Wölfflin's Affective Generative Analysis for Visual Art

Divyansh Jha **Hanna H Chang** **Mohamed Elhoseiny**
King Abdullah University of Sience and Technology (KAUST)
{divyansh.jha, hanna.chang, mohamed.elhoseiny}@kaust.edu.sa,

Abstract

We propose *Wölfflin Affective Generative Analysis (WAGA)* as an approach to understand and analyze the progress of machine-generated artworks in contrast to real art and their connection to our human artistic heritage, and how they extend the shape of art history. Specifically, we studied the machine-generated art after integrating creativity losses in the state-of-the-art generative models e.g., StyleGAN v1 and v2. We denote these models as Style Creative Adversarial Networks v1 and v2; in short, StyleCAN v1 and v2. We contrasted the learned representation between real and generated artworks through correlation analysis between constructed emotion (collected through Amazon MTurk) and Heinrich Wölfflin (1846-1945)'s principles of art history. Analogous to the recent ArtEmis dataset, we collected constructed emotions and explanations on generated art instead of real art to study the contrast. To enable Wölfflin Affective Generative Analysis, we collected 45,000 annotations (1800 paintings $\times 5$ principles $\times 5$ participants) for each of the five Wölfflin principles on 1800 artworks; 1000 real and 800 generated. Our analysis shows a correlation exists between the Wölfflin principles and the emotions. The collected dataset, analysis, and code is made publicly available at https://vision-cair.github.io/WAGA.

Introduction

With the development of computational creativity, machines are capable of classifying real artworks styles. However, the machine's ability to assess and classify its AI-generated artworks is less understood and requires further scholarly scrutiny (Colton 2008). *The main question we address in this paper is to quantitatively and qualitatively analyze the contrast between real and generated artworks from deep neural representations' perspectives.* We study AI-generated art in three analysis dimensions: 1) Likeability evaluated by human ratings, 2) Their learning representation connection to Wölfflin's principles (Wölfflin 1915), 3) emotions constructed by human participants. In contrast to (Elgammal et al. 2018), which used only real art for its analysis, our study focuses on the contrast between real and AI art generated using state-of-the-art GAN models, i.e., StyleGAN1 (Karras, Laine, and Aila 2019), and StyleGAN2 (Karras et al. 2020). We add the CAN loss (Elgammal et al. ; Sbai et al. 2018) to these architectures. We denote the corresponding models as StyleCAN1 and StyleCAN2. We also collect data of Wölfflin's principles on 1000 real art pieces and 800 generated art pieces from StyleGAN2 and StyleCAN2. Results of our study provide insights into the emotion of generated artworks. Figure 1 shows examples of AI-generated artwork with StyleCAN2 that constructs emotional experiences in survey participants.

Disgust: For me this is a harvest machine or something like that. However, it is kind of irritating to look at this painting where one can´t really know if this is a machine or not.
Something Else:The deserted vehicle near the sea creates a lonely feeling,
Amusement: sea shore where a roadside shop is left for a long time
Fear: The water looks turbulent and rough in the background.
Contentment: The white and greys are the water, capture the seas rage.

Figure 1: AI art constructing diverse emotional experiences.

Contribution: (1) We introduce StyleCAN v1 and v2 by integrating the CAN loss StyleGAN v1 and StyleGAN v2 models and observe that StyleCAN v1 and v2 have higher mean average likeability compared to StyleGAN v1 and StyleGAN v2. (2) We present a novel study on how AI Art generative models learn inherent features of our art heritage like Wölfflin's principles. We also study the ability of these models to constructs our emotional experiences compared to real art. (3) We collect Wölfflin principles annotations on real and AI art. We also collect emotion labels and their explanation on AI art. (4) Using the collected data, we performed detailed analysis that contrast real art and AI art based on Wölfflin's principles, constructed emotion categories, and corresponding explanations. We also observed connections between Wölfflin's principles and the constructed emotional experiences. Since we study both the Wölfflin's and the affective perspective for visual art, we denote our approach to analyse visual art

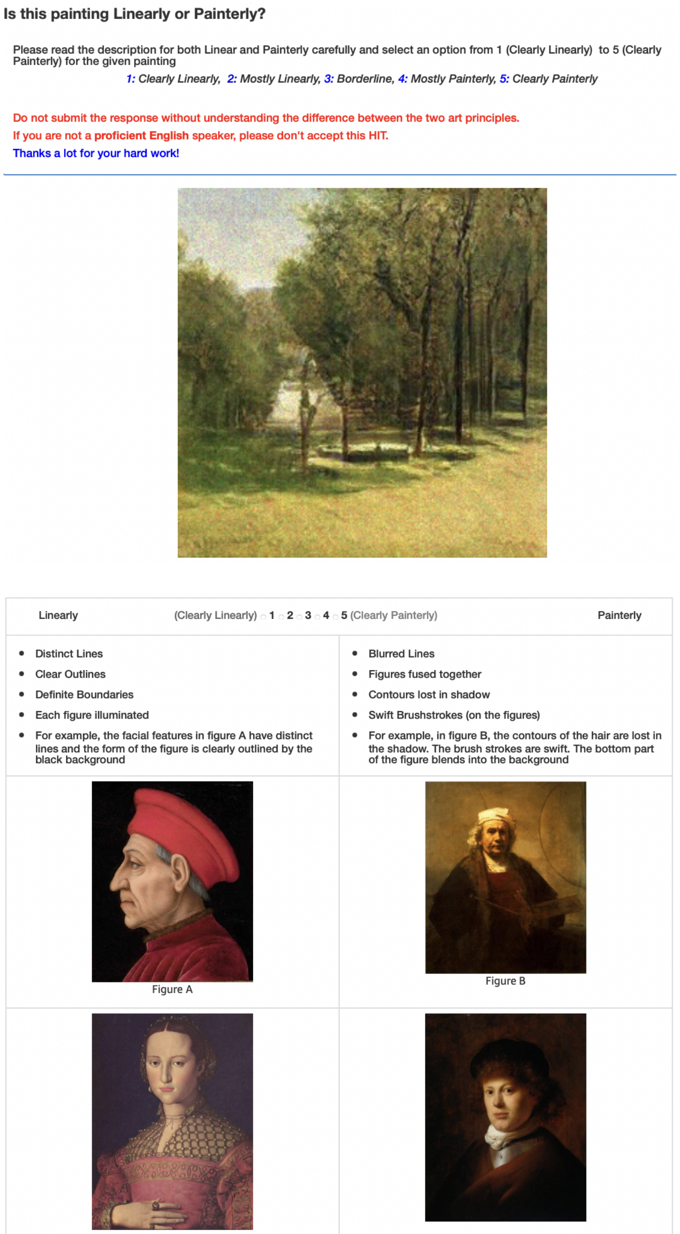

Figure 2: The data collection interface of Wölfflin's principles showing the examples used to train surveyors to classify between the two opposing concepts: linearly and painterly.

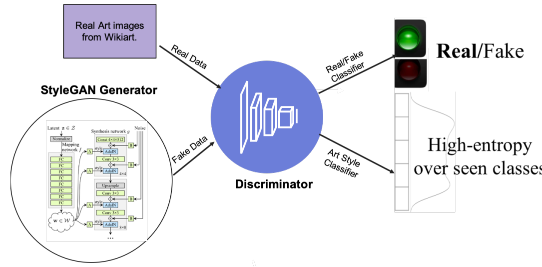

Figure 3: GAN architecture (StyleGAN1 or StyleGAN2) after adding CAN loss.

as **Wölfflin's Affective Generative Analysis (WAGA)** and hope this may encourage future comprehensive analysis of machine-generated art.

Related Work

Existing literature on AI for the art creation process has shifted from being emulative to being more creative (Elgammal et al. ; Hertzmann 2020; Sbai et al. 2018). Although recent creative AI models can produce novel quality artworks, it is less understood whether these models have all the characteristics of a creative system. (Colton 2008) defined three main characteristics that creative systems should have: (a)"the ability to produce novel artifacts (imagination), (b) "the ability to generate quality artifacts", and (c) "the ability to assess its creation". Wölfflin's Principles of Art History (Wölfflin 1915) are one of the key methodologies in art history that differentiates art styles. They have five categories to classify the stylistic component of the painting: 1) Linear and Painterly, 2) Planar and Recessional, 3) Closed Form and Open Form, 4) Multiplicity and Unity, and 5) Absolute Clarity and Relative Clarity. (Elgammal et al. 2018) has demonstrated a connection between Wölfflin's Principles and machine's learning representations of artworks. Specifically, Wölfflin's Principles were shown to be implicitly learning that each principle was shown to have a strong correlation by one or more neurons in their Neural Network. Our work extends this analysis by collecting performing Wölfflin's Principles analysis on Machine generated artworks.

The theory of Constructed Emotions (Barrett 2017) suggested that emotions are constructed rather than triggered. In line with the theory, (Achlioptas et al. 2021) collected responses of emotions constructed by Human participants who gett exposed to real artworks from the WikiArt dataset. Our work also aims at understanding how people construct emotions from visual art created by AI and contrast that to real art. Hence, we collected emotional responses for generated art as well.

Data Collection

Heinrich Wölfflin proposed five principles for visual art (Wölfflin 1950):

1. **Linearly and Painterly**: Linearly paintings depict isolated objects and clear boundaries and have all the figures illuminated. Painterly depicts blurry outlines and swift brushstrokes.
2. **Planar and Recessional**: The composition of objects in planar are arranged in planes parallel to the plane of the canvas. In recessional, these objects can be in angle and focus on spatial depth.
3. **Closed-form and Open-form**: All figures in closed-form are balanced within the frame, while in open-form, the figures are cut off. While the former is mostly self-contained, the latter indicates space beyond the frame.
4. **Multiplicity and Unity**: In multiplicity, we have distinguished parts, and each part demonstrates independent

Table 1: Pearson's correlation coefficients of features of various architectures on real art and generated art computed for all the Wölfflin principles. The term "vs" is used in the table to compare opposing concepts of each Wölfflin principle.

Architecture	**Linearly**	**vs**	**Painterly**	**Planar**	**vs**	**Recessional**	**Closed Form**	**vs**	**Open Form**	**Multiplicity**	**vs**	**Unity**	**Absolute Clarity**	**vs**	**Relative Clarity**
	Real art	StyleGAN2	StyleCAN2	Real art	StyleGAN2	StyleCAN2	Real art	StyleGAN2	StyleCAN2	Real art	StyleGAN2	StyleCAN2	Real art	StyleGAN2	StyleCAN2
ResNet101 + 2	-0.20	-0.29	-0.27	-0.15	-0.22	0.21	0.13	-0.16	0.14	-0.15	0.15	-0.24	-0.26	0.17	0.14
ResNet101	0.29	-0.39	-0.37	0.26	-0.32	-0.24	-0.33	-0.27	**-0.31**	0.35	0.21	0.42	0.54	**-0.44**	0.18
ResNet50 + 2	0.20	-0.41	-0.25	0.10	-0.26	0.15	-0.13	-0.16	-0.16	0.15	0.15	0.26	-0.18	0.29	0.12
ResNet50	0.33	0.36	0.32	0.32	0.28	**0.31**	-0.33	0.25	0.28	**0.36**	-0.24	0.42	0.67	0.44	-0.12
VGG16 + 2	-0.20	-0.18	0.25	-0.17	-0.17	0.18	-0.19	-0.16	-0.11	0.18	-0.12	-0.16	-0.10	-0.12	0.13
VGG16	0.41	0.38	**-0.44**	-0.18	0.21	-0.21	0.35	0.27	-0.19	-0.26	0.17	0.27	0.41	0.22	-0.13
StyleGAN1 Disc	-0.19	-0.30	-0.34	-0.22	-0.20	-0.17	-0.42	-0.24	-0.22	-0.26	-0.14	-0.40	-0.38	-0.28	-0.14
StyleCAN1 Disc	-0.30	-0.39	-0.47	-0.18	**-0.33**	-0.24	-0.32	-0.24	-0.21	-0.28	**-0.31**	0.38	0.41	-0.32	0.19
StyleGAN2 Disc	**-0.36**	**-0.47**	-0.42	-0.26	-0.28	-0.21	-0.37	-0.26	-0.23	0.31	0.23	0.42	**-0.69**	0.33	0.20
StyleCAN2 Disc	-0.27	0.33	-0.29	**0.35**	0.26	-0.26	**0.43**	**0.29**	-0.26	0.35	0.22	**-0.47**	0.56	0.35	**-0.24**

features. In unity, figures weld together, and colors blend in.

5. **Absolute clarity and Relative clarity**: While absolute clarity has realistic representation, relative clarity has representations enhanced with visual effects.

We collected Wölfflin's principles annotations by training people to learn one of these principles and then ask them to identify it in a painting. For example, the survey interface design provides descriptions of Wölfflin's principles (one at a time). Figure 2 shows the interface design for identifying linearly and painterly paintings. Based on the shown explanation of linearly and painterly paintings, the viewer selects a rating scale from 1 to 5 (1: Clear Linearly, 2: Mostly Linearly, 3: Borderline, 4: Mostly Painterly, 5: Clear Painterly). Using this method, we train the survey participants. We average the five ratings for every artwork and normalize the resulting score between 0 and 1, where normalized scores closer to 0 are linear, and scores closer to 1 define painterly characteristics. This way, an artwork has five floating-point numbers corresponding to its five Wölfflin principles. We conduct these experiments for both real and generated arts for each Wölfflin principle. We release the web interfaces for collection of the Wölfflin principles here

Methodology

Generative Adversarial Networks (GANs) (Goodfellow et al. 2014; Radford, Metz, and Chintala 2015; Ha and Eck 2018) is a popular modeling choice. However, the classic GAN training objective does not promote the generation of novel content beyond the training data. A GAN trained on artwork can generate Da Vinci's "Mona Lisa" again, but it will not produce a painting of a new style. Recent work has been able to encourage GANs to produce novel images. Inspired by (Elgammal et al. ; Sbai et al. 2018), we adapted GANs to generate novel paintings by encouraging the model to deviate from existing art styles. We attach a head on the GAN's Discriminator D, which predicts the *style* of an art piece. The Generator is then encouraged to generate real-looking examples, which is hard for D to assign a class.

StyleCAN Model (Figure. 3): We train a StyleGAN model (Karras, Laine, and Aila 2019), (Karras et al. 2020) using the creativity loss (Sbai et al. 2018) on the WikiArt dataset. In contrast, (Sbai et al. 2018) experiments it for fashion dataset. Concretely, our generator loss becomes:

$$\mathcal{L}_G = \mathcal{L}_{G\ \text{StyleGAN}} + \lambda \mathcal{L}_{G\ \text{creativity (Sbai et al. 2018)}} \tag{1}$$

$$\mathcal{L}_D = \mathcal{L}_{D\ \text{StyleGAN}} + \lambda \mathcal{L}_{G\ \text{style classification}} \tag{2}$$

We denote the resulting model as StyleCAN. StyleGAN v1 and v2 are then dubbed as StyleCAN v1 and v2.

Experiments

Generated Art Setup: Initially, we generate a set of 10,000 paintings from each trained GAN model. We then select 400 images representing every model and divide them into four groups containing 100 generated art pieces. (1) Highest Nearest Neighbour (NN $\uparrow$) - We computed the NN on the WikiArt dataset and selected the top 100 with the highest NN distance from the closest training image. (2) Lowest NN (NN $\downarrow$)- We selected another 100, which had the lowest NN distance from the closest training image. (3) Highest shape entropy (Entropy $\uparrow$) - We selected the artworks with the highest confusion in art style computed from a trained style classifier. (4) Random - A set of 100 random images.

Real Art Setup: For real art pieces, we selected around 170 art pieces from each century from 1400 to 2000 totaling 1000 real artworks. For these 1000 art pieces, we computed features from several different models trained on the art style classification task. We used ResNet50, ResNet101, VGG16 base architectures. Following the literature (Elgammal et al. 2018), we also added two additional layers and fine-tuned it further and we denote them as the $+2$ versions of these models. Also, we used the trained discriminators from StyleGAN models with and without CAN loss.

We conducted the following survey experiments, with 5 responses collected for each example, covering a total 500 participants.

Likeability Experiment: We follow (Elgammal et al.) for likeability experiment. We ask participants two questions. Q1) Rate the art on the scale of 5. Q2) Whether art is created by an artist or machine (we name this Turing Test).

Emotion Experiment: Follow (Achlioptas et al. 2021) setup, we ask participants to select one of 9 emotions (amusement, awe, contentment, excitement, anger, disgust, fear, sadness, and something else). We also ask them to describe in text what made them feel so.

Results

In Table 3, we find that participants prefer generations from models integrated with CAN loss more than the generations from vanilla versions. The mean likeability of StyleCAN1 is 2.5% more than its vanilla counterpart. For StyleCAN2, it is 7.0%. We also find that more people think that an artist creates generations from the model trained with CAN loss for both StyleGAN1 and StyleGAN2 models. The generated images achieved a higher likeability score than Art Basel.

Classifier features' principal components correlation with Wölfflin's principles: We find that learning to classify style makes the model also learn Wölfflin's principles inherently. We compute the Pearson's correlation coeffi-

Table 2: Weights of the linear classifier when trained on Wölfflin's principle to different emotions.The term "vs" is used in the table to compare opposing concepts of each Wölfflin principle.

Emotion	Linearly vs Painterly	Planar vs Recessional	Closed Form vs Open Form	Multiplicity vs Unity	Absolute Clarity vs Relative Clarity
Amusement	-0.512	0.218	0.101	**-1.235**	**-0.684**
Anger	**-0.532**	0.043	-0.079	-0.316	**0.996**
Awe	**-0.287**	0.11	**-0.386**	-0.108	0.276
Contentment	0.224	**-0.92**	**0.556**	0.284	-0.449
Disgust	0.096	**0.345**	-0.304	-0.001	**-0.49**
Excitement	0.098	-0.281	**0.992**	-0.045	**0.369**
Fear	**-0.701**	0.392	**-0.567**	0.72	-0.018
Sadness	-0.245	-0.062	**-0.413**	0.39	**-0.392**
Overall Negative	-0.018	**0.804**	**-0.612**	0.41	0.164
Overall Positive	0.073	-0.464	**0.75**	**-0.476**	-0.232

cient using the principal components of the artwork features. We compute this coefficient for all the classifiers for both real and generated art (StyleGAN2 and StyleCAN2 generated art). Table 1 summarizes the maximum coefficients we computed in the top 10 principal components. There was almost no correlation with components after the 10^{th} component. We also observe that ResNet50 features' have significant correlation coefficients for all generations and Wölfflin's principles.

Wölfflin's Principles Correlation to Emotions: We trained a linear classifier to predict emotions from Wölfflin's pair values collected from our human subject experiment. We used the emotion labels of real art from (Achlioptas et al. 2021). We trained one linear classifier, one for each emotion, and collected the weights. Table 2 summarizes the resultant classifier weights. From the experiment setup, the averaged floating-point value ranging in 0-1 that we calculate for a Wölfflin's pair if its less then its more of the first value in pair if its more then its more of the second value. For example, in Linearly vs. Painterly, if an artwork has a value of 0.1, it means it's more "Linearly" than "Painterly" and vice-versa. People feel amused when the painting is more towards "Multiplicity" than "Unity." because of the negative correlation. Similarly, people feel anger when the artwork has more "Relative Clarity" than "Absolute Clarity."; see more correlations in Table 2.

Emotion Distribution: We plot group bar-charts of distribution of emotions for real and generated art produced by StyleGAN2 and StyleCAN2 in Figure 4. We observe that AI-generated art is capable of constructing diverse sets of emotions that are similar to real ones. However, there are some differences in the distribution. When compared to real art bar graph, we find that "Excitement" emotion increased from 8.3% to 18.1% for StyleCAN2, while maintaining a

Table 3: Human experiments on generated art from Vanilla GAN and CAN losses. Models trained on CAN loss have a higher mean likeability in all the groups. More people believed the generated art to be real for artwork generated from the model trained on CAN loss.

		Likeability Mean					Turing Test
Loss	Architecture	Q1-mean(std)	NN $\uparrow$	NN $\downarrow$	Entropy $\uparrow$	Random	Q2(% Artist)
CAN (Elgammal et al.)	DCGAN	3.20(1.50)	-	-	-	-	53
Abstract art	Human	3.3(0.43)	N/A	N/A	N/A	N/A	85
Art Basel	Human	2.8(0.6)	N/A	N/A	N/A	N/A	41
All Art sets	Human	3.1(0.63)	N/A	N/A	N/A	N/A	62
GAN (Vanilla)	StyleGAN	3.12(0.58)	3.07	3.36	3.00	3.06	55.33
CAN	StyleGAN	3.20(0.62)	3.01	**3.61**	3.05	3.11	56.55
GAN (Vanilla)	StyleGAN2	3.02(1.15)	2.89	3.30	2.79	3.09	54.01
CAN	StyleGAN2	**3.23(1.16)**	**3.27**	3.34	**3.11**	**3.21**	**57.9**

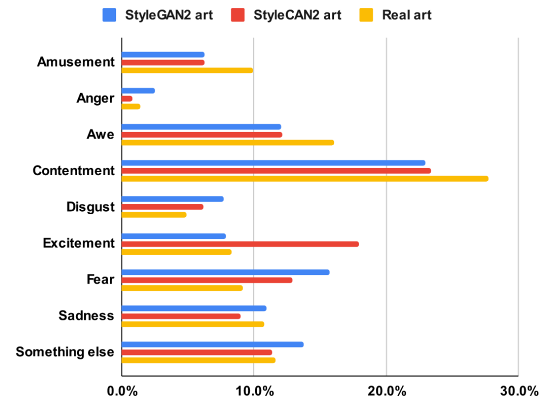

Figure 4: Distribution of emotions by emotion survey experiments on both real and generated art

similar percentage in StyleGAN2 (7.9%). Fear emotion percentage increased from 9.1% for real art to 15.8% for StyleGAN2 and 12.9% for StyleCAN2.

Likeability experiment qualitative analysis: We observe that AI-generated paintings with high likeability (Q1) score and high Turing test percentage (Q2) in table 3 were from NN↑ group. This shows that the generated is both novel (because of high NN distance) and likeable. We can see some examples in Figure 5. Artworks with high likeability in the NN ↓ group were mostly natural landscapes, vibrant colors, and distinct brushstrokes. They looked like Monet's "Impression Sunrise." NN ↑ group portrayed abstract figures with contrasting solid colors. Artworks from the NN ↓ group, which had a low Turing test percentage were abstract paintings, reminding the viewer of Mondrian's "Broadway Boogie Woogie." The style of these paintings may have prompted survey participants to assume a machine did these paintings, whereas artworks from NN ↑ with lower Turing test numbers looked like images seen under a microscope. The paintings' cell-like, scientific feel made the paintings appear like screenshots of microscopic scans, contributing to a lack of artistic intent.

Emotions experiments qualitative analysis: We derived and analyzed common elements of the AI-generated artworks that constructed various emotions.

Awe: Artworks that portrayed familiar subject matters like "brown coat", "man", "plant", and "sky" tend to construct emotions of awe within the audience. Many participants referred to the use of complementary colors and soft color scheme to be the underlying cause that constructed emotions of awe;*example response by our participants: "the play of colors in the sky of this painting is magnificent")*.

Anger: The disorderly arrangement of figures tends to create a sense of unease and construct anger (e.g., *"The painting looks confusing and shows no representation"*). The confusion caused by the abstract subject matter caused discomfort and points out the participants' annoyance in understanding the meaning of the painting (e.g., *"The color (red) used seems like a man bleeding with anger"*.

Contentment: We observed that floral color schemes constructed contentment emotions, as quoted by the participants:(e.g., *"mix of colors between green and yellow re-*

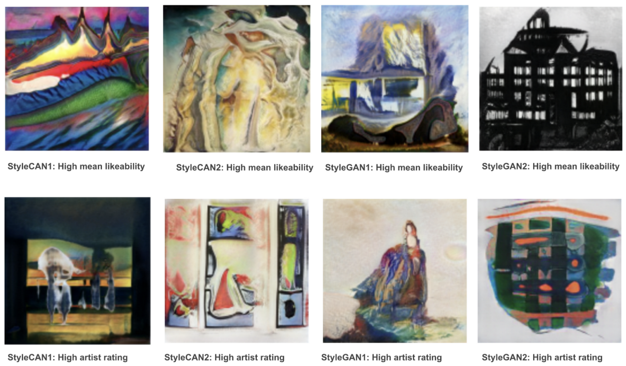

Figure 5: High likeability and high Turing test percentage artworks from NN $\uparrow$ group.

minds the changing of seasons, brings peace and tranquility"). Participants also underlined the role of depth, layers, and orderly composition played in constructing their awe feeling (e.g., *"This painting makes me feel relaxed because the items are well ordered and displayed in a coherent fashion"*). Positive past experiences evoked by the artworks contributed to the participant's selection of awe (e.g., *"I liked the image I felt pleasure because reminder my childhood"*).

Amusement: Artworks associated with feelings of amusement interestingly reminded the audience of animals *(e.g., "The white cat is hiding behind the building").* Participants were also amused by depiction of human characters. The portrayal of human subjects created both a sense of familiarity and beauty which contributed the construction of amusement e.g., "there seems to be a pretty girl dancing across this image".

Disgust: Dark color schemes and visual effects tend to construct emotions of disgust, as stated by the participants: "Too much darkness on the sea", and "looks like a dark cloud about to eat a human". Also, when an artwork didn't clearly convey its message or meaning, the audience felt emotions of disgust. Participants stated that the lack of expression came from "a lifeless representation".

Discussions

In our experiments, we notice the following key observations.

1. *From Table 1 we observe that the discriminator of GANs learn the Wölfflin principles inherently as we observe stronger correlation values as compared to traditional classifiers.*
2. *In Figure 4, we find that the generated art works construct diverse set of emotions in the viewer.*
3. *In Table 2, we also find strong correlation between these emotions and Wölfflin's principles. We find these principles can be used to predict the emotion of an art piece. Hence, being able to compute specific Wölfflin priciple can predict the emotion that the art piece will construct.*

Conclusion

We introduced *Wölfflin Affective Generative Analysis* as an approach to understand and analyze machine-generated artworks in contrast to real art. To perform this analysis, we collected Wofflin principles annotations on 1000 real paintings and 800 generated artworks. We also collected Affective responses on artworks produced by state-of-the-art generative models, including StyleGAN2 and our improved version of it StyleCAN2 after applying CAN loss. We show that models inherently learn the stylistic principles and emotions during the learning process. We showed that Wölfflin's principle coefficients are similar for generated and real artwork, showing that the generated art also contains the historical styles studied in the past. By training a multi-label classifier to predict emotion from Wölfflin principles, we observed that some Wölfflin principles have solid connections for constructing certain emotions. We release our models, analysis, and data on real and generated art to facilitate future research.

References

[Achlioptas et al. 2021] Achlioptas, P.; Ovsjanikov, M.; Haydarov, K.; Elhoseiny, M.; and Guibas, L. 2021. Artemis: Affective language for visual art. *CVPR*.

[Barrett 2017] Barrett, L. F. 2017. The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience*.

[Colton 2008] Colton, S. 2008. Creativity versus the perception of creativity in computational systems. In *AAAI spring symposium*.

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

[Elgammal et al. 2018] Elgammal, A.; Liu, B.; Kim, D.; Elhoseiny, M.; and Mazzone, M. 2018. The shape of art history in the eyes of the machine. In *AAAI*.

[Goodfellow et al. 2014] Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*.

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

[Hertzmann 2020] Hertzmann, A. 2020. Visual indeterminacy in gan art.

[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 *CVPR*.

[Karras, Laine, and Aila 2019] Karras, T.; Laine, S.; and Aila, T. 2019. A style-based generator architecture for generative adversarial networks. In *CVPR*.

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

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

[Wölfflin 1915] Wölfflin, H. 1915. *Principles of art history*.

[Wölfflin 1950] Wölfflin, H. 1950. Heinrich wölfflin principles of art history. `https://www.csus.edu/indiv/o/obriene/art192b/heinrich-wolffin_principles-of-art-history.pdf`. Accessed: 2021-03-10.