Autolume 2.0: A GAN-based No-Coding Small Data and Model Crafting Visual Synthesizer

Arshia Sobhan*
Simon Fraser University, CA
asobhans@sfu.ca

Ahmed M. Abuzuraiq*
Simon Fraser University, CA
aabuzura@sfu.ca

Philippe Pasquier
Simon Fraser University, CA
pasquier@sfu.ca

Abstract

Autolume is a no-code, generative AI tool that leverages the artistic potential of Generative Adversarial Networks (GANs). It empowers artists to train and manipulate their own models using small datasets, thereby fostering unique aesthetic exploration and personal creative control. By streamlining the full creative workflow—covering data preprocessing, model training, real-time latent space navigation, and output upscaling—Autolume makes the artistic potential of generative AI accessible to non-technical users. It also supports interactive applications, such as audio-reactive visuals, through OSC integration. As a free and open-source platform, Autolume expands creative possibilities for artists, enabling them to blend and customize aesthetic styles in a flexible and efficient manner.

1 Introduction

Recent advances in large-scale text-to-image diffusion models, such as DALL·E 3 by OpenAI [17] and Stable Diffusion by Stability AI [18], have significantly enhanced the ability to generate visuals from textual descriptions and led to the development of user-friendly tools that further democratize access to AI-driven art creation. Prior to that, small and personally-trainable models such as Generative Adversarial Networks (GANs) were more commonly used by AI artists with coding proficiency, such as Helena Sarin, Memo Akten, Mario Klingemann, and Anna Ridler, among others. [†]

Personal small models offer a distinctive value for artists over diffusion models which warrant continued support and attention from the AI community. Firstly, smaller models can be trained on smaller datasets, utilizing computing resources available to a wider range of artists. Working with smaller, personal or curated, datasets in tandem with personally-trainable models such as GANs offers a highly controllable and intimate artistic workflow that fosters creative ownership and provides a sense of craftsmanship [20, 6]. Secondly, they avoid perpetuating biases associated with using large-scale models [20], which may be trained on datasets of artworks lacking proper attribution to creators. Thirdly, from an aesthetics point of view, models such as GANs provide a smoothly navigable and serendipitous latent space [11] as well as a unique aesthetic of visual indeterminacy [13].

Fourthly, unlike large-scale text-based diffusion models, models such as GANs do not require the verbal articulation of tacit artistic intent [11], which is important for some art forms (e.g. Arabic calligraphy) where there is an inherent lack of vocabulary for describing the pieces, whether due to the nature of the art form or because they are not present in training datasets. Along the same point, the under-representation of some art forms in the training of large generative AI models may lead to inaccurate or culturally insensitive results. Lastly, AI artists work with algorithms and data as materials for creation [10], and so by virtue of being faster to train or fine-tune, variations of smaller models can become easier to explore[10]. Smaller models also offer high levels of creative control as

*Both authors contributed equally to this paper.

†To learn more about these artists and others, please visit https://aiartists.org/

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

they are amenable to hacking [7], adapting, and crafting, e.g. through active divergence techniques such as network bending or blending [9].

2 Autolume 2.0

Latent spaces can be challenging to navigate without guidance, and manipulating generative models can be challenging and often requires technical expertise, which many artists may lack. To make the artistic potential of personally-trainable generative AI models, particularly GANs, more accessible, and to expand the possibilities of working with them in an interactive real-time setting, we developed Autolume. Initially developed as a live VJing tool [16], Autolume is a no-coding, user-friendly visual synthesizer leveraging the artistic potential of GANs. It simplifies the entire GAN workflow, from dataset preprocessing and model training to real-time latent space exploration, feature extraction, and output upscaling, making it accessible to non-technical creatives. Moreover, Autolume introduces interactive generation via Open Sound Control (OSC) messaging, enabling audio-reactive and other forms of interactive art, thus broadening its creative possibilities.

Many no-code diffusion-based AI art tools have been introduced [2, 3, 4], but only a few exist for GAN-based art creation [5, 1]. Autolume is distinguished by being free, open-source, and locally run. It provides real-time interaction with direct manipulation including latent space navigation and model mixing. Finally, it offers options for adapting and crafting generative AI models [6]. Autolume offers both offline (pre-processing, post-processing) and real-time (latent space navigation, OSC messaging, semantic manipulation) modules. Autolume can be accessed on `https://www.metacreation.net/autolume`, and examples of artworks produced with Autolume can be found on the website as well as in Figure 2.

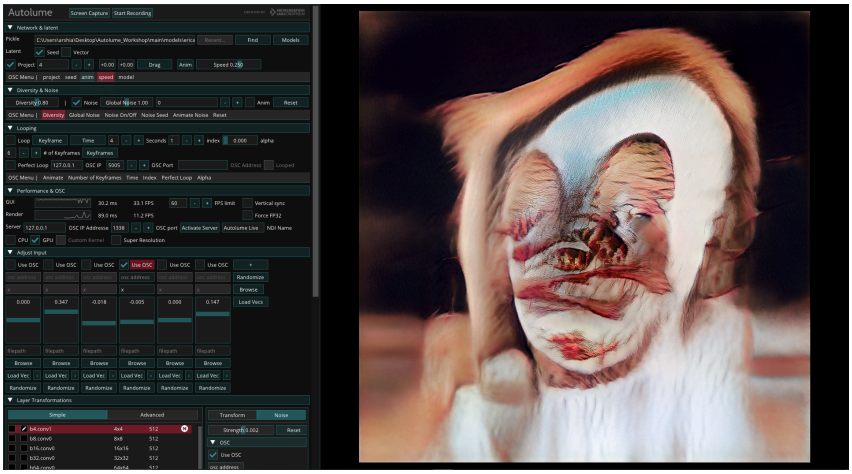

Figure 1: Screenshot of the Autolume-live interface.

2.1 Offline System Features

Autolume offers five offline modules, not presented in a figure here, for different processing features.

2.1.1 Model Training

The training module enables users to train StyleGAN2 [15] models from scratch or fine-tune pre-tained models, using either image datasets or frames extracted from uploaded videos. It provides data preprocessing, including resizing and framing options, and supports both square and non-square datasets. To enhance training with smaller datasets, it offers Adaptive Discriminator Augmentation (ADA) [14] and Differentiable Augmentation (DiffAug) [22] methods, each with various augmentation pipelines. These features facilitate training with smaller datasets and allows artists to use datasets as small as 100 images [‡]. Users can adjust training hyperparameters like batch size, learning rates,

[‡]The quality of training also depends on the visual coherency of the dataset and its size.

and gamma to achieve desired results, and the training runs on the user's local GPU. While achieving optimal results may require multiple iterations, a robust GPU, such as NVIDIA GeForce RTX 3090, can yield meaningful outcomes in about 12 hours. A real-time progress monitoring is available via a grid display of generated images to help visually evaluating the results during training.

2.1.2 Latent Space Projection

The projection module in Autolume allows users to enter a reference image, text prompt, or both to find the closest latent vector in a trained model that resembles them, as well as generate a video of the search process. The resulting latent vector can be loaded into the real-time module for further explorations.

2.1.3 Feature Extraction

The semantic feature extraction module uses GANSpace [12] to identify and extract interpretable feature directions in the latent space which provides an alternate approach for navigating the latent space. This feature includes options for feature sparsity and the number of directions to be calculated, and the calculated feature directions can be loaded and explored in the real-time module.

2.1.4 Super Resolution

Due to computational limitations, artists may opt to train their models on smaller resolutions (e.g., 256x256 or 512x512) and then upscale the results to achieve results with higher fidelity. The super-resolution module enables high-quality upscaling of images and videos using the Real-ESRGAN algorithm [21]. Users can choose from three modes—Quality, Balance, and Speed—based on their preference for output fidelity and processing time. Additionally, users can specify the upscaling factor or define exact output pixel dimensions. This module supports batch processing for images and videos. The Speed option is available in Autolume-live, allowing for upscaling during real-time generation.

2.1.5 Model Mixing

The model mixing module enables users to blend two trained models by combining their layers [19], creating and saving a new model for future use. This functionality allows for the fusion of aesthetic features from different models, generating unique visual styles and new aesthetic spaces. This feature is also available in the real-time module for a live preview of the results. The resulting hybrid models can be saved locally for future use.

2.2 Real-time System Features: Autolume-live

The Autolume-live module allows artists to load trained models and interact with them in real-time, offering a suite of features for exploring latent space and experimenting with network parameters to achieve unique aesthetics.

2.2.1 Latent Space Navigation and Controls

Artists can control the visual features by adjusting parameters such as global noise and truncation. Various latent space exploration options are available, including setting specific seed numbers as starting points of the exploration, performing random walks with adjustable speeds, and utilizing different looping techniques. For instance, artists can identify interesting points in the latent space and loop between them as keyframes, which allows them to create curated experiences of the latent space. Keyframes can also be saved to be loaded into other parts of the interface. Moreover, pre-calculated directions from Feature Extraction, derived in the offline module, can be loaded and explored in real-time.

2.2.2 Model Bending and Blending

The Layer Transformations panel enables custom noise control and transformations (e.g. translation, scaling, and inversion) on individual model layers, allowing for highly detailed control over visual outputs and expanding creative possibilities [8]. The model mixing feature, available in the offline module, can also be used in real-time for a live preview of the results.

2.2.3 Interactivity

Autolume-live also supports Open Sound Control (OSC) which is a protocol for networking different applications, such as sound/visual synthesizers or multimedia devices, to enable data exchange real-time control. OSC enables artists to control the real-time module parameters for audio-reactive and other interactive applications. The OSC feature also offers fine-tuning capabilities, such as applying mathematical expressions to OSC signals, for precise mapping within Autolume. Lastly, the generated visuals can be sent through a Network Device Interface (NDI) video streaming protocol to other software for additional post-processing.

2.2.4 Presets

To save the progress made during navigation and enable quickly revisiting preferred settings, artists can save all their current settings, including OSC mappings or sliders' values, as presents that are linked to their specific models.

3 Artworks

Autolume has been used in multiple artworks by various artists. Erica Lapadat-Janzen, in collaboration with Philippe Pasquier, used Autolume to train a model based on her artworks and explore latent space aesthetics, crafting unique stills and video loops in "Dreamscape." Another example is "Alpha Prism" by the Vancouver-based photographer Kris Krüg, who investigates identity and ethics in AI by training models exclusively on his analog portrait photography. Autolume has also been applied in interactive works; Jonas Kraasch and Philippe Pasquier used it to create Mzton, which explores themes of rebirth in audio-reactive visuals. Also, "Reprising Elements" showcases a live performance where Arshia Sobhan and Joshua Rodenberg integrate Autolume in an interactive, co-creative process exploring the interplay between Persian calligraphy, generative AI, and sound art.

(a) (b) (c) (d)

Figure 2: Works created with Autolume. (a) Dreamscape (b) Alpha Prism (c) Mzton (d) Reprising Elements

4 Conclusion

We presented Autolume, a tool for art creation that adopts model crafting and small-data mindsets. The goal of this project is to give artists the means to work in ways that maximize creative ownership and minimize ethical concerns associated with large-scale generative models. In the future, we will continue to develop Autolume and explore its potential through collaborations with artists.

5 Acknowledgment

We would like to acknowledge the Canadian funding agencies the Social Sciences and Humanities Research Council (SSHRC) and Natural Sciences and Engineering Research Council (NSERC) for their support in funding this research.

References

[1] Autolume: A Neural-network based Visual Synthesizer. Metacreation Lab. https://www.metacreation.net/autolume, .

[2] Automatic1111 WebUI. https://github.com/AUTOMATIC1111/stable-diffusion-webui, .

[3] ComfyAI. https://github.com/comfyanonymous/ComfyUI, .

[4] InvokeAI (community-version). https://github.com/invoke-ai/InvokeAI, .

[5] Playform: No-Code AI for Creative People. https://www.playform.io/train, .

[6] A. M. Abuzuraiq and P. Pasquier. Seizing the Means of Production: Exploring the Landscape of Crafting, Adapting and Navigating Generative AI Models. In *In the 3rd Generative AI and HCI Workshop*.

[7] T. Broad. Using generative AI as an artistic material: A hacker's guide. In *In Proceedings of Explainable AI for the Arts Workshop 2024 (XAIxArts 2024)*, 2024.

[8] T. Broad, F. F. Leymarie, and M. Grierson. Network bending: Expressive manipulation of deep generative models. In *Artificial Intelligence in Music, Sound, Art and Design: 10th International Conference, EvoMUSART 2021, Held as Part of EvoStar 2021, Virtual Event, April 7–9, 2021, Proceedings 10*, pages 20–36. Springer.

[9] T. Broad, S. Berns, S. Colton, and M. Grierson. Active divergence with generative deep Learning–A survey and taxonomy. In *Proceedings of the 12th International Conference on Computational Creativity (ICCC '21)*, 2021. ISBN 978-989-54-1603-5. doi: arXivpreprintarXiv:2107.05599.

[10] B. Caramiaux and S. Fdili Alaoui. "Explorers of Unknown Planets" Practices and Politics of Artificial Intelligence in Visual Arts. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–24, 2022.

[11] A. Elgammal. Text-to-Image Generators Have Altered the Digital Art Landscape—But Killed Creativity. Here's Why an Era of A.I. Art Is Over. https://news.artnet.com/art-world-archives/ahmed-elgammal-op-ed-ai-art-is-over-2304028.

[12] E. Härkönen, A. Hertzmann, J. Lehtinen, and S. Paris. Ganspace: Discovering interpretable gan controls. *Advances in neural information processing systems*, 33:9841–9850, 2020.

[13] A. Hertzmann. Visual indeterminacy in gan art. In *ACM SIGGRAPH 2020 Art Gallery*, pages 424–428. 2020.

[14] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila. Training generative adversarial networks with limited data. *Advances in neural information processing systems*, 33: 12104–12114, 2020.

[15] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.

[16] J. Kraasch and P. Pasquier. Autolume-Live: Turning GANs into a Live VJing tool. In *Proceedings of the 10th Conference on Computation, Communication, Aesthetics & X*, pages 152–169, 2022.

[17] OpenAI. DALL-E 3. https://openai.com/dall-e-3, .

[18] OpenAI. Stable Diffusion. https://stability.ai/stable-image, .

[19] J. N. M. Pinkney and D. Adler. Resolution dependent gan interpolation for controllable image synthesis between domains. In *Machine Learning for Creativity and Design Workshop*. doi: arXivpreprintarXiv:2010.05334.

[20] G. Vigliensoni, P. Perry, and R. Fiebrink. A small-data mindset for generative AI creative work. In *In Generative AI in HCI Workshop, CHI '22*, page 5, 2022.

[21] X. Wang, L. Xie, C. Dong, and Y. Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021.

[22] S. Zhao, Z. Liu, J. Lin, J.-Y. Zhu, and S. Han. Differentiable augmentation for data-efficient gan training. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 7559–7570. Curran Associates, Inc., 2020.