# Artist Discovery with Stable Evolusion

Simon Colton,<sup>1</sup> Amy Smith,<sup>1</sup> Blanca Pérez Ferrer<sup>2</sup> and Sebastian Berns<sup>1</sup>

<sup>1</sup>School of Electronic Engineering and Computer Science, Queen Mary University of London, UK

<sup>2</sup>Etopia Center for Arts and Technology, Zaragoza, Spain

s.colton@qmul.ac.uk

### Abstract

We describe the Stable Evolusion system, with which users can evolve text prompts for use in the Stable Diffusion textto-image generator, and view the resulting imagery. The system is designed for simplicity, enabling users to explore a space of styles visualising content of their choosing. We automate elements of the process with both a semantic search and a novelty-based search, to direct the user towards images of interest, and to help maintain diversity respectively. In addition, by combining the approach with Google Lens image searching, we enable the discovery of human artists and their artwork via the pre-generation of images similar to theirs.

## **Introduction and Background**

A new way of producing high-quality generative art has recently emerged, namely the employment of neural textto-image generators such as MidJourney (midjourney.com), DALL-E (openai.com/product/dall-e-2) and Stable Diffusion (dreamstudio.ai). The reaction to this from artists has been mixed, with some rightly upset that their artwork has been used to train generative deep learning models without permission, possibly infringing copyrights. While finding the right prompt for the generators is often not easy, the ease of use of these systems is perceived to threaten the livelihoods of commercial artists. To make matters worse, the image generators are able to fairly faithfully reproduce certain artists' styles, if prompted with their name, which could further effect livelihoods and legacies.

Other artists have embraced the new creative affordances that have arisen. For instance, photographer Boris Eldagsen recently won a category in the Sony World Photography competition with an AI-generated image, which caused some controversy (Williams 2023). To Eldagsen, the new way to generate images: "... is setting me free ... the boundaries I had in the past – material boundaries, budgets – no longer matter". He points out that the art of choosing the right prompt is not as easy as critics such as (McCormack et al. 2023) suggest. Indeed, dozens of websites where prompts can be downloaded, exchanged or purchased have sprung up. Eldagsen further points out that:

"for the first time in history, the older generation has an advantage, as AI is a knowledge accelerator. Two thirds of the prompts are only good if you have knowledge and skills, when you know how photography works, when you know art history." (Williams 2023) We describe here the *Stable Evolusion* system which helps novice users of the Stable Diffusion image generator to produce images via the evolution of prompts. This addresses somewhat the difficulty people have in writing prompts to achieve imagery of their liking, as the system supplies terminology from art practice and history without the user needing to know these. While not difficult to write a prompt for text-to-image generation, it is difficult to write the right prompt to achieve the kind of imagery required for a particular project. We apply Stable Evolusion to the discovery of human-produced artworks via Google image search, which could in a small way benefit commercial artists, balancing somewhat the difficulties they've encountered recently.

Stable Evolusion is written in a Colab notebook (Bisong 2019), built on top of the following technologies:

• **Stable Diffusion**. Released by Stability AI, this is a textto-image generation system employing a *latent diffusion* model (Rombach et al. 2021) which iteratively de-noises a Gaussian noise image conditioned with a text prompt.

• **CLIP**. Released by OpenAI, this comprises two models which can encode text and images respectively into the same latent space (Radford et al. 2021). As described in (Colton et al. 2021), CLIP can be used to calculate semantic similarities between images and texts.

• Vendi. This is a method for estimating the diversity in a set of media such as images (Friedman and Dieng 2022). It can employ any similarity function, such the distance between image embeddings in a latent space. We use the inception machine vision model for this (Szegedy et al. 2016).

• **Google Lens**. This is a suite of image recognition systems which can be employed to search the internet for images similar to a given image (Conditt 2017).

## **System Description**

The Stable Evolusion system has two ways to employ Stable Diffusion. Firstly, users can supply a key for the DreamStudio API offered by Stability AI (dreamstudio.ai), hence the notebook runs on a CPU, with image generation in the cloud. Alternatively, the notebook has code from the *HuggingFace Diffusers* package (Patil et al. 2022), so image generation can be performed on a Colab-supplied GPU. Generation of a 512x512 pixel image with the API takes around 5 seconds per image (network dependent); a standard Colab GPU takes around 11 seconds; a premium GPU takes around 3 seconds. For the experiments here, we used the Diffusers package.



Figure 1: Screenshot from the Stable Evolusion notebook.

Stable Evolusion employs a straightforward evolutionary approach where a genome consists of:

- A random seed as an integer between 1 and 100,000
- A **content text** describing the scene to be depicted
- A list of visual **modifiers** which are short text phrases

At the start of a session, the user supplies an initial content text such as "painting of a chair", and then starts the process. The system produces a first generation of 8 images using the unmodified content text as a prompt to Stable Diffusion and presents these to the user. The user can then select any image(s) they want to develop and produce a new generation from; if they choose none then it is assumed the next generation should be produced from all of the current generation. Users are free to select images from any previous generation with selections wiped after they have been used. Five generations are shown in the user interface, as per figure 1.

We have experimented with two evolutionary mechanisms for producing novel prompts, namely extension and *mutation*. The extension process simply adds a modifier to the list for a selected genome. The prompt for generating an image is produced by concatenating the modifiers to the content. Users can decide whether to keep the random seeds of their chosen genomes when generating the next round. As we see in experimentation below, child images generated with the same seed, but slightly different modifiers are substantially more similar to their parent than those where the seed is changed. Hence, keeping seeds enables users to make smaller steps in the possibility space, while changing seeds affords bigger steps. Buttons in the UI marked (image) direct Stable Evolusion to keep seeds, those with (prompt) direct it to change them. When a new generation is made, the older generations move down in the GUI to make space.

The pre-selected modifiers have been hand-curated from visual art history and practice, with adjectives and short phrases describing: media (e.g., oil painting); movements (e.g., impressionism); styles (e.g., unfinished); moods (e.g., melancholic); colours (e.g., vibrant colors); patterns (e.g., *tartan*); textures (e.g., *denim*); lighting (e.g., *moonlight*); and materials (e.g., plastic). After 25 such modifiers have been added to a genome, the entire prompt usually comprises more than the 77 tokens which can be accepted by Stable Diffusion, so modifiers 26, 27, etc., will not change the image. Hence at this stage, instead of extending the modifiers, one of them is changed (mutated) to a different one. As we see in the experiments below, this produces similar results to extending modifiers. If desired, the user can change the text prompt during a session, in which case, the system produces 8 new images which have the same seed and modifiers as in the most recent generation, but with the new content text. If the content text is related to the previous one, new images usually look quite similar to those in the row below it, which we found gives a satisfying level of control.

In the session depicted in figure 1, the user chose the prompt "Vase of flowers". In the first generation (at the bottom of the screenshot), 8 images were generated from just this prompt. The user chose images 1.3, 1.6 and 1.8 from these (marked with an asterix) by clicking on the buttons above them. In the next generation, single modifiers were added to the prompt and 8 more images were generated. The user chose images 2.2, 2.7 and 2.8, and the modifiers for these were extended into 8 images for generation three. For generation four, the user changed the prompt to "Vase of white flowers" and we can see that the images look somewhat like those directly below them, but with (more) white flowers. The user chose a single image (4.6) for the fifth generation and produced 8 variations of this by clicking on the 'Choice (Image)' button to retain the seed while extending the modifiers in 8 different ways. The large image showing is 5.5, with the modifiers neon, tartan and cubism.

To increase ease of use, the bottom part of the GUI enables users to force a search over multiple generations. These are guided by one of the following processes, which are each iterated for a user-given number of generations:

• Novelty-based search. Here, Stable Evolusion chooses to evolve the three most novel images in the current generation and extends/mutates their modifiers into the next generation. Novelty for an image, I, is determined by how much the Vendi diversity score reduces when calculated before and after I is removed. We experiment with alternatives below.

• Semantic search. Here, the user can supply a secondary *target phrase*, T, which is used in selecting images for evolution. In particular, the CLIP similarity to T for each of the images in the current generation is calculated, and the three most similar are evolved in the next round. An example of how this drives the search is given in figure 2, showing five generations of chair images progressing towards the target phrase 'psychedelic'. Naturally, the user could instead supply content text including the target phrase, but this can quite drastically change the images, and it is often preferable to evolve towards the target using a semantic search.



Figure 2: Images from a session automatically evolving chair images with the target word 'psychedelic'.

### **Experiments and Results**

To study the effects of the extension and mutation evolutionary operators, we generated images for five content texts, namely 'downtown manhattan' (shortened here to city); 'painting of a chair' (chair); 'seascape with a boat' (seascape); 'vase of flowers' (flowers) and 'modern architecture building' (building). For each content text, T, we generated a parent image from a random genome with nmodifiers, where n ranges from 0 to 24. We then extended these into a child genome by adding a single modifier, then produced a child image. For each (T, n), we produced parent/child pairs both keeping and changing the seed in the child, repeated for 5 different random genomes. This produced 1,250 parent/child pairs, over which we calculated the CLIP similarity and plotted relevant averages in figure 3(a).

We undertook a similar experiment with mutation rather than extension evolving the parent genome into the child, again with results plotted on figure 3(a). As expected, we see that when the child images are produced using the same seed as the parent, the CLIP similarity is substantially higher than when the seed is changed. We expected the similarity of parent/child pairs to increase in line with the position of the altered/added modifier, as later words in a prompt have lower effect on images in general. This trend is certainly observed when the child shares its parent's seed, with the similarity raising from around 0.88 to around 0.94. However, when changing the seed, this appears to be sufficiently disruptive to images that this trend is not observed.

[	Keen seed			Change seed		
Content	Intra	Inter	Darent	Intra	Inter	Darent
Content	mua	muer	1 arcm	mua	mu	1 arcm
City	0.833	0.659	0.916	0.789	0.663	0.826
Chair	0.864	0.678	0.933	0.828	0.685	0.842
Seascape	0.851	0.654	0.923	0.800	0.658	0.826
Flowers	0.847	0.647	0.922	0.805	0.659	0.838
Building	0.839	0.661	0.910	0.782	0.659	0.810
Average	0.847	0.660	0.921	0.801	0.665	0.828

Table 1: Average CLIP similarities over pairs of images.



Figure 3: (a) Child/Parent image CLIP similarity, as the number of modifiers increases from 1 to 25, and the position of the mutated modifier increases. (b) Vendi and CLIP diversity measures over the different selection methods. (c) Comparison of CLIP similarities to target texts for semantic (green bars) and random (blue bars) searches.

For further context about CLIP similarities, in table 1, we record the average CLIP similarities over all pairs of images with (a) the same content text [intra] (b) different content texts [inter] and (c) a [parent]/child relationship. We see that CLIP similarities go as low as 0.65 inter categories, which highlights that relatively high similarities of up to 0.94 seen in figure 3(a) indicate strong visual similarity between parent and child images. This is borne out under visual inspection: often when changing the 25th modifier in a genome and keeping the seed for the child, it looks very similar indeed to its parent. Finally, we note that the graphs in figure 3(a) for extension and mutation are roughly similar. Hence there shouldn't be a noticable difference when mutation takes over. In practice, when mutations occur after 25 modifiers are added, they are applied randomly to a modifier in the final 5, and we've found this provides good continuity.

A reasonable use case for the novelty search is when a user starts a session and wants a diverse set of images to choose from initially, produced over, say, five generations. We experimented with the following six different mechanisms for producing 5 generations at the start of a session:

• **Random:** each genome is generated randomly with *n* modifiers for generation number *n*.

• All: all genomes in the current generation are evolved.

• **Random Evolving:** three genomes are selected randomly for evolution.

• Vendi Novelty: genomes for the three images which reduce the vendi diversity score the most are selected.

• **CLIP Novelty:** genomes for the three images with least total similarity to the other images are selected.

• **CLIP Clustering:** each genome is given a 7-entry vector *profile* by calculating the CLIP similarity between its image and the other images. This profile is used in a K-means clustering process to produce 3 clusters, from each of which a genome is selected randomly for evolution.

For each of the content texts above, over 5 trials each, an initial 8 images were generated from modifier-free genomes. The 8 genomes were then evolved over five generations via extending the modifiers (changing the seed), in six seperate sessions, i.e., one for each of the above generation mechanism. The diversity of each generation was estimated in two different ways: (a) using the Vendi diversity score, which essentially calculates the exponential of the Shannon entropy of a similarity matrix's eigenvalues (Friedman and Dieng 2022), with similarity being the cosine distance between embeddings of images in the inception latent space (Szegedy et al. 2016), and (b) the reciprocal of the average CLIP similarity over every pair of images in the generation.

The results are collated in figure 3(b). The findings are inconclusive, partly because the two diversity measures often didn't agree. In earlier experiments, we found that, subjectively, the CLIP diversity estimation was more accurate than the Vendi score, and so we concentrate on that here. As expected, purely random generation and extending all the genomes in each round produces reliably diverse sets, with some exceptions. However, this gives no continuity or progression from generation 1 to 5, which can be useful in showing users how prompts and images evolve. Indeed, an initial motivation for introducing more sophisticated novelty searches was to have slightly less diversity in order to increase continuity. Of the non-random approaches, the CLIP clustering mechanism appears to have performed the best, with the CLIP novelty approach also performing well.

A reasonable use-case for the semantic search is for a user to choose a single image and then evolve it (keeping the seed) over five generations, using CLIP-guidance with respect to a target text. For each of the content texts above, and each of these targets: fiery, abstract, minimal, yellow and *psychedelic*, we simulated this use case 4 times. For comparison, we did likewise but using random choice rather than the semantic choice. At each generation, we recorded the highest CLIP similarity between an image and the target text for both semantic and random search. The average CLIP similarities over the sessions are recorded in figure 3(c). As expected, the semantic search always produced CLIP similarities (on average) higher than the random search. An example where the semantic search worked well is given in figure 2. Note that the random session is given down the left hand side. However, on inspection, we found that less than a quarter of the semantic searches produced images reflecting the target, hence there is much room for improvement.

## **Application to Artist Discovery**

When Stable Evolusion is run in the Chrome browser, rightclicking any generated image allows that image to be used in a search for similar images, via Google Lens. In a small pilot study with two participants, we explored the potential for this to be used to discover artists that were previously unknown to the participant. The first participant was asked to imagine they were decorating a new apartment and wanted to find some human-painted physical artwork available for purchase online. They used Stable Evolusion for around 1 hour, starting with the content phrase 'downtown manhattan'. They produced 320 images over 40 generations, and searched online for physical artworks 17 times, each time finding something interesting from the traditional art world. They highlighted three artists whose work was particularly interesting, with details given in the appendix.

The second participant is an art historian, curator and cultural mediator with 15 years experience. They were asked to imagine a theme for a new exhibition and to use Stable Evolusion to find potential artists. The theme chosen was 'underwater world' and the session lasted 20 minutes, with the participant finding numerous potential artists, with 6 highlighted in the appendix. The participant pointed out that this approach helps break a chicken-and-egg problem in finding artists: it's hard to know in advance what to ask for a preliminary text search, but without such a search, images can't be found for image search to discover artists. With Stable Evolusion, they said, many different styles are offered to visualise the content text, some of which were new to them. Both participants expressed satisfaction in the ease of use of the system and the ability to discover artists. Both participants also noted that the user interface was cumbersome.

## **Conclusions and Future Work**

Prompting image generators is currently a sought-after skill, and numerous approaches have been developed to automate prompt engineering. In (Martins et al. 2023), the authors implemented a similar evolutionary approach to prompt discovery as ours, but focused on quality of the results and match to user preferences, rather than artistic visualisations. Also, reverse engineering images to suggest prompts that will produce similar images is available via the CLIP Interrogator (huggingface.co/spaces/pharma/CLIP-Interrogator), and numerous other tools such as the Prompt Builder (promptomania.com/stable-diffusion-prompt-builder) are available to help in writing prompts.

We presented here the Stable Evolusion system which helps users find artistic visualisations of chosen content material via an evolutionary search which constructs prompts for text-to-image generation. We experimented to understand better how prompt evolution affects the images generated, and to evaluate the automated novelty and semantic searches. We demonstrated that the approach can help people to find human artists that they perhaps would not be able to through standard search methods. We plan to improve the user interface (moving to a HuggingFace Space), the search strategies, possibly using crossover techniques, and to increase functionality in discovering human artists.

## **Demonstration**

### The Stable Evolusion colab notebook is available here:

https://colab.research.google.com/ drive/17sqwISmLbcpw3DEzMSzw1mbd8I1XBK4Z

### Acknowledgements

We would like to thank Stability AI for providing £500 of free credits for the DreamStudio API, enabling much preliminary experimentation. We also thank the anonymous reviewers for the helpful and insightful comments. Amy Smith and Sebastian Berns are funded by the IGGI centre for doctoral training [UKRI grant EP/S022325/1].

### References

Bisong, E. 2019. Google colaboratory. In Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress.

Colton, S.; Smith, A.; Berns, S.; and Murdock, R. 2021. Generative search engines: First experiments. In Proceedings of the International Conference on Computational Creativity.

Conditt, J. 2017. Google lens is a powerful, AI-driven visual search app. Engadget, May 17.

Friedman, D., and Dieng, A. B. 2022. The vendi score: A diversity evaluation metric for machine learning. arxiv 2210.02410.

Martins, T.; Cunha, J. M.; Correia, J.; and Machado, P. 2023. Towards the evolution of prompts with MetaPrompter. In Johnson, C.; Rodríguez-Fernández, N.; and Rebelo, S. M., eds., Proceedings of EvoMusArt: Artificial Intelligence in Music, Sound, Art and Design.

McCormack, J.; Cruz Gambardella, C.; Rajcic, N.; Krol, S.; Llano, T.; and Yang, M. 2023. Is writing prompts really making art? In Johnson, C.; Rodríguez-Fernández, N.; and Rebelo, S. M., eds., Proceedings of EvoMusArt: Artificial Intelligence in Music, Sound, Art and Design.

Patil, S.; Cuenca, P.; Lambert, N.; and von Platen, P. 2022. Stable diffusion with diffusers. Hugging Face Blog. huggingface.co/blog/stable\_diffusion.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning transferable visual models from natural language supervision. arxiv 2103.00020.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2021. High-resolution image synthesis with latent diffusion models. arxiv 2112.10752.

Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; and Wojna, Z. 2016. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

2023. Williams, Z. 'AI isn't a threat' - Boris Eldagsen, whose fake photo duped the Sony judges, hits back. Guardian, 18th April 2023.















### Random seed: 94932 Prompt: downtown

Appendix

manhattan, cubism

Artist found: Hanna Kryvolap

#### Random seed: 16797

Prompt: downtown manhattan, davlight, upbeat, pointillism, abstract expressionism, cyberpunk unfinished, wood, art nouveau colors, cotton, newsprint, abstract, moonlight

Artist found: Mona Edulesco

#### Random seed: 16797

Prompt: downtown manhattan daylight, upbeat, pointillism, abstract expressionism erpunk, unfinished, wood, a nouveau colors, cotton, newsprint, abstract, finger painting, grayscale

> Artist found: Maria Asunción Raventós









Prompt: underwater world, cotton, impressionism, quiet, glowing, black and white Artist found: Alex Turco



#### Prompt: underwater world, cotton, impressionism, quiet checkerboard, cubism

Random seed: 6242

Artists found: Carl Robert Holty Andre Lanskov



Prompt: underwater world, argyl, cubism, quiet, polka dot, saturated colors

Artists found: Irina Kirk Robert Margetts

#### Random seed: 57491

Prompt: underwater world, cotton, impressionism, quiet glowing, black and white, bright, color photograph

> Artist found: Sebastien Zanella



Figure 4. Example artist discoveries by study participants 1 (top) and 2 (bottom). Left: generated image; Centre: seed, prompt and discovered artist(s); Right: retrieved image(s).

