Redefining Creativity: An Improved Creative Adversarial Network

Md. Shihab Shahriar Khan, Fazle Rabbi, Md. Aquib Azmain, Dr. Naushin Nower, Dr. Mohammad Shoyaib

Institute of Information Technology, University of Dhaka, Dhaka, Bangladesh

Abstract

Generative models for high dimensional data, in particular images have made tremendous advances in recent years. Although various methodologies, notably among them Generative Adversarial Network, have started to produce images with high fidelity and diversity, creativity has drawn considerably less attention in the literature. In this work, a novel method for painting generation that explicitly seeks to maximize creativity of produced paintings is proposed. Our method reinterprets and formalizes the notion of creativity and borrows ideas from adversarial sample generation to propose a simple and efficient algorithm for creative art generation. We experiment both on synthetic and real world datasets and demonstrate the effectiveness of our proposed method.

Keywords: Neural network, Generative adversarial network, Creativity, Creative adversarial network, Decision boundary, Adversarial attacks

1. Introduction

Solving creativity, a broad term that encapsulates understanding what makes any particular piece of art creative and applying these principles to generate new creative products in an automated way has long been pursued by scientists, specially the artificial intelligence community. One particular branch of this pursuit is automated painting generation, which is now undergoing a renewed wave of interest after discovery of some very promising techniques in image domain, in particular deep learning.

Since Alexnets [1] astonishing performance in ImageNet [2] classification challenge demonstrated potential of deep learning in discriminative tasks, interest grew in applying same technology in generative modeling tasks. Arguably the most famous success in this endeavour is generative adversarial network (GAN) [3] , which pitted two neural network against each other in Minimax fashion. GAN can produce very high dimensional and complex distributions in an unsupervised way. Although the images generated by GAN is approaching natural photographs in both fidelity and diversity[4], focus on explicitly making these samples creative has seen comparatively slower progress.

A notable work in this regard is Creative Adversarial Network (CAN) proposed in [5]. In it, they show for an artwork to be creative, it needs to be ambiguous in terms of style. To that end, they modify GAN, where discriminator is augmented with a new classification loss. Generator on the other hand try to keep discriminator confused by trying to maintain classification probability distribution of discriminator close to uniform.

In this paper, we build on the works of [5] to tackle the issue of explicitly modeling creativity on generative adversarial networks. We begin by formalizing creativity as defined in [5] from decision boundary perspective. We assume a dataset of painting to be multimodal, each mode comprising of a single style. In this framework, creative paintings as defined by CAN are the ones that lie at the decision boundary of the classification component of discriminator. We then show the uniform probability distribution loss of generator doesnt align well with above objective, and severely restricts the space of creative samples. To solve this, we propose a new generator loss that expands the creative space, while conforming to art-historic definition of creativity as provided in [5].

Email addresses: bsse0703@iit.du.ac.bd (Md. Shihab Shahriar Khan), bsse0725@iit.du.ac.bd (Fazle Rabbi), bsse0718@iit.du.ac.bd (Md. Aquib Azmain)

2. Literature Review

2.1. *Generating Art*

As early as 1990, there has been extensive work on rendering paintings. Stroke-based rendering (SBR) is the field of digitally generating paintings by arranging brushstrokes on a canvas according to some optimization goal [6]. A closely related field is texture transfer, there exist a large range of powerful non-parametric algorithms that can synthesise photorealistic natural textures by resampling the pixels of a given source texture [7, 8]. More recently, there is growing interest in neural style transfer[9], which given a pair of content and style images, try to stylize the content image along the style of style image. A common theme in all of these algorithms is that they start from a random noise image or already existing picture, then gradually render it by referencing a particular painting. This reliance on a single painting (style) is far simpler objective than trying to teach abstract, high-level artistic characteristics to a model.

Very recently, some works have attempted to apply generative modeling in neural style transfer [10, 11]. Although they can capture more abstract concepts like Cubism, Impressionism etc., these models are still limited to just one such concept.

2.2. *Deep Generative Models*

Deep Generative Models are unsupervised or semi-supervised methods to model the distribution of very high dimensional data like text, image, speech etc. Common approaches in this field include Variational Autoencoder (VAE) [12], Generative Adversarial Network (GAN) [3], autoregressive models[13], and normalizing flow models[14]. These are use parametric approaches to model data distribution, and through iterative optimization, try to reduce the distance between model distribution with data distribution. In image domain, which is of relevance to our work, GAN tends to consistently outperform the alternatives [3].

2.3. *Generating Decision Boundary Samples*

Decision boundary of a trained classifier is the region of data distribution where two or more classes have equal probability in predicted class probability distribution. Generating samples from near decision boundary region has drawn special attention in adversarial attack literature [15, 16], which try to find minimum perturbation to fool a classifier for a particular sample. These methods however start with a sample of training data, and gradually drive it towards decision boundary. [17] uses GAN to model decision boundary distribution, and can generate new boundary samples very efficiently.

3. Methodology

CAN paper [5] interpreted creative artwork as one that is stylistically ambiguous. They argued While GAN is able to emulate training data well, the images it generates are not very creative. To remedy this, they proposed a style ambiguity loss.

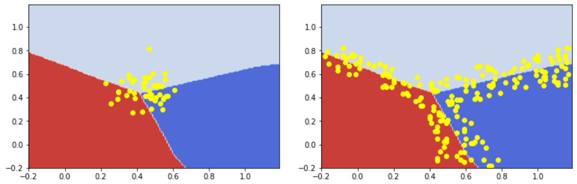
A vanilla GAN has two components, Generator (G) and Discriminator (D). The discriminator tries to discriminate between real images of the training set and fake images generated by the generator. The generator tries to generate images similar to the training set without seeing these images. It does so by mapping a random noise z to image space. Both of these networks are trained simultaneously, in minimax fashion (1).

$$\begin{aligned}\min_G \max_D \; V(D,G) = E_{x\sim p_{data}(x)}[logD(x)] \\ + E_{z\sim p_z(z)}[log(1 - D(G(z)))]\end{aligned} \tag{1}$$

CAN [5] augments both of these components with additional loss. Given a dataset of paintings along with their style labels, discriminator in CAN also tries to classify the style of paintings. Generator on the other hand, apart from trying to make generated paintings realistic,, it also tries to make their style ambiguous. To achieve that, it guides the class probability distribution of generated paintings towards uniform distribution (2).

$$\begin{aligned}&\min_G \max_D \; V(D,G) = \\ &\quad E_{x,\hat{c}\sim p_{data}}[logD_r(x) + logD_c(c = \hat{c}|x)]+ \\ &\quad E_{z\sim p_z}[log(1 - D_r(G(z))) - \sum_{k=1}^{K}(\frac{1}{K}log(D_c(c_k|G(z))+ \\ &\quad (1 - \frac{1}{K})log(1 - D_c(c_k|G(z)))]\end{aligned} \tag{2}$$

In this paper, we argue this uniform distribution loss is too strict and narrows the creative subspace of painting space. Style ambiguity, as defined in [5], is achieved when discriminator assigns equally high probability to just two styles, even if the rest are assigned low probability. This ensures, according to discriminator, generated painting doesnt entirely conform to any particular style, giving it the desired ambiguity. While equal probability for all styles is



(a) (b)

Figure 1: Comparison of creative subspace. 1(a) shows the samples from CAN's creative subspace (yellow). 1(b) shows the samples from subspace of our proposed method

of course in some sense more ambiguous, this is not necessary to conform to ambiguity definition as provided in psychology-based theory of creativity of [18].

This difference will perhaps be better clarified from decision boundary perspective. In figure 1 is the decision boundary of a neural network, trained on PCA transformed 2D space of Wine dataset [19].

Under uniform distribution loss of [5], creative samples can only come from the region around the intersection of all style regions, as depicted in figure 1a. In our method, creative samples can come from the intersection of any two style regions,as depicted in 1b. So while fully conforming to the ambiguity definition of [18], this new method significantly expands the creative subspace to sample paintings from, as can be seen from figure 1b.

Technically, we replace the class ambiguity loss of generator and keep the classification loss. To find a suitable loss function that is low around decision boundary but high everywhere else, we turn to adversarial attack literature. In particular, we use the Carlini-Wagner loss proposed in [20], which quantifies the difference in probability of two highest probable styles (3):

$$D(x)_y - \max_{\hat{y}\neq y}(D(x)_{\hat{y}}) \tag{3}$$

Where where x is image, y is the label with highest assigned probability and $D(x)_j$ represents probability assigned to j-th label by discriminator. So our final combined loss is (4):

$$\begin{aligned}\min_G \max_D \; & V(D,G) = \\ & E_{x,\hat{c}\sim p_{data}}[logD_r(x) + logD_c(c=\hat{c}|x)]+ \\ & E_{z\sim p_z}[log(1-D_r(G(z))) - (D(G(z))_c - \\ & \max_{\hat{c}\neq c}(D(G(z))_{\hat{c}}))]\end{aligned} \tag{4}$$

So to summarize, our complete training algorithm is shown in algorithm 1.

Algorithm 1 Proposed training algorithm with step size α, using mini-batch SGD for simplicity

```
Input:mini-batch images $x$, matching label $\hat{c}$, number of training batch steps $S$
for $n = 1$ to $S$ do
    $z \sim N(0,1)^Z$ {Draw sample of random noise}
    $\hat{x} \leftarrow G(z)$ {Forward through generator}
    $s_D^r \leftarrow D_r(x)$ {real image, real/fake loss}
    $s_D^c \leftarrow D_c(\hat{c}|x)$ {real image, multi class loss}
    $s_G^f \leftarrow D_r(\hat{x})$ {fake image, real/fake loss}
    $s_G^c \leftarrow (D(G(z))_c - \max_{\hat{c}\neq c}(D(G(z))_{\hat{c}})$ {Prposed decision Boundary loss, 'c' is the highest probable label}
    $L_D \leftarrow log(s_D^r) + log(s_D^c) + log(1 - s_G^f)$
    $D \leftarrow D - \alpha\partial L_D/\partial D$ {Update discriminator}
    $L_G \leftarrow log(s_G^f) - s_G^c$
    $G \leftarrow G - \alpha\partial L_G/\partial G$ {Update generator}
end for
```

Table 1: Artistic Styles Used in Training

Style name	Image number
Abstract-Expressionism	2782
Minimalism	1337
Naive Art-Primitivism	2405
Art-Nouveau-Modern	4334
Baroque	4241
Color-Field-Painting	1615
Contemporary-Realism	481
Cubism	2236
Early-Renaissance	1391
Expressionism	6736
Fauvism	934
Mannerism-Late-Renaissance	1279
High-Renaissance	1343
Impressionism	13060
New-Realism	314
Northern-Renaissance	2552
Pointillism	513
Pop-Art	1483
Post-Impressionism	6452
Realism	10733
Rococo	2089
Romanticism	7019

4. Result

We use a Wikiart dataset for to train our model. This dataset have around 80,000 paintings from 1,119

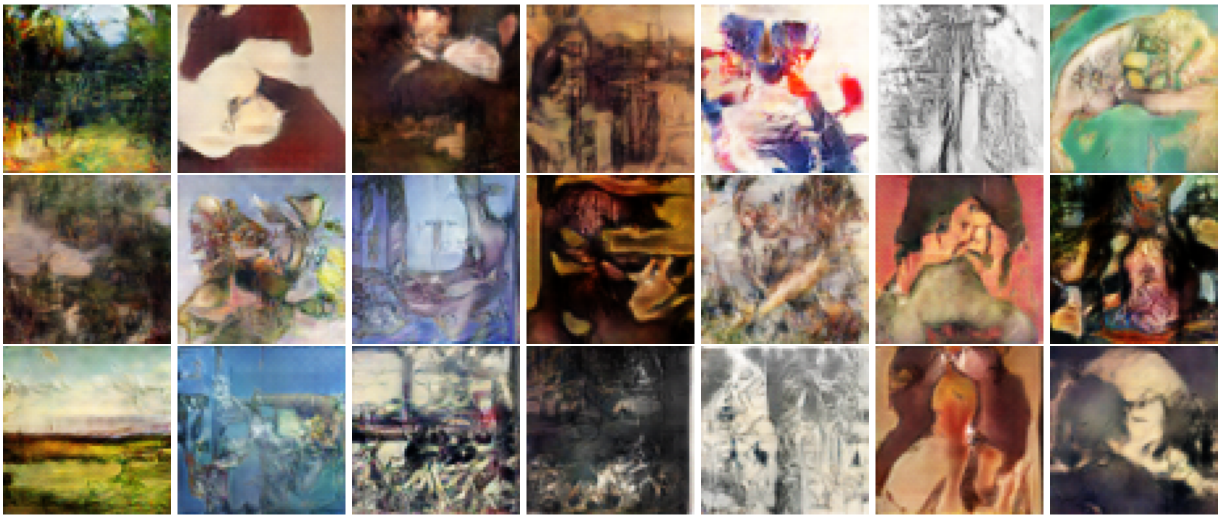


Figure 2: Images generated by CAN

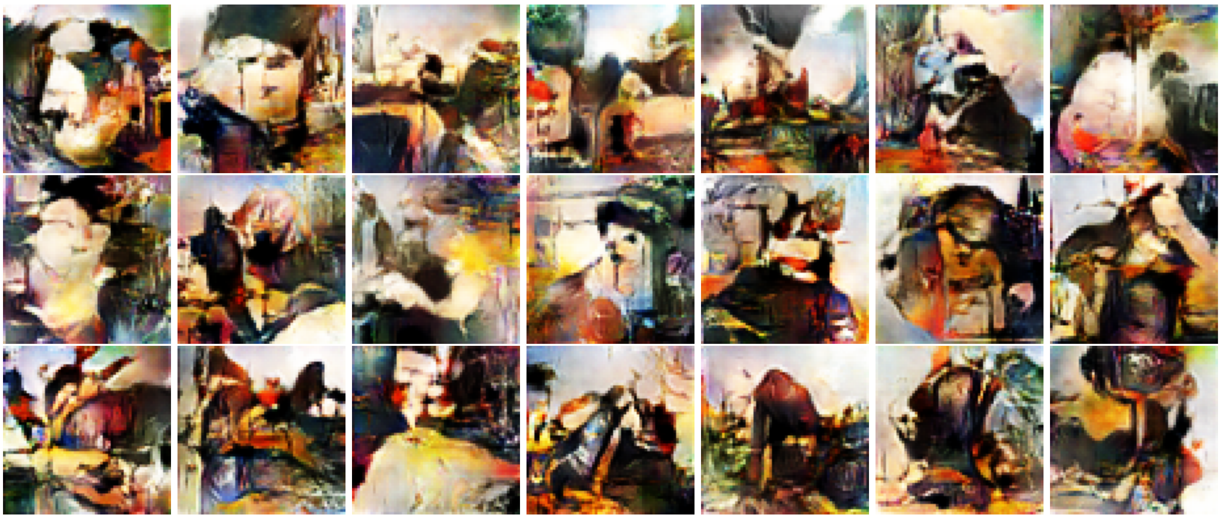


Figure 3: Images generated by our proposed method

artists and of 25 styles. We removed some styles of very few number of samples and trained on 22 classes (Table 1) . In figure 2 and 3, we compare the output of both methods. Figure 2 shows the images generated by CAN, figure 3 shows the images generated by our method. Please note that we dont claim to produce superior quality paintings compared to CAN, rather more diverse paintings due to extended space.

In the experiment below (Figure 4), we ran paintings generated by both our model and CAN through the trained discriminator. By looking at final class probability distribution, our goal was to measure the ambiguity of generated paintings.

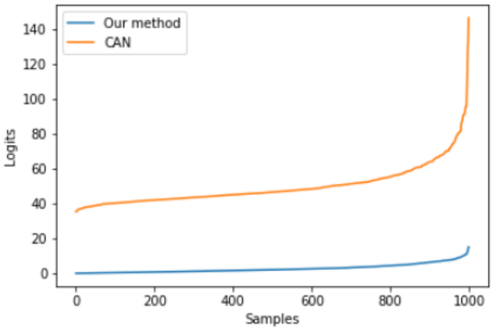


Figure 4: Comparison of difference of unnormalized logits between highest and second highest probable styles. Lower values indicate increased uncertainty by classifier in assigning style

5. Conclusion

In this paper, we presented a new method to generate art by proposing a new loss function that properly captures the creative subspace of painting distribution. While our method doesnt aim to or have managed to produce better quality art compared to baseline, this allows sampling more diverse and creative paintings. It has also formalized the notion of creativity using decision boundary analysis and has presented this problems connection with adversarial attack literature.

One potential way to extend this work is to utilize recent advances in adversarial attack field in boundary sample generation. Ideas and optimization techniques from style transfer literature can also hopefully be incorporated in current work to provide even richer, higher-quality creative paintings.

6. References

[1] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Neural Information Processing Systems 25 (2012).

[2] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, Li Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, pp. 2672–2680.

[4] A. Brock, J. Donahue, K. Simonyan, Large scale gan training for high fidelity natural image synthesis, arXiv preprint arXiv:1809.11096 (2018).

[5] A. Elgammal, B. Liu, M. Elhoseiny, M. Mazzone, Can: Creative adversarial networks, generating "art" by learning about styles and deviating from style norms.

[6] A. Hertzmann, A survey of stroke-based rendering, IEEE Computer Graphics and Applications (2003) 70–81.

[7] A. A. Efros, T. K. Leung, Texture synthesis by non-parametric sampling, in: Proceedings of the seventh IEEE international conference on computer vision, volume 2, IEEE, pp. 1033–1038.

[8] A. A. Efros, W. T. Freeman, Image quilting for texture synthesis and transfer, in: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM, pp. 341–346.

[9] L. A. Gatys, A. S. Ecker, M. Bethge, Image style transfer using convolutional neural networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2414–2423.

[10] A. Sanakoyeu, D. Kotovenko, S. Lang, B. Ommer, A style-aware content loss for real-time hd style transfer, in: Proceedings of the European Conference on Computer Vision (ECCV), pp. 698–714.

[11] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE international conference on computer vision, pp. 2223–2232.

[12] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, L. Carin, Variational Autoencoder for Deep Learning of Images, Labels and Captions, arXiv e-prints (2016) arXiv:1609.08976.

[13] H. Akaike, Fitting autoregressive models for prediction, Annals of the Institute of Statistical Mathematics 21 (1969) 243247.

[14] D. Jimenez Rezende, S. Mohamed, Variational inference with normalizing flows (2015).

[15] B. Biggio, I. Corona, D. Maiorca, B. Nelson, P. Laskov, G. Giacinto, F. Roli, Evasion attacks against machine learning at test time, pp. 387–402.

[16] F. Tramr, N. Papernot, I. Goodfellow, D. Boneh, P. McDaniel, The space of transferable adversarial examples (2017).

[17] C. Xiao, B. Li, J.-Y. Zhu, W. He, M. Liu, D. Song, Generating adversarial examples with adversarial networks, pp. 3905–3911.

[18] C. Martindale, The clockwork muse: The predictability of artistic change, Journal of the American Statistical Association 88 (1993).

[19] M. Forina, R. Leardi, C. Armanino, S. Lanteri, PARVUS: An Extendable Package of Programs for Data Exploration, Classification and Correlation, volume 4, 1988.

[20] N. Carlini, D. Wagner, Towards evaluating the robustness of neural networks, pp. 39–57.