Creative Walk Adversarial Networks: Novel Art Generation with Probabilistic Random Walk Deviation from Style Norms

Divyansh Jha*, **Kai Yi, Ivan Skorokhodov, Mohamed Elhoseiny***
King Abdullah University of Science and Technology (KAUST)
{divyansh.jha, kai.yi, ivan.skorokhodov, mohamed.elhoseiny}@kaust.edu.sa, * denotes Equal contribution.

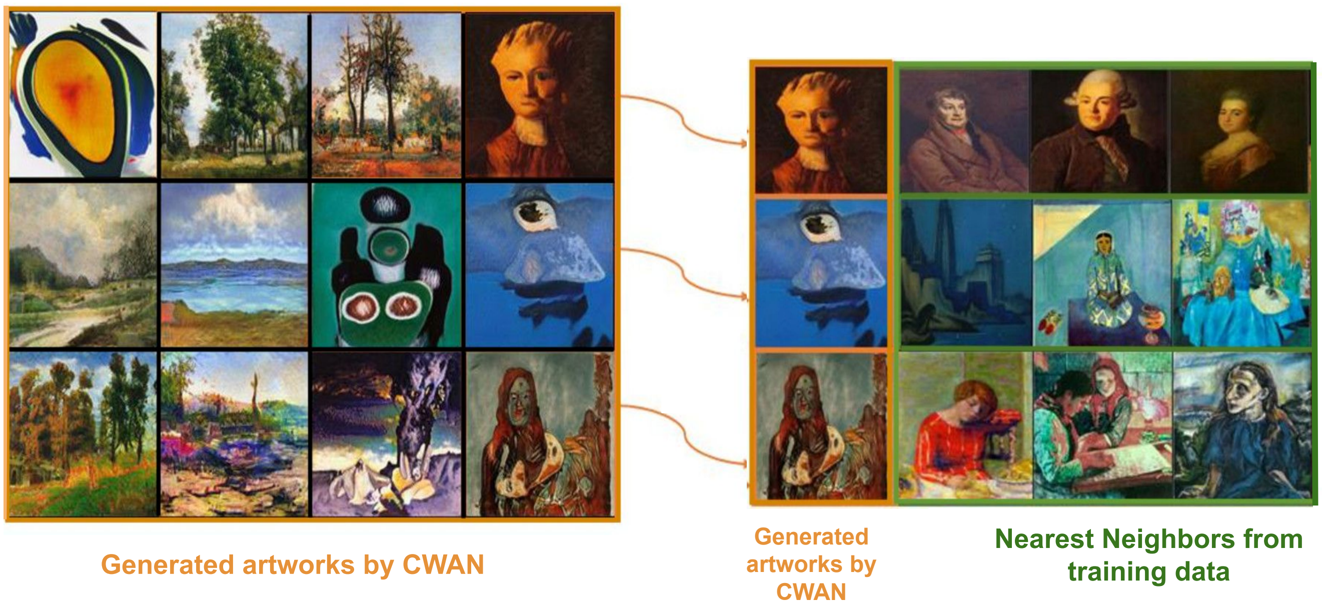

Figure 1: Art images on left with orange borders are generated using Creative Walk Adversarial Networks. The right part shows the Nearest Neighbors (NN) from the training set on the WikiArt dataset (with green borders), which are different indicating our generations' novelty. Nearest neighbor distance is computed on ResNet-18 space (He et al. 2016).

Abstract

We propose Creative Walk Adversarial Networks (CWAN) for novel art generation. Quality learning representation of unseen art styles is critical to facilitate generation of new meaningful artworks. CWAN learns an improved metric space for generative art by exploring unseen visual spaces with probabilistic random walks. CWAN constructs a dynamic graph that includes the seen art style centers and generated samples in the current minibatch. We then initiate a random walk from each art style center through the generated artworks in the current minibatch. As a deviation signal, we encourage the random walk to eventually land after T steps in a feature representation that is difficult to classify as any of the seen art styles. We investigate the ability of the proposed loss to generate meaningful novel visual art on the WikiArt dataset. Our experimental results and human evaluations demonstrate that CWAN can generate novel art that is significantly more preferable compared to strong state-of-the-art methods, including StyleGAN2 and StyleCAN2. The code is publicly available at: https://vision-cair.github.io/CWAN/

Introduction

Generative models like Generative Adversarial Networks (GANs) (Goodfellow et al. 2014a) and Variational Auto Encoders (VAEs) (Kingma and Welling 2013) are excellent tools for generating images due to their ability to represent high-dimensional probability distributions. However, they are not explicitly trained to go beyond distribution seen during training. Hence, the generations tends to be more emulative than creative. To generate likable novel visual content, GANs' training has been augmented with explicit losses that encourages careful deviation from existing classes, as first demonstrated in Creative Adversarial Networks (CANs) (Elgammal et al. 2017a). These models were shown to have some capability to produce unseen aesthetic art (Elgammal et al. 2017a; Hertzmann 2018; Jha, Chang, and Elhoseiny 2021), fashion (Sbai et al. 2018), design (Nobari, Rashad, and Ahmed 2021), and sculpture (Ge et al. 2019). Producing these creative generations is mainly leveraged by the generative model's improved ability to learn visual representations of novel generations that are distinguishable from seen ones. Similar deviation mechanisms was shown to have generalization benefit, improving performance on the task of unseen class recognition by encouraging discrimination explicitly between seen and unseen generations(Elhoseiny and Elfeki 2019; Elhoseiny, Yi, and Elfeki 2021).

We propose Creative Walk Adversarial Networks (*CWAN*) as a new learning system for generating artworks. We build our method on top of the state-of-the-art GAN architectures, StyleGANs (Karras, Laine, and Aila 2019a; Karras et al. 2020), due to their superior performance as compared to VAEs. We augment StyleGANs with *parameter-free* graph-based loss, dubbed as Creative Walk loss, to improve learning representation of unseen Artworks generatively. We first represent each art style class (e.g., cubism, High renaissance) by its center, representing the mean neural representation of the given Art style. Our Creative Walk loss then starts from the center of each seen art style class and performs a random walk through the generated images for T steps. Then, we encourage the landing representation to be distant and distinguishable from the seen art style centers. In summary, the Creative Walk loss is computed over a similarity graph involving the centers of seen art styles and the generated images/art pieces in the current minibatch. Thus, Creative Walks takes a global view of the data manifold compared to existing deviation losses that are local/per example; e.g., (Sbai et al. 2018; Elgammal et al. 2017a). Our work can be connected to recent advances in semi-supervised learnin, that leverage unlabeled data within the training classes, e.g., (Zhang et al. 2018)(Ayyad et al. 2020)(Ren et al. 2018)(Haeusser, Mordvintsev, and Cremers 2017)(Li et al. 2019). In these methods, unlabeled data are encouraged to attract existing classes. In contrast, our goal is the *opposite*, deviating from seen styles. Also, creative walks operate on generated images instead of provided unlabeled data.

Contribution. We propose Creative Walk Adversarial Networks(CWAN) for novel art generation. CWANs augment state-of-the-art adversarial generative models with a Creative Walk loss that learns an improved metric space for novel art generation. The loss generatively explores unseen art discriminatively against the existing art style classes. The augmented loss is unsupervised on the generative space and can be applied to any GAN architectures; e.g., DCGAN (Radford, Metz, and Chintala 2016), StyleGAN (Karras, Laine, and Aila 2019a), and StyleGAN2 (Karras et al. 2020). We show that Creative Walk Adversarial Networks helps understand unseen visual styles better, improving the generative capability in unseen space of liked art as compared to state-of-the-art baselines including StyleGAN2(Karras et al. 2020) and StyleCAN2(Jha, Chang, and Elhoseiny 2021); see Fig. 1.

Related Work

Generative Models with Deviation Losses. In the context of computational creativity, several approaches have been proposed to produce original items with aesthetic and meaningful characteristics (Machado and Cardoso 2000; Mordvintsev, Olah, and Tyka 2015; DiPaola and Gabora 2009; Tendulkar et al. 2019). Various early studies have made progress on writing pop songs (Briot, Hadjeres, and Pachet 2017), and transferring styles of great painters to other images (Gatys, Ecker, and Bethge 2016; Date, Ganesan, and Oates 2017; Dumoulin et al. 2017; Johnson, Alahi, and Li 2016; Isola et al. 2017) or doodling sketches (Ha and Eck 2018). The creative space of the style transfer images is limited by the content image and the stylizing image, which could be an artistic image by Van Gogh. GANs (Goodfellow et al. 2014a; Radford, Metz, and Chintala 2016; Ha and Eck 2018; Reed et al. 2016; Zhang et al. 2017; Karras et al. 2018; Karras, Laine, and Aila 2019a) have a capability to learn visual distributions and produce images from a latent z vector. However, they are not trained explicitly to produce novel content beyond the training data. More recent work explored an early capability to produce novel art with CAN (Elgammal et al. 2017b), and fashion designs with a holistic CAN (an improved version of CAN) (Sbai et al. 2018), which are based on augmenting DCGAN (Radford, Metz, and Chintala 2016) with a loss encouraging deviation from existing styles. The difference between CAN and holistic-CAN is that the deviation signal is Binary Cross Entropy over individual styles for CAN (Elgammal et al. 2017b) and Multi-Class Cross-Entropy (MCE) loss overall styles in Holistic-CAN (Sbai et al. 2018). (Jha, Chang, and Elhoseiny 2021) recently proposed *StyleCAN* model, which augments the Holistic CAN loss on StyleGANs, showing an improved performance compared to StyleGANs in art generation.

In contrast to these deviation losses, our Creative Walk loss is global. It establishes dynamic messages between generations produced in every mini-batch iteration and seen visual spaces. These generations deviate from style norms represented by the centers of the seen art style classes. In our experiments, we added the proposed loss to StyleGAN1 and StyleGAN2 architectures to produce novel visual artworks, showing superior likeability compared to existing losses.

Creative Walk Adversarial Networks

We start this section by the formulation of our *Creative Walk* loss. We will show later in this section how state-of-the-art deep-GAN models can be integrated to encourage novel visual generations. We denote the generator as $G(z)$ and its corresponding parameters as θ_G. As in (Goodfellow et al. 2014b; Karras, Laine, and Aila 2019a), the random vector $z \in \mathbb{R}^Z$ sampled from a Gaussian distribution $p_z = \mathcal{N}(0, 1)$ to generate an image. Hence, $G(z)$ is an generated image from the noise vector z. We denote the discriminator as D and its corresponding parameters as θ_D. The discriminator is trained with two objectives: (1) predicting real images from the training images and fake for generated ones. (2) identify the art style class of the input artwork. The discriminator then has two classification heads. The first head is for binary real/fake classification; $\{0, 1\}$ classifier. The second head is a K-way classifier over the seen art style classes, where K is the number of style classes in the training dataset. We denote the real/fake probability produced by D for an input image as $D^r(\cdot)$, and the classification score of a seen style class $k \in \mathcal{S}$ given the image as $D^c(\cdot)$, where $\mathcal{S}$ is the set of seen art styles.

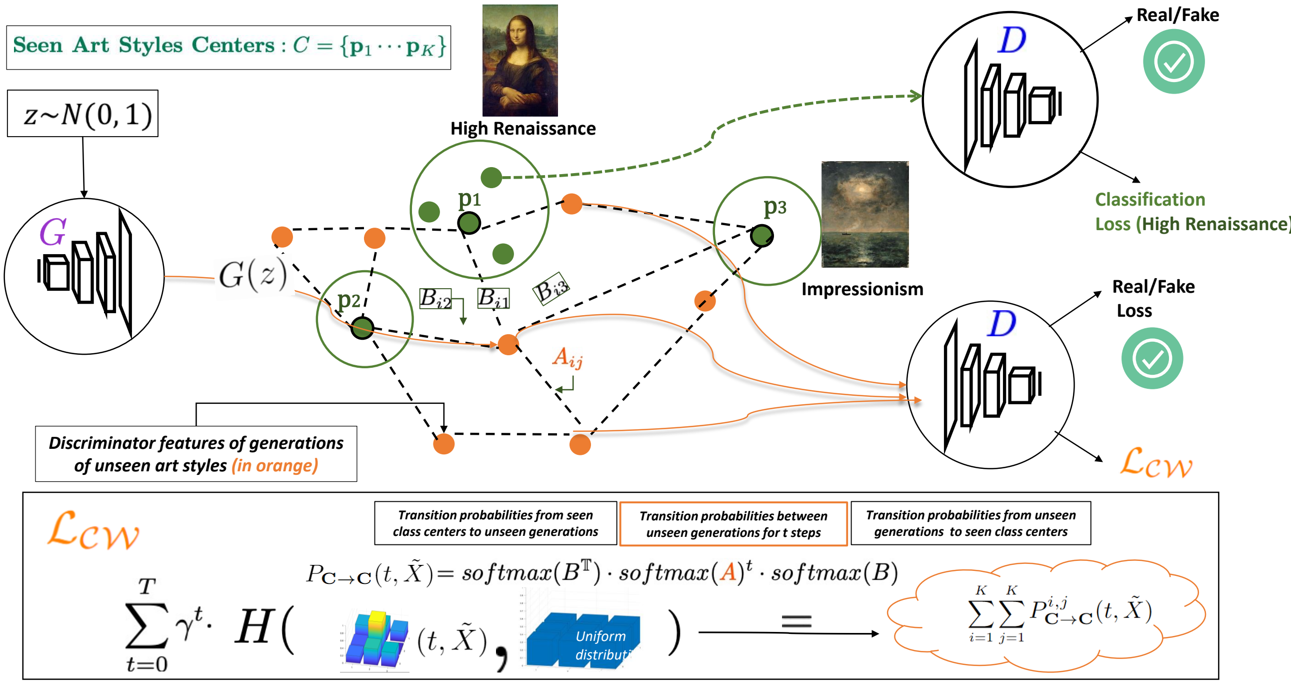

Figure 2: Creative Walk loss starts from each seen style class center (i.e., $\mathbf{p}_i$). It then performs a random walk through generated examples of hallucinated unseen classes using $G(z)$ for T steps. The landing probability distribution of the random walk is encouraged to be uniform over the seen classes. For careful deviation from seen classes, the generated images are encouraged to be classified as real by the Discriminator D. H indicates relative entropy; see Eq. 4 for detailed definition.

Creative Walk Loss

We denote the seen class centers, or prototypes[1], that we aim to deviate from as $C = \{\mathbf{p}_1 \cdots \mathbf{p}_K\}$, where $\mathbf{p}_i$ represents center of seen class/style i and K is the number of seen art styles that we aim to deviate from. We compute $C = \{\mathbf{p}_1 \cdots \mathbf{p}_K\}$ by sampling a small episodic memory of size m for every class and computing $\mathbf{p}_i$ from the discriminator representation. Concretely, we randomly sample $m = 10$ examples per class once and compute at each iteration its mean discriminator features, computed as activations from the last layer of the Discriminator D followed by scaled L2 normalization $L2(\mathbf{v}, \beta) = \beta \frac{\mathbf{v}}{\|\mathbf{v}\|}$, $\beta = 3$.

With the generator $G(\cdot)$, we sample generated images $\tilde{X}$ of size $\tilde{N}$ that we aim them to deviate from the seen art styles. $\tilde{X}$ is then embedded to the same feature space as style centers with the discriminator. Let $B \in \mathbb{R}^{\tilde{N}\times K}$ be the similarity matrix between the features of the generations ($\tilde{X}$) and the seen class centers (C). Similarly, let $A \in \mathbb{R}^{\tilde{N}\times\tilde{N}}$ be the similarity matrix between the generated images. In particular, we use the negative Euclidean distances between the embeddings as a similarity measure as follows:

$$B_{ij} = -\|\tilde{x}_i - \mathbf{p}_j\|^2, \quad A_{i,j} = -\|\tilde{x}_i - \tilde{x}_j\|^2 \qquad (1)$$

where $\tilde{x}_i$ and $\tilde{x}_j$ are i^{th} and j^{th} features in the set $\tilde{X}$; see Fig. 2. To avoid self-cycle, The diagonal entries $A_{i,i}$ are set to a small number ϵ.

[1] we refer alternatively between prototypes and centers

Hence, we defined three transition probability matrices:

$$P_{\mathbf{C}\to\tilde{X}} = \sigma(B^{\mathbb{T}}), \quad P_{\tilde{X}\to\mathbf{C}} = \sigma(B), \quad P_{\tilde{X}\to\tilde{X}} = \sigma(A) \qquad (2)$$

where σ is the softmax operator is applied over each row of the input matrix, $P_{\mathbf{C}\to\tilde{X}}$ and $P_{\tilde{X}\to\mathbf{C}}$ are the transition probability matrices from each seen class over the $\tilde{N}$ generated images and vice-versa respectively. $P_{\tilde{X}\to\tilde{X}}$ is the transition probability matrix from each generated image over other generated images. We hence define our generative random walker probability matrix as:

$$P_{\mathbf{C}\to\mathbf{C}}(t, \tilde{X}) = P_{\mathbf{C}\to\tilde{X}} \cdot (P_{\tilde{X}\to\tilde{X}})^t \cdot P_{\tilde{X}\to\mathbf{C}} \qquad (3)$$

where $P^{i,j}_{\mathbf{C}\to\mathbf{C}}(t, \tilde{X})$ denotes the probability of ending a random walk of a length t at a seen class j given that we have started at seen class i; t denotes the number of steps taken between the generated points, before stepping back to land on a seen art style.

Creative Walk Loss. Our random walk loss aims at boosting the deviation of unseen visual spaces from seen art style classes. Hence, we define our loss by encouraging each row in $P_{\mathbf{C}\to\mathbf{C}}(t)$ to be hard to classify to seen classes as follows

$$\begin{aligned} L_{CW} = &-\sum_{t=0}^{T} \gamma^t \cdot \sum_{i=1}^{K}\sum_{j=1}^{K} U_c(j) log(P^{i,j}_{\mathbf{C}\to\mathbf{C}}(t, \tilde{X})) \\ &- \sum_{j=1}^{N_u} U_x(j) log(P_v(j)) \end{aligned} \qquad (4)$$

where the first term minimizes cross entropy loss between every row in $P_{\mathbf{C}\to\mathbf{C}}(t, \tilde{X}) \forall t = 1 \to T$ and uniform distribution over seen classes $U_c(j) = \frac{1}{K^s}, \forall j = 1 \cdots K^s$, where

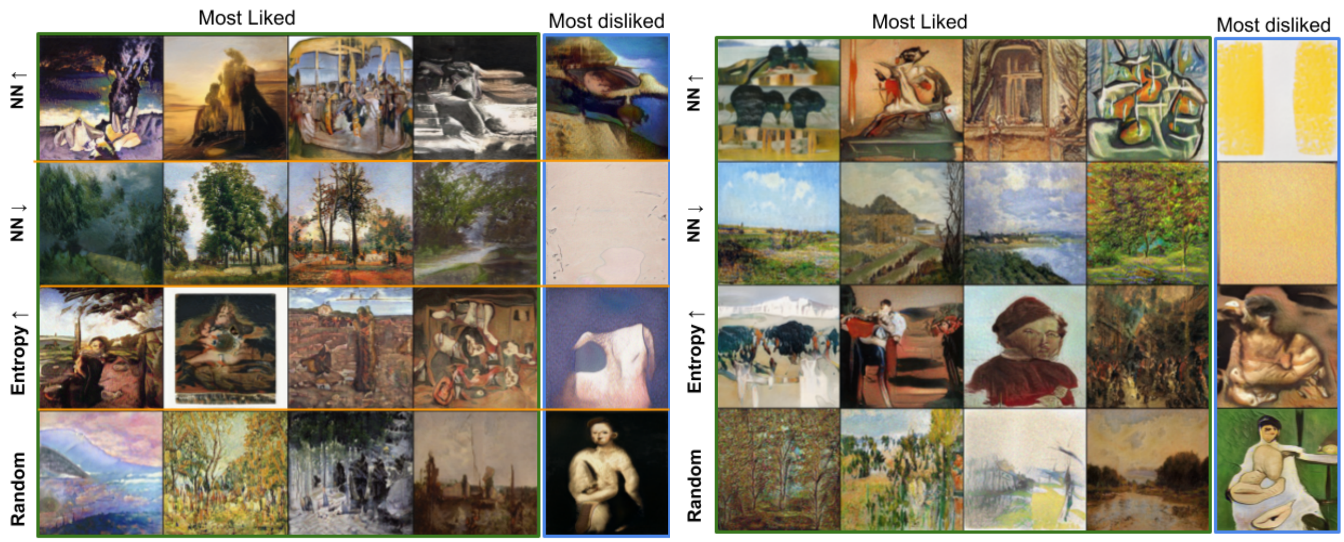

Figure 3: Most liked and disliked art generated using StyleGAN1 + CWAN(left) and StyleGAN2 + CWAN(right) architectures.

T is a hyperparameter and γ is exponential decay set to 0.7 in our experiments. In the second term, we maximize the probability of all the generations $\tilde{x}_i \in \tilde{X}$ to be equality visited by the random walk; see Fig. 2. This term is called the "visit loss" and was proposed in (Haeusser, Mordvintsev, and Cremers 2017) to encourage random walker to visit a large set of unlabeled points. We compute the overall probability that each generated point would be visited by any of the seen class $P_v = \frac{1}{K}\sum_{i=0}^{K} P^i_{\mathbf{C}\to\tilde{X}}$, where $P^i_{\mathbf{C}\to\tilde{X}}$ represents the i^{th} row of the $P^{\mathbf{C}\to\tilde{X}}$ matrix. The visit loss is then defined as the cross-entropy between P_v and the uniform distribution $U_x(j) = \frac{1}{\tilde{N}}, \forall j = 1 \cdots \tilde{N}$. Hence, visit loss encourages to visit as many examples as possible from $\tilde{X}$ and hence improves learning representation.

Note that, if we replace U_c by an identity matrix to encourage landing to the starting seen class, the loss becomes an attraction signal similar to (Haeusser, Mordvintsev, and Cremers 2017), which defines its conceptual difference to the Creative Walk loss. We integrated our loss with StyleGAN1 (Karras, Laine, and Aila 2019a) and StyleGAN2 (Karras et al. 2020) by simply adding L_{GRW} in Eq. 4 to the generator loss. The generator and discriminator losses are defined as follows

$$\mathcal{L}_G = \mathcal{L}_{G \text{ GAN}} + \lambda \mathcal{L}_{CW} \quad (5)$$

$$\mathcal{L}_D = \mathcal{L}_{D \text{ GAN}} + \lambda \mathcal{L}_{\text{style\_classification}} \quad (6)$$

where $\mathcal{L}_{G \text{ GAN}}$ and $\mathcal{L}_{D \text{ GAN}}$ are the default generator and discriminator loss, used in the adopted GAN architecture (e.g., DCGAN, StyleGAN1. StyleGAN2). Similar to (Elgammal et al. 2017a; Sbai et al. 2018), we add art style classification loss, $\mathcal{L}_{\text{style\_classification}}$, on real art images to $\mathcal{L}_D$.

Experiments

Dataset. We performed our experiments on the WikiArt datasets (WikiArt 2015), which contains approximately 81k art works from 27 different art styles and over 1k artists.

Nomenclature. Our models are referred as CWAN-T(value), where CWAN means Creative Walk Adversarial Network, with Creative Walk loss of T time steps. We name our models according to this convention throughout this section. We perform human subject experiments to evaluate generated art. We set value of the loss coefficient λ as 10 in all our experiments. We divide the generations from these models into four groups, each containing 100 images; see examples in Fig. 3.

- **NN↑.** Images with high nearest neighbor (NN) distance from the training dataset.
- **NN↓.** Images with low nearest neighbor (NN) distance from the training dataset.
- **Entropy ↑.** Images with high entropy of the probabilities from a style classifier trained on WikiArt dataset.
- **Random (R).** A set of random images.

For example, we denote generations using CWAN with T=10, and NN↑ group as CWAN-T10_NN↑. Fig. 3 shows top liked/disliked paintings according to human evaluation on StyleGAN1 and StyleGAN2 with our Creative Walk loss.

Baselines. We performed comparisons with two baselines, i.e., (1) the vanilla GAN for the chosen architecture, and (2) adding Holistic-CAN loss (Sbai et al. 2018) (an improved version of CAN (Elgammal et al. 2017b)). For simplicity, we refer the Holistic-CAN as CAN. We also compared to StyleCAN(Jha, Chang, and Elhoseiny 2021) model, an adaptation of the holistic CAN loss on the state-of-the-art StyleGAN (Karras, Laine, and Aila 2019b) and StyleGAN2 (Karras et al. 2020) architectures.

Human Evaluation. We performed our human subject MTurk experiments based on StyleGAN1 (Karras, Laine, and Aila 2019b) & StyleGAN2 (Karras et al. 2020) architecture's vanilla GAN, CAN, and CWAN variants. We conducted three types of experiments; see Fig. 5.

1. **Likeability Experiment:** Following(Elgammal et al. 2017a), we performed the likeability experiments on

Table 1: Human experiments on generated art from vanilla GAN, and CAN, and CWAN. CWAN obtained the highest mean likeability in all the groups. Here Q1 is asking for a likeability score and Q2 is asking whether the art work is created by a computer/human. See Likeability Experiment for more details. More people believed the generated art to be real for the artwork generated from model trained using the Creative Walk loss.

		Likeability Mean					Turing Test
Loss	Architecture	Q1-mean(std)	NN ↑	NN ↓	Entropy ↑	Random	Q2(% Artist)
CAN (Elgammal et al. 2017b)	DCGAN	3.20(1.50)	-	-	-	-	53
GAN (Vanilla) (Karras, Laine, and Aila 2019a)	StyleGAN	3.12(0.58)	3.07	3.36	3.00	3.06	55.33
CAN (Jha, Chang, and Elhoseiny 2021)	StyleGAN	3.20(0.62)	3.01	3.61	3.05	3.11	56.55
CWAN-T3 (Ours)	StyleGAN	**3.29(0.59)**	3.05	3.58	3.13	**3.38**	54.08
CWAN-T10 (Ours)	StyleGAN	**3.29(0.63)**	**3.15**	**3.67**	**3.15**	3.17	**58.63**
GAN (Vanilla) (Karras et al. 2020)	StyleGAN2	3.02(1.15)	2.89	3.30	2.79	3.09	54.01
CAN (Jha, Chang, and Elhoseiny 2021)	StyleGAN2	3.23(1.16)	3.27	3.34	3.11	3.21	57.9
CWAN-T3 (Ours)	StyleGAN2	**3.40(1.1)**	**3.30**	**3.61**	**3.33**	**3.35**	**64.0**

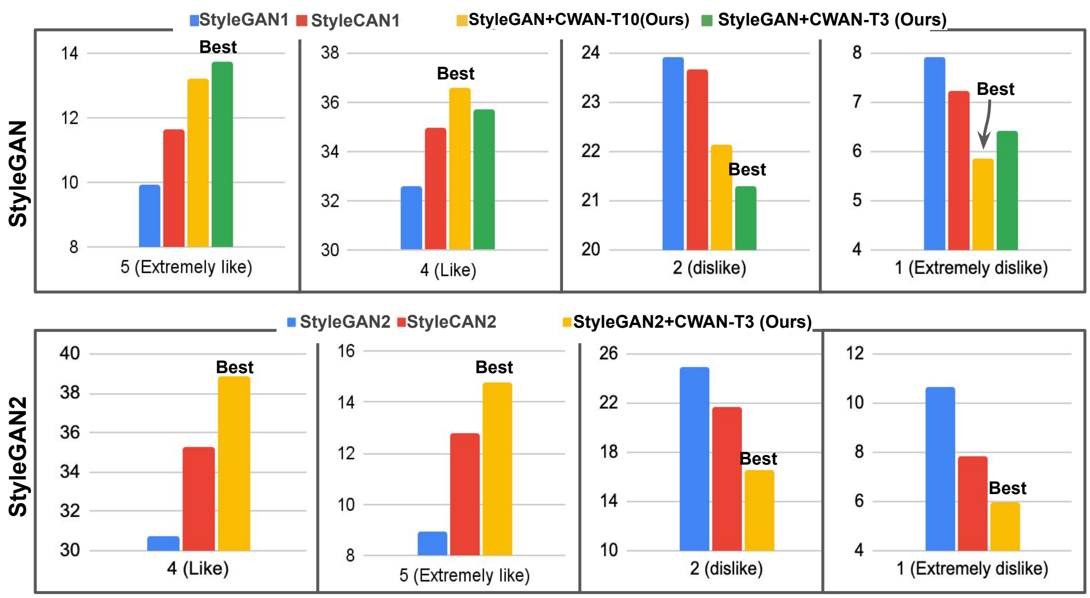

Figure 4: Percentage of each rating from human subject experiments on generated images. Compared to CAN, images generated using CWAN are rated (5,4) by a significantly larger share of people, and are rated (1,2) by fewer people.

Amazon Mechanical Turk by asking the surveyors the following questions.

(a) **Q1.** How much do you like this image? 1-5 rating; 5 is best rating.

(b) **Q2.** Do you think this image was created by artist or generated by computer? (yes/no)

The user interface of this experiment is shown in Figure 5 (top). We divide the generations into four groups described in nomenclature. We collect five responses for each art piece (400 images), totaling 2000 responses per model by 341 unique workers. Table 1 summarizes the likeability of CWAN generated artworks in comparison to vanilla GAN and StyleCAN variants (Jha, Chang, and Elhoseiny 2021). We find that images generated from our model is more likeable in all the groups described earlier. Figure 4 shows how our paintings are given higher ratings by more share of participants and lower ratings by less participants. We found that artworks from the trained StyleGAN1 and StyleGAN2 with our Creative Walk loss were more likeable and more people believed them to be real art, as shown in Table 1. For StyleGAN1, adding the Creative Walk loss resulted in 38% and 18% more people giving a full rating of 5 over vanilla StyleGAN1 and StyleGAN1 + CAN (StyleCAN1) loss, respectively, see Fig. 4. For StyleGAN2, these improvements are 65% and 15%. Table 2 shows that images generated by CWAN on StyleGAN1 and StyleGAN2 architectures have better ranks when combined with sets from other baselines.

2. **Comparison Experiment:** We performed experiments where given an artwork from a model trained with our Creative Walk loss vs an artwork with CAN loss, we ask people, which one they prefer. The pairing of the im-

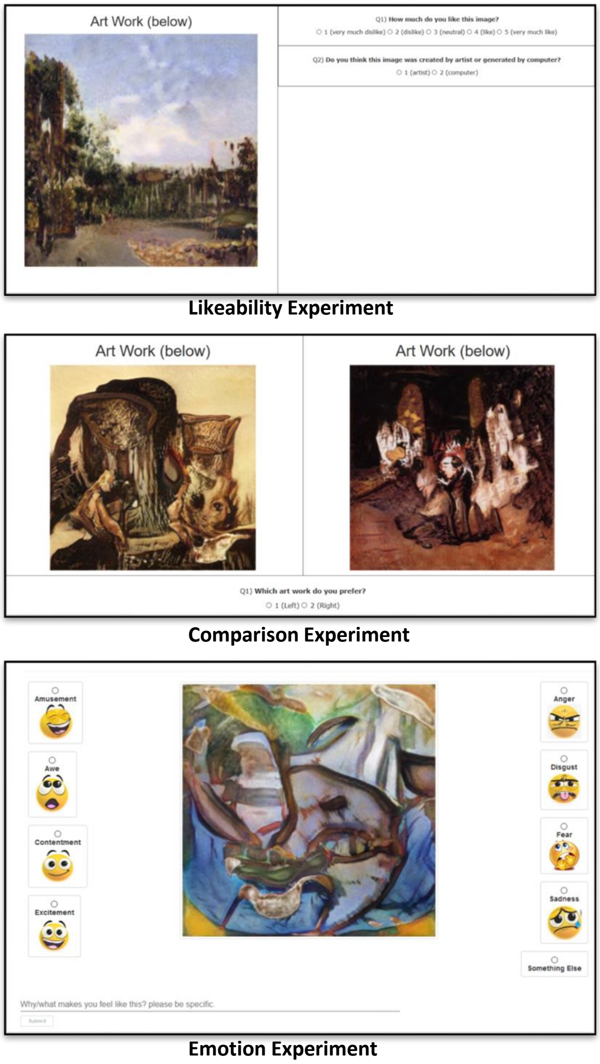

Figure 5: User interfaces of the likeability experiment(top), comparison experiment(middle) and emotion experiment(bottom).

Table 2: Normalized mean ranking (lower the better) calculated from the likeability experiment. We take the mean rating of each artwork on both CAN and CWAN losses. We then stack, sort, normalize them to compute the normalized mean rank. The numbers are corresponding normalized ranks from the models in the row above them.

	Normalized Mean Ranks		
	CAN/**CWAN-T10**	CAN/**CWAN-T3**	CAN/**CWAN-T10**/**CWAN-T3**
StyleGAN1	0.53/**0.47**	0.53/**0.47**	0.52/**0.48**/**0.50**
	CAN/**CWAN-T3**	GAN/**CWAN-T3**	CAN/GAN/**CWAN-T3**
StyleGAN2	0.54/**0.46**	0.59/**0.41**	0.49/0.59/**0.42**

ages was done on the basis of nearest neighbour. So, for each image generated from a StyleGAN model trained on Creative Walk loss, we found the nearest neighbour from images of model trained on CAN loss. Several qualitative results from these experiments are shown in Figure 6. The nearest neighbour was computed based on features that were extracted from a pretrained ResNet-18 (He et al. 2016). This is to make sure that the images we give out for comparison looks similiar as possible. We took random pairs of images from generations from StyleGAN model trained with CAN and CWAN; see the user interface for this experiment in Figure 5 (middle). The results for this experiment on StyleGAN 1 and 2 model on CWAN and CAN losses are summarized in Table 3. We collected 5 responses each for 600 pairs of artworks by 300 unique workers. Table 3 shows that CWAN loss is significantly more preferred compared to art work from CAN losses.

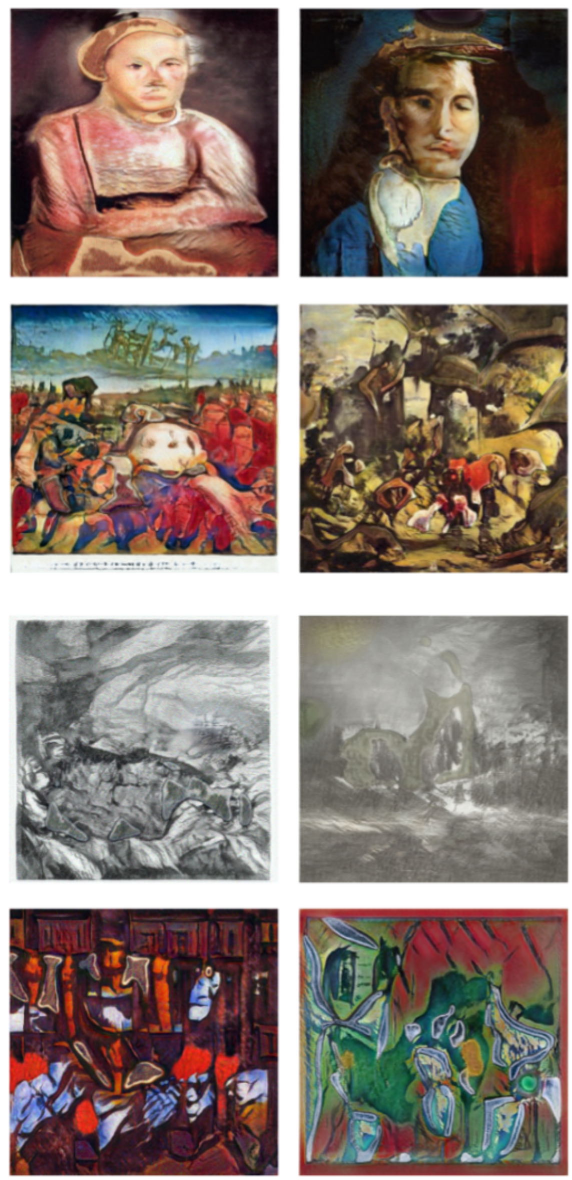

Figure 6: Figure shows CWAN (left) preferred more than CAN (right) for each pair of columns (random selection).

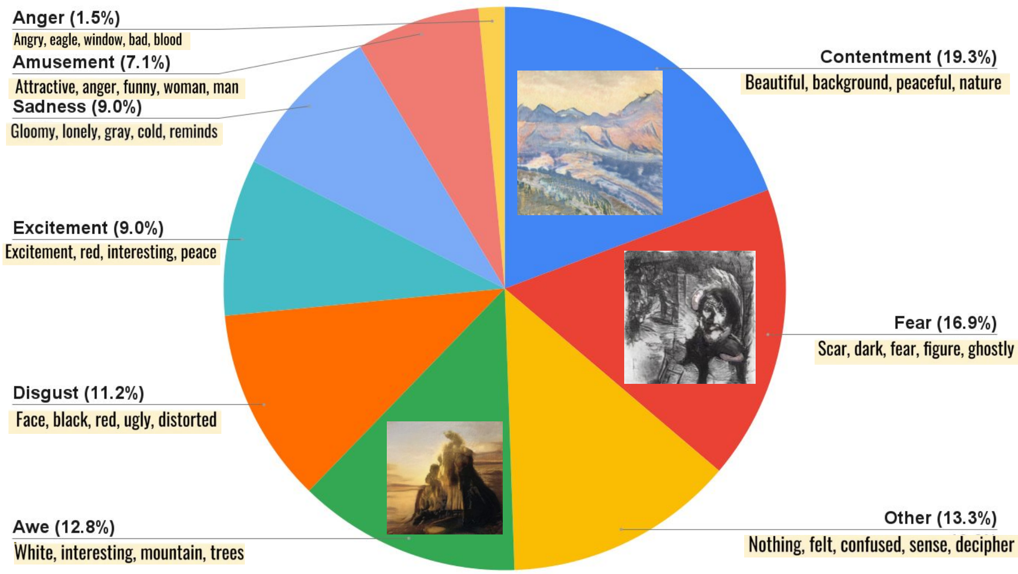

Figure 7: Distribution of emotional responses for generated art from StyleGAN1 + CWAN. Example image for fear, awe, and contentment is shown. The box beneath shows the most frequent words used by evaluators to describe their feeling. These responses were collected from a survey on Amazon Mechanical Turk.

3. **Emotion Human Subject Experiments:** To measure the emotional influence of AI generated art on the participants similar to (Jha, Chang, and Elhoseiny 2021), we asked participants to record their constructed emotion when exposed to a generated artwork. Following (Machajdik and Hanbury 2010; Achlioptas et al. 2021; Mohamed et al. 2022), we allowed these options of emotion categories 1) Amusement 2) Awe 3) Contentment 4) Excitement 5) Anger 6) Disgust 7) Fear 8) Sadness and 9) Something Else ("Other" in Fig 7). People were also asked to describe why they feel that particular emotion in text, so that survey participants chose the emotion after properly looking at the art work; see the user interface Figure 5 (bottom). We collected 5 responses each for a set of 400 generated artworks from 250 unique workers. Despite the model being trained unconditionally, it was able to produce generations that constructed diverse feelings in the viewer. Fig. 7 shows the distribution over the opted emotions, which are diverse but mostly positive. However, some generations construct negative emotions like fear. Fig. 7 also shows the most frequent words for each emotion after removing stop words. Notable positive words include "funny", "beautiful", "attractive", and negative words include "dark", "ghostly" which are associated with feelings like fear and disgust. Compared to the emotion experiments on Real Art and StyleCAN reported in (Jha, Chang, and Elhoseiny 2021), emotional responses to StyleGAN +CWAN art are more entropic (diverse).

Table 3: Evaluator preference percentage for generated images for both CWAN and CAN loss on the StyleGAN architectures. We split the preferred images into two groups based on their NN distance, and then the preference percentage is calculated for these groups.

	Architecture	Low NN distance split	High NN distance split
CAN	StyleGAN1	0.46	0.48
CWAN-T10	StyleGAN1	**0.54**	**0.52**
CAN	StyleGAN2	0.46	0.43
CWAN-T3	StyleGAN2	**0.54**	**0.56**

Emotional Descriptions by people. In Fig. 9, we can see a sample of the emotional descriptions that we collected on the art generated by CWAN in the emotion human subject experiment. One of the interesting descriptions we collect by a survey participant where they describe an artwork with a old looking female as "Zombie grandma". Another survey participant describes a artwork generated as "super relaxing" because of the sunset like colors in the artwork. More examples are shown in Fig. 9

Wundt Curve Analysis. Wundt curve (Packard 1975; Wundt 1874) illustrates Collin Martinale's "The principle of least efforts", a theory that explains human behavior towards creativity in artworks (Martindale 1990). The curve shows that as the originality/novelty of the work increases, people like the work. After a certain threshold, people start disliking it due to the difficulty of understanding, which leads to a lack of appreciation. We approximate Wundt curve by fitting a degree 3 polynomial on a scatter plot of normalized likeability vs. mean NN distance (novelty measure). Generations are more likable if the deviation from existing art is moderate but not too much; see Fig. 8. We observe that likeability responses to image sets with higher NN distance (i.e., Random (R) and NN↑) are generally lower compared to NN↓. Compared to CAN and GAN, CWAN achieves on

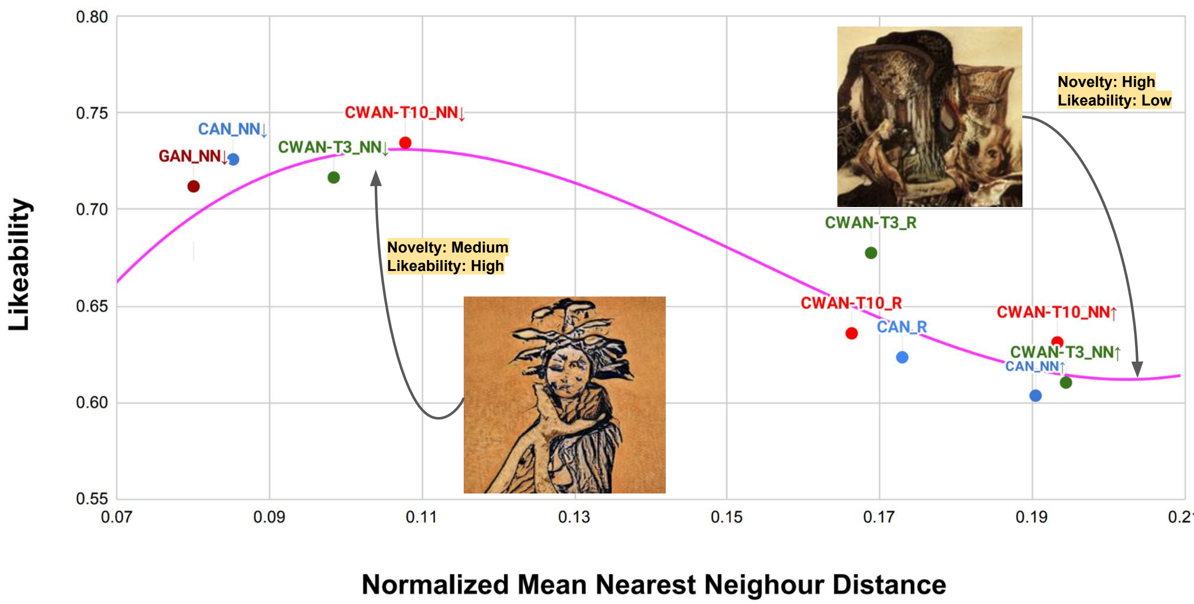

Figure 8: Empirical approximation of Wundt Curve (Packard 1975; Wundt 1874). The color of the data point represents a specific model and its label specifies the group named according to nomenclature. Art from the NN ↑ group has lower likeability than the NN ↓ group. Examples of a high and low likeability artwork and its novelty are shown. The NN distance is computed from features of resnet-18 and are normalized by scaling down by 20 (to be < 1). We select 20 because it was around the higher NN distances we observe in our generations

Figure 9: Descriptions given by people when asked to describe the why they felt a particular emotion while looking at artworks generated by CWAN (our method)

balance novel images that are more preferred.

Key Observations

In the experiments and the analysis conducted above, we noted the following key observations.

1. *The creative walk loss used in CWAN has performed better than CAN on two SOTA base architectures i.e. StyleGAN1 and StyleGAN2.*
2. *From Table 1 we find that the artworks generated by our proposed CWAN model are more likeable than those artworks by CAN in all the evaluation groups.*
3. *From Fig. 3 we see that artworks by CWAN have a significantly higher percentage of people giving a rating of 5 and least amount for people giving a rating of 1.*
4. *In Fig. 8, we approximated the Wundt Curve from artworks generated from CWAN.*
5. *The generated artworks were able to construct meaningful and diverse emotional experiences for our human participants as shown in Figures 7 and 9*

Conclusion

We propose Creative Walk Adversarial Networks. Augmenting Generative Adversarial Networks with a creative random walk loss. Through our experiments and analysis, we showed that CWAN improves generative models' capability to better represent the unseen artistic space and generate preferable novel artworks. We think the improvement is due to our learning mechanism's global nature, which operates at the minibatch level producing generations that are message-passing to each other to facilitate better deviation of generated artworks from seen art style classes.

Author Contributions

Mohamed Elhoseiny came up with the idea and Divyansh Jha was involved more in the implementation. Both Divyansh and Mohamed invested time in writing different sections of the paper. Where Mohamed helped in the theoretical aspects, Divyansh worked on experiments, figures and observations. Mohamed also played a key role in validation of results and observations. The other two authors Kai and Ivan helped us in paper writing for some duration of the life of the project.

Acknowledgements

This work is funded KAUST BAS/1/1685-01-01. We thank the IBEX computing cluster team at KAUST for their continual support in providing us resources and helping with technical challenges. We also wish to emphasize our gratitude to all the hard working Amazon Mechanical Turkers without whom this work would not be possible.

References

[Achlioptas et al. 2021] Achlioptas, P.; Ovsjanikov, M.; Haydarov, K.; Elhoseiny, M.; and Guibas, L. J. 2021. Artemis: Affective language for visual art. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11569–11579.

[Ayyad et al. 2020] Ayyad, A.; Navab, N.; Elhoseiny, M.; and Albarqouni, S. 2020. Semi-supervised few-shot learning with prototypical random walks.

[Briot, Hadjeres, and Pachet 2017] Briot, J.-P.; Hadjeres, G.; and Pachet, F. 2017. Deep learning techniques for music generation-a survey. *arXiv:1709.01620*.

[Date, Ganesan, and Oates 2017] Date, P.; Ganesan, A.; and Oates, T. 2017. Fashioning with networks: Neural style transfer to design clothes. In *KDD ML4Fashion workshop*.

[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.

[Dumoulin et al. 2017] Dumoulin, V.; Shlens, J.; Kudlur, M.; Behboodi, A.; Lemic, F.; Wolisz, A.; Molinaro, M.; Hirche, C.; Hayashi, M.; Bagan, E.; et al. 2017. A learned representation for artistic style. *ICLR*.

[Elgammal et al. 2017a] Elgammal, A.; Liu, B.; Elhoseiny, M.; and Mazzone, M. 2017a. Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. *arXiv preprint arXiv:1706.07068*.

[Elgammal et al. 2017b] Elgammal, A.; Liu, B.; Elhoseiny, M.; and Mazzone, M. 2017b. Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. In *International Conference on Computational Creativity*.

[Elhoseiny and Elfeki 2019] Elhoseiny, M., and Elfeki, M. 2019. Creativity inspired zero-shot learning. In *Proceedings of the IEEE International Conference on Computer Vision*, 5784–5793.

[Elhoseiny, Yi, and Elfeki 2021] Elhoseiny, M.; Yi, K.; and Elfeki, M. 2021. Cizsl++: Creativity inspired generative zero-shot learning. *arXiv preprint arXiv:2101.00173*.

[Gatys, Ecker, and Bethge 2016] Gatys, L. A.; Ecker, A. S.; and Bethge, M. 2016. Image style transfer using convolutional neural networks. In *CVPR*.

[Ge et al. 2019] Ge, S.; Dill, A.; Kang, E.; Li, C.-L.; Zhang, L.; Zaheer, M.; and Poczos, B. 2019. Developing creative ai to generate sculptural objects. *arXiv preprint arXiv:1908.07587*.

[Goodfellow et al. 2014a] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014a. Generative adversarial nets. In *NIPS*, 2672–2680.

[Goodfellow et al. 2014b] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014b. Generative adversarial networks. In *NIPS*.

[Ha and Eck 2018] Ha, D., and Eck, D. 2018. A neural representation of sketch drawings. *ICLR*.

[Haeusser, Mordvintsev, and Cremers 2017] Haeusser, P.; Mordvintsev, A.; and Cremers, D. 2017. Learning by association — a versatile semi-supervised training method for neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 626–635.

[He et al. 2016] He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *CVPR*.

[Hertzmann 2018] Hertzmann, A. 2018. Can computers create art? In *Arts*, volume 7, 18. Multidisciplinary Digital Publishing Institute.

[Isola et al. 2017] Isola, P.; Zhu, J.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. *CVPR*.

[Jha, Chang, and Elhoseiny 2021] Jha, D.; Chang, H. H.; and Elhoseiny, M. 2021. Wölfflin's affective generative analysis of visual art. *The International Conference on Computational Creativity (ICCC)*.

[Johnson, Alahi, and Li 2016] Johnson, J.; Alahi, A.; and Li, F. 2016. Perceptual losses for real-time style transfer and super-resolution. *ECCV*.

[Karras et al. 2018] Karras, T.; Aila, T.; Laine, S.; and Lehtinen, J. 2018. Progressive growing of GANs for improved quality, stability, and variation. *ICLR*.

[Karras et al. 2020] Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8110–8119.

[Karras, Laine, and Aila 2019a] Karras, T.; Laine, S.; and Aila, T. 2019a. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4401–4410.

[Karras, Laine, and Aila 2019b] Karras, T.; Laine, S.; and Aila, T. 2019b. A style-based generator architecture for generative adversarial networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).*

[Kingma and Welling 2013] Kingma, D. P., and Welling, M. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

[Li et al. 2019] Li, X.; Sun, Q.; Liu, Y.; Zhou, Q.; Zheng, S.; Chua, T.-S.; and Schiele, B. 2019. Learning to self-train for semi-supervised few-shot classification. In *Advances in Neural Information Processing Systems*, 10276–10286.

[Machado and Cardoso 2000] Machado, P., and Cardoso, A. 2000. Nevar–the assessment of an evolutionary art tool. In *Proc. of the AISB00 Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science*, volume 456.

[Machajdik and Hanbury 2010] Machajdik, J., and Hanbury, A. 2010. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the 18th ACM international conference on Multimedia*, 83–92.

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

[Mohamed et al. 2022] Mohamed, Y.; Khan, F. F.; Haydarov, K.; and Elhoseiny, M. 2022. It is okay to not be okay: Overcoming emotional bias in affective image captioning by contrastive data collection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

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

[Nobari, Rashad, and Ahmed 2021] Nobari, A. H.; Rashad, M. F.; and Ahmed, F. 2021. Creativegan: editing generative adversarial networks for creative design synthesis. *arXiv preprint arXiv:2103.06242*.

[Packard 1975] Packard, S. 1975. Aesthetics and psychobiology by de berlyne. *Leonardo* 8(3):258–259.

[Radford, Metz, and Chintala 2016] Radford, A.; Metz, L.; and Chintala, S. 2016. Unsupervised representation learning with deep convolutional generative adversarial networks. *ICLR*.

[Reed et al. 2016] Reed, S. E.; Akata, Z.; Mohan, S.; Tenka, S.; Schiele, B.; and Lee, H. 2016. Learning what and where to draw. In *NIPS*.

[Ren et al. 2018] Ren, M.; Ravi, S.; Triantafillou, E.; Snell, J.; Swersky, K.; Tenenbaum, J. B.; Larochelle, H.; and Zemel, R. S. 2018. Meta-learning for semi-supervised few-shot classification. In *International Conference on Learning Representations*.

[Sbai et al. 2018] Sbai, O.; Elhoseiny, M.; Bordes, A.; LeCun, Y.; and Couprie, C. 2018. Design: Design inspiration from generative networks. In *ECCV workshop*.

[Tendulkar et al. 2019] Tendulkar, P.; Krishna, K.; Selvaraju, R. R.; and Parikh, D. 2019. Trick or treat: Thematic reinforcement for artistic typography. In *ICCC*.

[WikiArt 2015] WikiArt, O. 2015. Wikiart dataset. https://www.wikiart.org/. Accessed: 2020-05-30.

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

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

[Zhang et al. 2018] Zhang, R.; Che, T.; Ghahramani, Z.; Bengio, Y.; and Song, Y. 2018. Metagan: An adversarial approach to few-shot learning. In *Advances in Neural Information Processing Systems*, 2371–2380.