CAN: Creative Adversarial Networks

Ahmed Elgammal[1] **Bingchen Liu**[1] **Mohamed Elhoseiny**[2] **Marian Mazzone**[3]
[1] Department of Computer Science, Rutgers University, NJ, USA
[2] Facebook AI Research, CA, USA
[3] Department of Art History, College of Charleston, SC, USA

Abstract

We propose a new system for generating art. The system generates art by looking at art and learning about style; and becomes creative by increasing the arousal potential of the generated art by deviating from the learned styles. We build over Generative Adversarial Networks (GAN), which have shown the ability to learn to generate novel images simulating a given distribution. We argue that such networks are limited in their ability to generate creative products in their original design. We propose modifications to its objective to make it capable of generating creative art by maximizing deviation from established styles and minimizing deviation from art distribution. We conducted experiments to compare the response of human subjects to the generated art with their response to art created by artists. The results show that human subjects could not distinguish art generated by the proposed system from art generated by contemporary artists and shown in top art fairs.

Introduction

Since the dawn of Artificial Intelligence, scientists have been exploring the machine's ability to generate human-level creative products such as poetry, stories, jokes, music, paintings, etc., as well as creative problem solving. This ability is fundamental in proving that Artificial Intelligent algorithms are in fact intelligent. In terms of visual art, several systems have been proposed to automatically create art, not only in the domain of AI and computational creativity (e.g. (Baker and Seltzer 1993; DiPaola and Gabora 2009; Colton et al. 2015; Heath and Ventura 2016)), but also in computer graphics (Sims 1991), and machine learning, (e.g. (Mordvintsev, Olah, and Tyka 2015; Johnson, Alahi, and Fei-Fei 2016)).

Within the computational creativity literature, different algorithms have been proposed mainly focused on investigating different and effective ways of exploring the infinite creative space. Several approaches used an evolutionary process where the algorithm iterates through generating some candidates and evaluate them using fitness function and then modifying them to improve the fitness score in the next iteration (e.g. (Machado, Romero, and Manaris ; DiPaola and Gabora 2009)). Typically this process is done within a genetic algorithm framework. As pointed out by DiPaola and Gabora 2009, the challenge these algorithms face is "how to write a logical fitness function that has an aesthetic sense". Some earlier systems utilized human in the loop with the role of guiding the process (e.g. (Baker and Seltzer 1993; Graf and Banzhaf 1995)). In these, interactive system, the computer plays the role of exploring the creative space, and the human plays the role of the observer whose feedback is essential in driving the process. Recent systems have emphasized the role of perception and cognition in the creative process (Colton 2008; Colton et al. 2015; Heath and Ventura 2016).

The goal of the paper is to investigate a computational creative system for art generation without involving a human artist in the creative process, however involving human creative products in the process. An essential component in art-generating algorithms is relating their creative process to art that already have been produced over the history and continue to be produced. We believe this is important because the human's creative process utilizes the prior experience and exposure. An artist is continuously being exposed to other artists' work, and has been exposed to different art all his/her life. This is problematic due to the lack of a clear answer about what is the underlying theory that explains what derive art progress over time. Such theory is needed to be able to integrate exposure to art with creation of art.

Colin Martindale (1943-2008) proposed a psychology-based theory that explains the progress (Martindale 1990). He hypothesized that at any point of time, creative artists try to increase the arousal potential of their produced art to push against habituation. However, this increase has to be minimal to avoid negative reaction by the observers (principle of least effort). Martindale also hypothesized that style break happens as a way of increasing arousal potential of art when artists exert other means within the roles of style. The approach proposed in this paper is inspired by Martindale's principle of least effort and his explanation of style break. Among the other theories that try to explain progress of art, we find Martindale's theory to be computationally feasible.

Deep neural networks have recently played a transformative role in advancing artificial intelligence across various application domains. In particular, several generative deep networks has been proposed with the ability to generate novel images to emulate a given training distribution (**?**). Generative Adversarial Networks (GAN) have been quite

successful in achieving this goal (Goodfellow et al. 2014). We argue that such networks are limited in their ability to generate creative products in their original design. Inspired by Martindale?s theory, in this paper we propose modifications to GAN's objective to make it able to generate creative art by maximizing deviation from established styles while minimizing deviation from art distribution.

Methodology

Background

The proposed approach is motivated from the theory suggested by D. E. Berlyne (1924-1976). Berlyne argued that the psychophysical concept of "arousal" has a great relevance on studying aesthetic phenomena (Berlyne 1971). "Level of arousal" measures how alert or excited a human being is. The level of arousal varies from lowest level, when a person is asleep or relaxed, to highest level when s/he is in violent, fury, or passion situations (Berlyne 1967). Among different mechanisms of arousal, of particular importance and relevance to art are properties of external stimulus patterns (Berlyne 1971).

The term arousal potential refers to the properties of stimulus patterns that lead to raising arousal. Besides other psychophysical and ecological properties of stimulus patterns, Berlyne emphasized that the most significant arousal-raising properties for aesthetics are *novelty, surprisingness, complexity, ambiguity, and puzzlingness*. He coined the term *collative variables* to refer to these properties collectively.

Novelty refers to the degree a stimulus differs from what an observer has seen/experienced before. Surprisingness refers to the degree a stimulus disagrees with expectation. Surprisingness is not necessarily correlated with novelty, for example it can stem from lack of novelty. Unlike novelty and surprisingness, which rely on inter-stimulus comparisons of similarity and differences, complexity is an intra-stimulus property that increases as the number of independent elements in stimulus grows. Ambiguity refers to the conflict between the semantic and syntactic information in a stimulus. Puzzlingness refers to the ambiguity due to multiple, potentially inconsistent, meanings.

Several studies have shown that people prefer stimulus with a moderate arousal potential (Berlyne 1967; Schneirla 1959). Too little arousal potential is considered boring, and too much activates the aversion system, which results in negative response. This behavior is explained by the Wundt curve that related the arousal potential with the hedonic response (Berlyne 1971; Wundt 1874).

Berlyne also studied arousal moderating mechanisms. Of particular importance in art is habituation, which refers to decreased arousal in response to repetitions of a stimulus (Berlyne 1971).

Martindale emphasized the importance of habituation in deriving the art-producing system(Martindale 1990). If the art producing system keeps producing similar works of arts this directly reduces the arousal potential and hence the likeness of that art. Therefore, at any point of time, the art-producing system will try to increase the arousal potential of produced art. In other words, habituation forms a constant pressure to change art. However, this increase has to be with the minimum amount necessary to compensate for habituation without falling into the negative hedonic range according to Wundt curve ("stimuli that are slightly rather than vastly supernormal are preferred"). Martindale called this the principle of "Least effort". Therefore, there is an opposite pressure that leads to a gradual change in the arts.

Art Generating Agent

We propose a model for an art-generating agent, and later we propose a realization of that model using a variant of GAN to make it creative. The agent's goal is to generate art with increased levels arousal potential in a constrained way to avoid activating the aversion system and falling into the negative hedonic range. In other words the agent tries to generate novel art but not too novel. This criterion is common in many computationally creative systems, however it is not easy to find a way to achieve that goal given the infinite possibilities in the creative space.

In our model the art-generating agent has a memory that encodes the art it has previously exposed to, which can continuously be updated with perception of new art. The agent utilizes this encoded memory in an indirect way while generating art with a restrained increase in arousal potential. While there are several ways to increase the arousal potential, in this paper we focus on building an agent that tried to increase the *stylistic ambiguity* and break from style norms while maintaining a force that pulls it back from moving away from what is accepted as art. In other words the agent tries to explore the creative space by deviating from the established style norms.

There are two types of ambiguities that are expected in the generated art by the proposed network; one is by design and the other one is inherited. Inherently, almost all computer-generated art is bound to be ambiguous from subject matter and figurative art point of view. The art generated will not have a clear figures or an interpretable subject matter. Along that line, Heath et al argued that the creative machine needs to have perceptual ability (learn to see) in order to be able to generate plausible creative art (Heath and Ventura 2016). This limited perceptual ability is what causes the inherited ambiguity. Typically, this type of ambiguity results in users being able to tell right away that the work is generated by a machine not an artist. Even though several styles of art developed in the 20th century might lack a clear figurative or lucid subject matter interpretation to the viewer, still, human observer would not be fooled to confuse a human-generated art from computer-generated art. Because of this inherited ambiguity people always think of computer-generated art as being hallucination-like. The Guardian commented on a the images generated by Google DeepDream (Mordvintsev, Olah, and Tyka 2015) by "Most, however, look like dorm-room mandalas, or the kind of digital psychedelia you might expect to find on the cover of a Terrence McKenna book"[1]. Others commented on it as being "dazzling, druggy, and creepy" [2]. This negative reaction might be explained as a

[1] Alex Rayner, the Guardian, March 28, 2016

[2] David Auerbach, Slate, July 23, 2015

result of too much arousal, which results in negative hedonic according to the Wundt curve.

The other type of ambiguity in the generated art by the proposed agent is stylistic-ambiguity which is intentional by design. The rational is that creative artists would eventually break from established styles and explore new ways of expression to increase the arousal potential of their art as Martindale suggested. As suggested by DiPaola and Gabora, "creators often work within a very structured domain, following rules that they eventually break free of" (DiPaola and Gabora 2009).

The proposed art-generating agent is realized by a model that we call Creative Adversarial Network which will describe next. The network is designed to generate art that does not follow established art movement or style, in contrast it tries to generate art that maximally confuses us as to which style it belongs to.

GAN: Emulative and not Creative

Generative Adversarial Network (GAN) has two sub networks, a generator and a discriminator. The discriminator has access to a set of image (training images). The discriminator tries to discriminate between "real" images (from 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. The generator starts by generating random images and receive a signal from the discriminator whether the discriminator finds them real or fake. At equilibrium the discriminator should not be able to tell the difference between the images generated by the generator and the actual images in the training set, hence the generator succeeds in generating images that come from the same distribution as the training set.

Let us now assume that we trained a GAN model on images of paintings. Since the generator is trained to generate images that fool the discriminator to believe it is coming from the training distribution, ultimately the generator will just generate images that look like already existing art. There is no motivation to generate any thing creative. There is no force that derives the generator to explore the creative space. Let us think about a generator that can cheat and already has access to samples from the training data. In that case the discriminator will right away be fooled to believe the generator is generating art, while in fact it is already existing art, and hence not novel and not creative.

There have been extensions to GANs that facilitate generating images conditioned on categories(e.g., (Radford, Metz, and Chintala 2016)) or captions (e.g., (Reed et al. 2016)). We can think of a GAN that can be designed and trained to generate images of different art styles or different art genres by providing such labels with training. This might be able to generate art that looks like, for example, Renaissance, Impressionist or Cubism. However that does not lead to any thing creative either. No creative artist will create art today that tries to emulate for Renaissance, Baroque, Impressionist style, or any traditional style. According to Berlyne and Martindale, artists would try to increase the arousal potential of their art by creating novel, surprising, ambiguous, puzzling art. This highlights the fundamental limitation of using GANs in generating creative works.

From being Emulative to being Creative

In the proposed Creative Adversarial Network, the generator is designed to receive two signals from the discriminator that act as two contradictory forces to achieve three points: First, generate novel works, Second the novel work should not too novel, i.e., it should not be far away from the distribution, otherwise it would generate too much arousal and would be activate the aversion system and hence will fall into the negative hedonic part according to the Wundt curve. Third, the generated work should increase the stylistic ambiguity.

Similar to Generative Adversarial Networks (GAN), the proposed network has two adversary networks, a discriminator and a generator. The discriminator has access to a large set of art associated with style labels (Renaissance, Baroque, Impressionism, Expressionism, etc.) and use it to learn to discriminate between styles. The generator does not have access to any art. It generates art starting from a random input, but unlike GAN, it receives two signals from discriminator for any work it generates. The first signal is the discriminator's classification to "art or not art". In traditional GAN, this signal enables the generator to change its weights to generate images that more and more will deceive the discriminator as coming from the same distribution. Since the discriminator in our case is trained on art, this will signal whether the discriminator think the generated art is coming from the same distribution as the actual art it knows about. In that sense, this signal flags whether the discriminator think the image presented to it is "art or not art". If the generator only receives this signal, it would eventually converge to generate images that will emulate art.

The second signal the generator receives is a signal about how well the discriminator can classify the generated art into established styles. If the generator generates images that the discriminator think it is art and also can classify it well to one of the established styles, then the generator would have fooled the discriminator to believe it generate actual art that lies within the established styles. In contrast, the creative generator will try to generate art that confuses the discriminator. In one hand it tries to fool the discriminator to think it is "art" and on the other hand it tries to confuse the discriminator about the style of the work generated.

These two signals are contradictory forces because, on one hand, the first signal pushes the generator to generate works that the discriminator accepts as "art". However, if it succeeded to do that within the rules of established styles, the discriminator will also be able to classify its styles, however, the second signal will heftily penalize the generator for doing that. This is because the second signal pushes the generator to generate style-ambiguous works. Therefore, these two signals together would expect to push the generator to explore parts of the creative space that lay close to the distribution of art (to maximize the first objective) and in the same time maximizes the ambiguity of the generated art with respect to how it fits in the realm of standard styles and art movements.

Technical Details

Generative Adversarial Networks

Generative Adversarial Networks (GANs) (Goodfellow et al. 2014) are one of the most successful image synthesis models in the past few years. GANs are typically trained by setting a game between two players. The first player is the called the generator G, which creates samples that are intended to come from the same probability distribution as the training data (i.e. p_{data}). The other player is the Discriminator D examines samples to determine whether they are real or fake. Both the discriminator and the generator are typically modeled as deep Neural Networks. The training procedure is similar to a two-player min-max game with the following objective function

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z)))], \quad (1)$$

where z is a noise vector sampled from distribution p_z (e.g., uniform or Gaussian distribution and x is a real image from the data distribution p_{data}. In practice, the discriminator D and the generator G are alternatively optimized for every batch. The discriminator D encourages maximizing Eq 1 by minimizing $-\mathbb{E}_{x\sim p_{data}}[\log D(x)] - \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z)))$, which improves the function of the D is a fake vs real image detector. Meanwhile, the generator G encourages minimizing Eq 1 by maximizing $log(D(G(z))$ which works better than $-log(1-D(G(z))$ since it provides stronger gradients for learning. By optimizing D and G alternatively, GANs are trained to generate images.

Creative Adversarial Networks

We modified the GAN loss function to achieve the vision explained in the previous section. We added a style classification loss to the discriminator and a style ambiguity loss to the discriminator. Maximizing the stylistic ambiguity can be achieved by maximizing the style class posterior probability. Hence, we need to design the loss such that the Generator G produces an image $x \sim p_{data}$ and meanwhile maximizes the entropy of $p(c|x)$ (i.e. the conditional distribution over the art classes given the generated image). The direct way to increase the stylistic ambiguity is to maximize the class posterior entropy. However, instead of maximizing the class posterior entropy, we minimize the cross entropy between the class posterior and a uniform target distribution. Similar to entropy that is maximized when the class posteriors (i.e., $p(c|G(z))$) are equi-probable, cross entropy with uniform target distribution will be minimized when the classes are equi- probable. So both objectives will be optimal when the classes are equi-probable. However, the difference is that the cross entropy will go sharply up at the boundary since it goes to infinity if any class posterior approaches 1 (or zero), while entropy goes to zero at this boundary condition. Therefore, using the cross entropy results in a hefty penalty if the generated image is classified to one of the classes with high probability. This in turn would generate very large loss, and hence large gradients if the generated images start to be classified to any of the style classes with high confidence. Hence, we can redefine the cost function with a different adversarial objective below

$$\min_G \max_D V(D,G) = \mathbb{E}_{x,\hat{c}\sim p_{data}}[\log D_r(x) + \log D_c(c=\hat{c}|x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D_r(G(z))) - \sum_{k=1}^{K}(\frac{1}{K} log(D_c(c_k|G(z)) + (1-\frac{1}{K}) log(1 - D_c(c_k|G(z)))], \quad (2)$$

where z is a noise vector sampled from distribution p_z (e.g., uniform or Gaussian distribution) and x and $\hat{c}$ are a real image and its corresponding label from the data distribution p_{data}, D_r is discriminate real and fake images, and D_c discriminate between different style categories (i.e., $D_c(c_k|G(z)) = p(c_k|G(z))$).

Discriminator Training : In Eq 2, the discriminator D encourages maximizing Eq 2 by minimizing $-\mathbb{E}_{x\sim p_{data}}[\log D_r(x) + \log D_c(c = \hat{c}|x)]$ for the real images and $-\mathbb{E}_{z\sim p_z}[\log(1 - D_r(G(z))) + \sum_{k=1}^{K} D_c(c_k|G(x)) log(D_c(c_k|G(z))]$ for the generated images. For the real images, the discriminator is encourages to classify the image x drawn from p_{data} as real by one loss and classify the image to the associated $\hat{c}$ by an additional K-way loss (where K is the number of classes).

Generator Training : the generator G encourages minimizing Eq 2 by maximizing $log(D_r(G(z)) + \sum_{k=1}^{K}(\frac{1}{K} log(D_c(c_k|G(z)) + (1-\frac{1}{K}) log(1 - D_c(c_k|G(z))$ which encourages the generates images to look real and meanwhile to have a big entropy for $p(c|G(z))$. Our hypothesis is that by optimizing D and G alternatively, our CANs are trained to generate creative images. Note that the CAN generator does not require any class label similar to unconditional Generative Model.

We denote the parameters for the real/fake discriminator D_r as θ_{D_r}, for the multi-label discriminator D_c as θ_{D_c}, and for the Generator G as θ_G. Let $D = \{D_r, D_c\}$. Algorithm 1 illustrates CAN training process.

Model Architecture : The Generator G and similar to architecture (Radford, Metz, and Chintala 2016), first $z \in \mathbb{R}^{100}$ normally sampled from 0 to 1 is up-sampled to a $4\times$ spatial extent convolutional representation with 2048 feature maps resulting in a $4 \times 4 \times 2048$ tensor. Then a series of four fractionally-stride convolutions (in some papers, wrongly called deconvolutions). Finally, convert this high level representation into a 256×256 pixel image. In other words, starting from $z \in \mathbb{R}^{100} \rightarrow 4 \times 4 \times 1024 \rightarrow 8 \times 8 \times 1024 \rightarrow 16 \times 16 \times 512 \rightarrow 32 \times 32 \times 256 \rightarrow 64 \times 64 \times 128 \rightarrow 128 \times 128 \times 64 \rightarrow 256 \times 256 \times 3$ (the generated image size). As described earlier, the discriminator has two losses (real/fake loss and multi-label loss). The discriminator in our work starts by a common body of convolution layers followed by two heads (one for the real/fake loss and one for the multi-label loss). *The common body* of convolution layers is composed of a series of six convolution layers (all

Algorithm 1 CAN 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**
  $z \sim \mathcal{N}(0,1)^Z$ {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}
  $s_G^c \leftarrow \sum_{k=1}^{K} \frac{1}{K} log(p(c_k|\hat{x}) + (1 - \frac{1}{K})(log(p(c_k|\hat{x}))$ {fake image Entropy loss}
  $\mathcal{L}_D \leftarrow \log(s_D^r) + \log(s_D^c) + \log(1 - s_G^f)$
  $D \leftarrow D - \alpha \partial \mathcal{L}_D / \partial D$ {Update discriminator}
  $\mathcal{L}_G \leftarrow \log(s_G^f) - s_G^c$
  $G \leftarrow G - \alpha \partial \mathcal{L}_G / \partial G$ {Update generator}
end for

with stride 2 and 1 pixel padding). conv1 (32 4×4 filters), conv2 (64 4×4 filters, conv3 (128 4×4 filters, conv4 (256 4×4 filters, conv5 (512 4×4 filters, conv6 (512 4×4 filters). Each convolutional layer is followed by a leaky rectified activation (LeakyRelU) (Maas, Hannun, and Ng 2013; Xu et al. 2015) in all the layers of the discriminator. After passing a image to the common conv D body, it will produce a feature map or size ($4 \times 4 \times 512$). *The real/fake D_r head* collapses the ($4 \times 4 \times 512$) by a fully connected to produce $D_r(c|x)$ (probability of image coming for the real image distribution). The *multi-label probabilities $D_c(c_k|x)$ head* is produced by passing the($4 \times 4 \times 512$) into 3 fully collected layers sizes 1024, 512, K, respectively, where K is the number of style classes. We plan to publish the source code of our work.

Initialization and Training parameters: The weights were initialized from a zero-centered Normal distribution with standard deviation 0.02. We used a mini-batch size of 128 and used mini-batch stochastic gradient descent (SGD) for training with 0.0001 as learning rate. In the LeakyReLU, the slope of the leak was set to 0.2 in all models. While previous GAN work has used momentum to accelerate training, we used the Adam optimizer and trained the model for 100 epochs (100 passes over the training data). To stabilize the training, we used Batch Normalization (Ioffe and Szegedy 2015) that normalizing the input to each unit to have zero mean and unit variance. We performed data augmentation by adding 5 crops within for each image (bottom-left, bottom-right, mid, top-left, top-right) on our image dataset. The width and hight of each crop is 90% of the width and the hight of the original painting.

Results and Validation

Training the model

We trained the networks using paintings from the publicly available WikiArt dataset [3]. This collection (as downloaded in 2015) has images of 81,449 paintings from 1,119 artists ranging from the fifteenth century to contemporary artists. Table 1 shows the number of images in each style.

[3]https://www.wikiart.org/

Table 1: Artistic Style Used in Training

Style name	Image number
cubism	2236
action-painting	98
impressionism	13060
expressionism	6736
art-nouveau-modern	4334
northern-renaissance	2552
analytical-cubism	110
synthetic-cubism	216
abstract-expressionism	2782
new-realism	314
contemporary-realism	481
baroque	4241
realism	10733
na-ve-art-primitivism	2405
early-renaissance	1391
high-renaissance	1343
pointillism	513
mannerism-late-renaissance	1279
rococo	2089
romanticism	7019
color-field-painting	1615
post-impressionism	6452
minimalism	1337
fauvism	934
pop-art	1483

Validation

Assessing the creativity of artifacts generated by the machine is an open and hard question. As noted by Colton 2008, aesthetic assessment of an artifact is different from the creativity assessment (Colton 2008).

We conducted human subject experiments to evaluate the creativity of the proposed model. We approach this assessment from a Turing test point of view. Human subjects are shown an image at a time and are asked whether they think it is created by the an artist or generated by a computer.

The goal of this experiment is to test whether human subjects would be able to distinguish art generated by the system from art generated by artists. However, the hard question is which art by human artists we should use for this comparison. Since the goal of this study is to evaluate the creativity of the artifacts produced by the proposed system, we need to compare human response to such artifacts with art that is considered to be novel and creative at this point in time. If we compare the produced artifacts to, for example, Impressionist art, we would be testing the ability of the system to emulate such art, and not the creativity of the system. Therefore we collected two sets of real artist works as well as two machine-generated as follows

1. Abstract Expressionist Set: A collection of 25 painting

Table 2: Means and standard deviations of responses of Experiment I

Painting set	Q1 (std)	Q2 (std)
CAN	53% (1.8)	3.2 (1.5)
GAN (Radford, Metz, and Chintala 2016)	35% (1.5)	2.8 (0.54)
Abstract Expressionist	85% (1.6)	3.3 (0.43)
Art Basel 2016	41% (2.9)	2.8 (0.68)
Artist sets combined	62% (3.2)	3.1 (0.63)

Table 3: Means and standard deviations of the responses of Experiment II

Painting set	Q1 (std)	Q2 (std)	Q3 (std)	Q4 (std)
CAN	3.3 (0.47)	3.2 (0.47)	2.7 (0.46)	2.5 (0.41)
Abstract Expressionist	2.8 (0.43)	2.6 (0.35)	2.4 (0.41)	2.3 (0.27)
Art Basel 2016	2.5 (0.72)	2.4 (0.64)	2.1 (0.59)	1.9(0.54)
Artist sets combined	2.7 (0.6)	2.5 (0.52)	2.2 (0.54)	2.1 (0.45)

by Abstract Expressionist masters made between 1945-2007, many of them by famous artists. This set was previously used in recent studies to compare human and machine?s ability to distinguish between abstract art by created artists, children or animals (Snapper et al. 2015; Shamir, Nissel, and Winner 2016). We use this set as a baseline set. Human subjects are expected to easily determine that these are created by artists.

2. Art Basel 2016 Set: This set consists of 25 paintings of various artists that were shown in Art Basel 2016, which is the flagship art fair for contemporary art world wide. Being shown in Art Basel 2016 is an indication that these are art works at the frontiers of human creativity in paintings, at least as judged by the art experts and the art market.
3. DC GAN Set: a set of 100 images generated by the state-of-the art Deep Convolution GAN (DCGAN) architecture trained on art (Radford, Metz, and Chintala 2016) which trained on our setup. It also uses the class-labels as we do for fair comparison.
4. CAN Set: a set of 125 images generated by the proposed model.

We used the same set of training images for both the GAN and CAN models and we conducted two human subject experiments as follows.

Experiment I:

The goal of this experiment is to test the ability of the system to generate art that human users would not distinguish from top creative art that is being generated by artists today.

In this experiment each subject is shown one image at time images from the four sets of images described above and asked:

Q1: Do you think the work is created by artist or generated by computer? The user has to choose one of two answers: artist or computer.

Q2: The user asked to rate how they like the image in a scale 1 (extremely dislike) to 5 (extremely like).

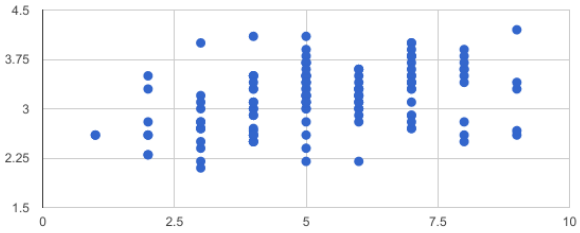


Figure 1: Experiment I (Q1 vs. Q2 responses)

18 users participated in this experiment.

Results: The results are summarized in Table 2. There are several conclusions we can draw from these results: 1) As expected, subjects rated the Abstract Expressionist set highly as being created by an artist (85%). 2) The proposed CAN model significantly out-perform GAN model in generating images that human think are generated by artist (53% vs. 35%). 3) More interestingly, human subject rated the images generated by CAN higher as being created by a human than the ones from the Art Basel set (53% vs. 41%) when combining the two sets of art created by artists, the images generated by CAN scored only less than 10% less (53% vs. 62%). Fig. 1 shows a scatter plot of the responses for the two quesions. Interestingly shows weak correlation between the likeness rating and whether subjects think it is by an artist or a computer.

Experiment II:

This experiment is similar to an experiment conducted by (Snapper et al. 2015) to determine to what degree human subject find the works of art to be intentional, having visual structure, communicative, and inspirational.

Q1: As I interact with this painting, I start to see the artists intentionality: it looks like it was composed very intentionally.

Q2: As I interact with this painting, I start to see a structure emerging.

Q3: Communication: As I interact with this painting, I feel that it is communicating with me.

Q4: Inspiration: As I interact with this painting, I feel inspired and elevated.

For each of the question the users answered in a scale from 1 (Strongly Disagree) to 5 (Strongly Agree). The users were asked to look at each image at least 5 second before answering. 21 users participated in this experiment.

(Snapper et al. 2015), hypothesized that subjects would rate works by real artists higher in these scales that works by children or animals, and indeed their experiments validated their hypothesis.

We also hypothesized that human subject would rate art by real artist higher on these scales than those generated by the proposed system. To our surprise the results showed that our hypothesis is not true! Human subjects rated the images generated by the proposed system higher than those created

by real artist, whether in the Abstract Expressionism set or in the Art Basel set (see Table 3).

While it might be debatable what higher score in each of these scales exactly means, the fact that human subjects found the images of generated by the machine intentional, having visual structure, communicative and inspiring indicate that human subjects see these images as art! Table 4 show several examples generated by our proposed CAN approach.

Discussion and Conclusion

We proposed a system for generating art with creative characteristics. We proposed a realization of this system based on a novel creative adversarial network. The system is trained using a large collection of images of art from 15th century to 21st century and their style labels. The system is able to generate art by optimizing a criterion that maximizes stylistic ambiguity while staying within the art distribution. The system was evaluated by human subject experiments which showed that human subjects similar level of confusion about the generated art and real art, and rated the generated art even higher than real art on different high-level scales.

What creative characteristics does the proposed system have? Colton 2008 suggested three criteria that a creative system should have: the ability to produce novel artifacts (imagination), the ability to generate quality artifacts (skill) and the ability to assess its own creation(Colton 2008). We shall discuss which of these criteria the proposed system possesses. The ability to produce novel artifacts is built-in in the system by construction through the interaction between the two signals that derive the generation, which forces the system to explore the space to find solutions that deviate from styles but stay close to the boundary of art. This interaction also provides a way for the system to self-assess its products. The quality of the artifacts are verified by the human subject experiments, which showed that subjects not only thought these artifacts are created by artists, but also rated them higher on different scales than real art.

One of the main characteristics of the proposed system is that it learns about art in its process to create art. However it does not have any semantic understanding of art behind the concept of styles. It does not know any thing about subject matter, or explicit models of elements or principle of arts. The learning here is only based on exposure to art and concepts of styles. In that sense the system has the ability to continuously learn from new art and would then adapt its generation based on what it learns.

It is left open how to interpret the subjects responses which ranked the CAN art than Art Basel Art samples in different aspects. It is because the users have typical style-backward bias? Are the subjects biased by their aesthetic assessment ? Would that mean that the results are not that creative? More experiments are definitely needed to help answering these questions.

References

[Baker and Seltzer 1993] Baker, E., and Seltzer, M. I. 1993. Evolving line drawings. 1

[Berlyne 1967] Berlyne, D. E. 1967. Arousal and reinforcement. In *Nebraska symposium on motivation*. University of Nebraska Press. 2

[Berlyne 1971] Berlyne, D. E. 1971. *Aesthetics and psychobiology*, volume 336. JSTOR. 2

[Colton et al. 2015] Colton, S.; Halskov, J.; Ventura, D.; Gouldstone, I.; Cook, M.; and Perez-Ferrer, B. 2015. The painting fool sees! new projects with the automated painter. In *Proceedings of the 6th International Conference on Computational Creativity*, 189–196. 1

[Colton 2008] Colton, S. 2008. Creativity versus the perception of creativity in computational systems. In *AAAI spring symposium: creative intelligent systems*, volume 8. 1, 5, 7

[DiPaola and Gabora 2009] DiPaola, S., and Gabora, L. 2009. Incorporating characteristics of human creativity into an evolutionary art algorithm. *Genetic Programming and Evolvable Machines* 10(2):97–110. 1, 3

[Goodfellow et al. 2014] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *Advances in neural information processing systems*, 2672–2680. 2, 4

[Graf and Banzhaf 1995] Graf, J., and Banzhaf, W. 1995. Interactive evolution of images. In *Evolutionary Programming*, 53–65. 1

[Heath and Ventura 2016] Heath, D., and Ventura, D. 2016. Before a computer can draw, it must first learn to see. In *Proceedings of the 7th International Conference on Computational Creativity, page to appear*. 1, 2

[Ioffe and Szegedy 2015] Ioffe, S., and Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*. 5

[Johnson, Alahi, and Fei-Fei 2016] Johnson, J.; Alahi, A.; and Fei-Fei, L. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, 694–711. Springer. 1

[Maas, Hannun, and Ng 2013] Maas, A. L.; Hannun, A. Y.; and Ng, A. Y. 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML*, volume 30. 5

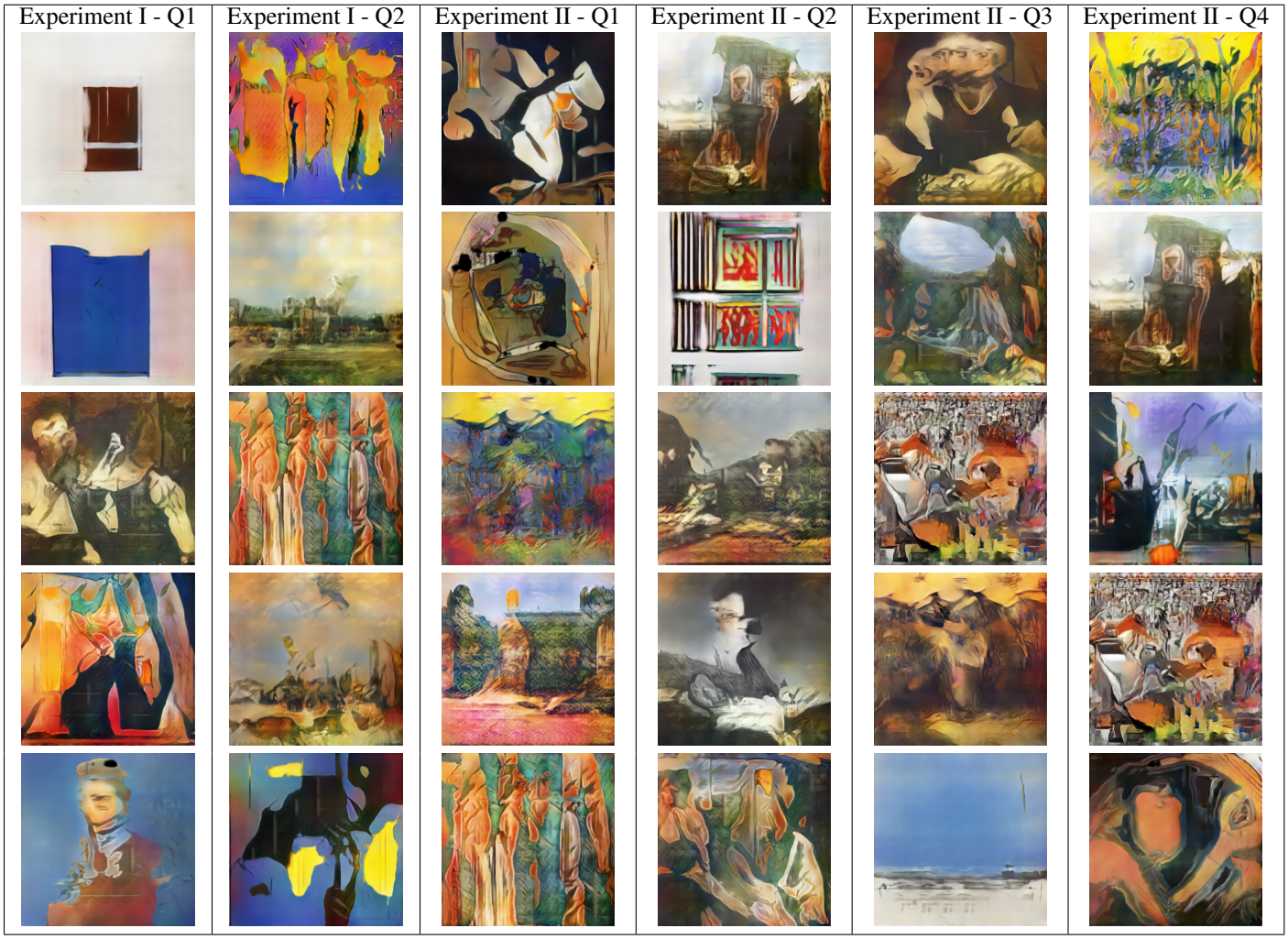
[Machado, Romero, and Manaris] Machado, P.; Romero, J.; and Manaris, B. An iterative approach to stylistic change in evolutionary art. 1

[Martindale 1990] Martindale, C. 1990. *The clockwork muse: The predictability of artistic change.* Basic Books. 1, 2

[Mordvintsev, Olah, and Tyka 2015] Mordvintsev, A.; Olah, C.; and Tyka, M. 2015. Inceptionism: Going deeper into neural networks. *Google Research Blog. Retrieved June* 20:14. 1, 2

[Radford, Metz, and Chintala 2016] Radford, A.; Metz, L.; and Chintala, S. 2016. Unsupervised representation learning

Table 4: Top Ranked Images From CAN in Human Subject Experiment I and II



Top ranked images generated by CAN according to the responses of human subjects for each question in experiments I and II (in top-down order). Since same image may appear on the top of different questions, the repeated images are replaced by the next image in their ranks.

with deep convolutional generative adversarial networks. 3, 4, 6

[Reed et al. 2016] Reed, S.; Akata, Z.; Yan, X.; Logeswaran, L.; Schiele, B.; and Lee, H. 2016. Generative adversarial text-to-image synthesis. In *Proceedings of The 33rd International Conference on Machine Learning*. 3

[Schneirla 1959] Schneirla, T. C. 1959. An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. 2

[Shamir, Nissel, and Winner 2016] Shamir, L.; Nissel, J.; and Winner, E. 2016. Distinguishing between abstract art by artists vs. children and animals: Comparison between human and machine perception. *ACM Transactions on Applied Perception (TAP)* 13(3):17. 6

[Sims 1991] Sims, K. 1991. *Artificial evolution for computer graphics*, volume 25. ACM. 1

[Snapper et al. 2015] Snapper, L.; Oranç, C.; Hawley-Dolan, A.; Nissel, J.; and Winner, E. 2015. Your kid could not have done that: Even untutored observers can discern intentionality and structure in abstract expressionist art. *Cognition* 137:154–165. 6

[Wundt 1874] Wundt, W. M. 1874. *Grundzüge de physiologischen Psychologie*, volume 1. W. Engelman. 2

[Xu et al. 2015] Xu, B.; Wang, N.; Chen, T.; and Li, M. 2015. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*. 5