9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Exploring the Evolution of Artistic Styles Using Generative AI and Influence Modeling

Keywords: generative AI, cultural evolution, artistic concepts, computational creativity, historic creativity

Extended Abstract

Understanding the evolution of human creative expression is central to art history, and instrumental to the progress of algorithmic creativity [1, 2]. Recent advances in generative AI, such as Stable Diffusion, Midjourney, and DALL-E, show great promise in generating detailed images based on textual prompts blending visual concepts and art styles. However, whether these models can produce truly novel outputs beyond recombination remains unclear. Measuring creativity and cultural progress in subjective domains like the arts is challenging, but generative AI can help deconstruct art into distinct concepts, such as style, content, and composition, and measure their similarity. In this ongoing work, we propose a method to measure similarity of visual concepts and thereby measure the cultural evolution of artist styles. We also present a simple influence model to measure "historic" creativity and discuss the potential of generative AI to explore new, unseen art styles.

Our work is based on the Stable Diffusion model [3], which comprises a text interpreter and an image generator. Specifically, CLIP, the text interpreter used by Stable Diffusion, encodes textual descriptions and images into a joint vector space. In a second step a diffusion model generates an image from this joint representation. Our focus is on the first layer of the CLIP model, which tokenizes the textual input and embeds each token into a token space. This process encodes linguistic concepts, such as "cat" or "Van Gogh," into a shared embedding. As a result, the model implicitly deconstructs images, including artworks, into concepts based on the objects depicted, specific colors, and various stroke techniques. When prompted to generate "a portrait in the style of Rembrandt," the model demonstrates an understanding of the concepts "portrait" and "Rembrandt." Moreover, by generating "a painting of a car in the style of Rembrandt," the model exhibits recombination of concepts not jointly present in the training data.

We introduce a new dataset of Stable Diffusion/CLIP artistic style embeddings for 1,114 artists from the 1400s to 2000s, using the refined WikiArt dataset as a basis [4]. While the original CLIP model has some understanding of the style of some artists as referenced in the example above, we found through empirical evaluation that it fails to accurately represent many other artists based on their name alone. Additionally, since CLIP uses a syllable-based tokenizer, artists are often encoded as multiple, jointly interpreted vectors. To address these issues, we applied textual inversion to recompute a single vector $\vec{v} \in \mathbb{R}^{768}$ representing the style of each artist in the WikiArt dataset. Textual inversion involves finding a representation of a concept $\vec{v}$ in token space based on a set of example images and a prompt blueprint (i.e. "a painting in the style of $\vec{v}$") [5]. In a first version of this dataset, we grouped the 81,444 digital images in the WikiArt dataset by artist and computed a single vector to represent each artist. Given the extensive resources needed for this computation (over 180 GPU days), we plan to share this dataset with the community for further analysis.

We conducted a preliminary analysis of the evolution of artistic styles in the latent space which we are currently expanding. As shown in Figure 1.a, we computed the first two principal components of the style vectors and color-coded them by the median year of the artist's works. Our results demonstrate that, on average, artists are located in the vicinity of their contemporary peers. Moreover, we observed that the convex hull of artistic styles expands over time, indicating that new styles emerge beyond the distribution of existing styles.

To assess the 'historic' creativity of artists, we model an artist's style as a linear combination of pre-dating artists using an influence model (Equation 1). Using Ridge Regularization (L1) to ensure a sparse influence matrix, we show that artist's style can be well described as being influenced by only 2.2 predating artist on average.

$$v_i = \sum_{i>j} w_{ij} v_j + \varepsilon, \forall w_{ij} > 0 \tag{1}$$

We define an artist's 'historic' creativity [2] as the deviation ε of an artist's style from what could be expected based on the artist's possible influence from pre-dating artists alone. An important caveat is that the term 'historic' acknowledges that the dataset we analyzed is curated and that the creativity we measure is from a contemporary, historical perspective.

In this ongoing work, we present an analysis of the evolution of visual styles in the latent space of generative AI. We propose novel methods for computing the stylistic similarity of different artists and measuring stylistic creativity. Our findings shed light on the prospects of discovering new art styles in the latent space of generative AI. We also discuss the limitations of our approach, including potential biases stemming from historical and cultural selection. Our study contributes to the growing field of computational creativity and holds implications for future research in generative AI and quantitative cultural evolution.

References

[1] Byunghwee Lee, Min Kyung Seo, Daniel Kim, In-Seob Shin, Maximilian Schich, Hawoong Jeong, and Seung Kee Han. Dissecting landscape art history with information theory. Proc. Natl. Acad. Sci. U. S. A., 117(43):26580–26590, October 2020.

[2] Ahmed Elgammal and Babak Saleh. Quantifying creativity in art networks. June 2015.

[3] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. CoRR, abs/1503.03585, 2015.

[4] Wei Ren Tan, Chee Seng Chan, Hernan Aguirre, and Kiyoshi Tanaka. Improved artgan for conditional synthesis of natural image and artwork. IEEE Transactions on Image Processing, 28(1):394–409, 2019.

[5] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing Text-to-Image generation using textual inversion. August 2022.

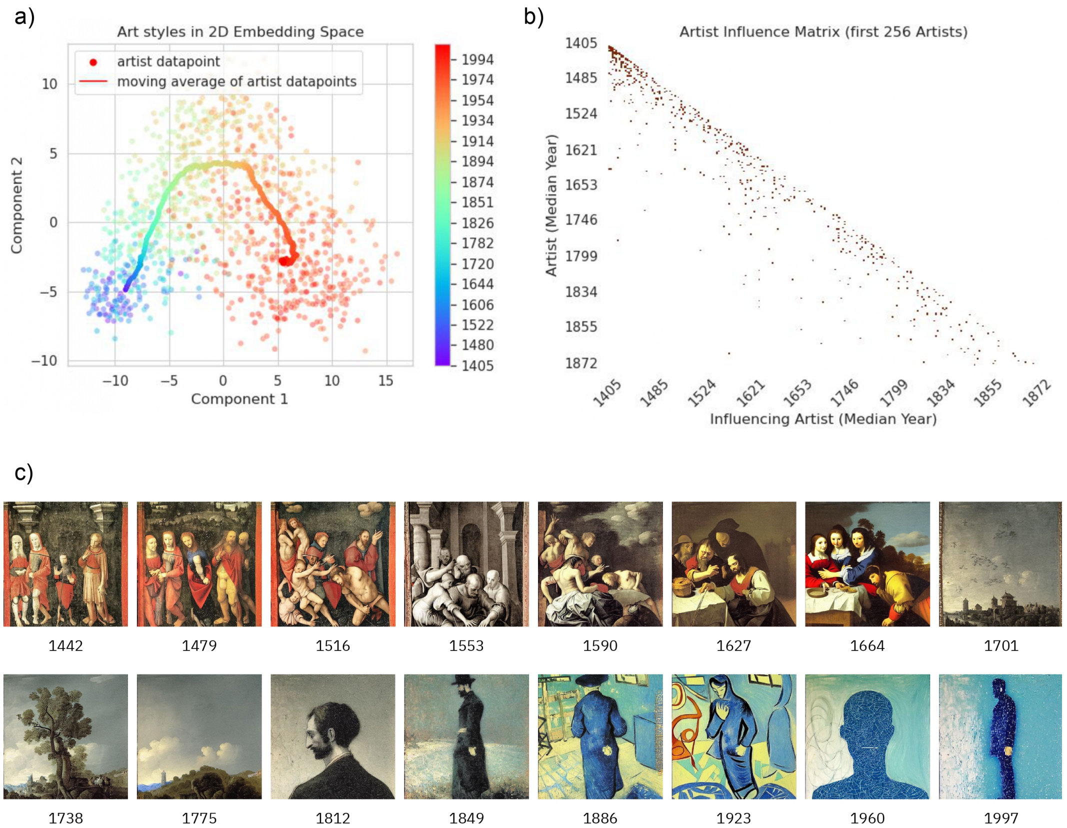

Figure 1: [Preliminary Results] (a) A projection of the first two PCA dimensions of the artists' embeddings with colors representing the median year of the artists' artworks. A rainbow-colored line shows the moving average of the embeddings. (b) Non-zero weights of the linear influence model, as described in equation 1. A non-zero weight can be interpreted as one artist influencing another artist style. Empirically we found a concentration close to the diagonal. Therefore the model suggests that artist in the dataset are predominantly influenced by peers of the same period. (c) Images generated with the prompt "a painting in the style of <avg. year>" with the vector <avg. year> corresponding to the moving average as depicted in (a).