.

[6] Y. Taigman, A. Polyak, and L. Wolf. Unsupervised cross domain image generation. In ICLR, 2017.

[7] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In ICCV, 2017.

[8] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In CVPR, 2017.

[9] C. K. Sønderby, J. Caballero, L. Theis, W. Shi, and F. Huszár. Amortised map inference for image super-resolution. In ICLR, 2017.

[10] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text-to-image synthesis. In ICML, 2016.

[11] S. Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele, and H. Lee. Learning what and where to draw. In NIPS, 2016.

[12] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In ICCV, 2017.

[13] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In ICLR, 2018.

[14] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. arXiv:1701.07875, 2017.

[15] T. Salimans, H. Zhang, A. Radford, and D.N. Metaxas. Improving gans using optimal transport. In ICLR, 2018.

[16] L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein. Unrolled generative adversarial networks. In ICLR, 2017.

[17] T. Che, Y. Li, A. P. Jacob, Y. Bengio, and W. Li. Mode regularized generative adversarial networks. In ICLR, 2017.

[18] J. Zhao, M. Mathieu, and Y. LeCun. Energy-based generative adversarial network. In ICLR, 2017.

[19] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of wasserstein gans. In NIPS, 2017.

[20] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida. Spectral normalization for generative adversarial networks. In ICLR, 2018.

[21] A. Odena, C. Olah, and J. Shlens. Conditional image synthesis with auxiliary classifier gans. In ICLR, 2017.

[22] T. Miyato and M. Koyama. Cgans with projection discriminator. In ICLR, 2018.

[23] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. arXiv:1409.0473, 2014.

[24] K. Gregor, I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra. DRAW: A recurrent neural network for image generation. In ICML, 2015.

[25] K. Xu, J. Ba, R. Kiros, K. Cho, A. C. Courville, R. Salakhutdinov, R. S. Zemel, and Y. Bengio. Show, attend and tell: Neural image caption generation with visual attention. In ICML, 2015.

[26] Z. Yang, X. He, J. Gao, L. Deng, and A. J. Smola. Stacked attention networks for image question answering. In CVPR, 2016.

[27] J. Cheng, L. Dong, and M. Lapata. Long short-term memory-networks for machine reading. In EMNLP, 2016.

[28] A. P. Parikh, O. Täckström, D. Das, and J. Uszkoreit. A decomposable attention model for natural language inference. In EMNLP, 2016.
[29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. arXiv:1706.03762, 2017.
[30] N. Parmar, A. Vaswani, J. Uszkoreit, Łukasz Kaiser, N. Shazeer, and A. Ku. Image transformer. arXiv:1802.05751, 2018.
[31] X. Wang, R. Girshick, A. Gupta, and K. He. Non-local neural networks. In CVPR, 2018.
[32] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. In CVPR, 2018.
[33] J. H. Lim and J. C. Ye. Geometric gan. arXiv:1705.02894, 2017.
[34] D. Tran, R. Ranganath, and D. M. Blei. Deep and hierarchical implicit models. arXiv:1702.08896, 2017.
[35] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In ICLR, 2015.