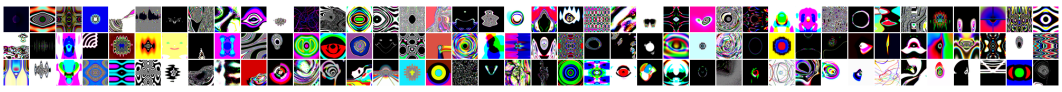
Navigating Neural Fields with Vision-Language Models

Neale Ratzlaff, Phillip Howard, Vasudev Lal
Intel Labs
Santa Clara, CA
{neale.ratzlaff, phillip.r.howard, vasudev.lal}@intel.com



Abstract

Generative art is an enduring discipline in the field of computer science that has traditionally taken on a wide variety of creative implementations. But if we view the current landscape of generative art without a discerning eye, the scope of techniques and methods may look quite flat – only diffusion models, LLMs, and their LoRAs to be seen. In this work we aim to showcase a variation of an older technique for image generation that can create striking visual art without relying on training data, exhaustive computation, or narrowly defined priors. Specifically, we revisit the CPPN-NEAT algorithm, and retool it to be more amenable to current generative model workflows. Instead of evolutionary augmentation, we generate random Watts-Strogatz graphs, convert them to neural fields, and generate the resulting image at an arbitrary resolution. We obtain high-quality samples by using an off-the-shelf VLM to make pairwise selections between generated examples. Images that survive multiple rounds are selected for final human review. This automated procedure is simple, and allows us to quickly and easily generate 12000px x 12000px images on a consumer desktop machine, in a style that is distinct from publicly-available image generation models.

1 Introduction

There has been a recent surge of interest in generative art due to the incredible capabilities of large language models [Achiam et al., 2023, Bai et al., 2022, Dubey et al., 2024] as well as frontier text-to-image generation models [Saharia et al., 2022, Peebles and Xie, 2023, Rombach et al., 2022]. However, its not yet clear to what extent these new generative models can be used to augment, rather than replace artistic workflows. It is straightforward to simply prompt for a style of painting, or for verse, or sound, but the frontier of generative art is wide open when it comes to generation beyond prompting. In the spirit of building upon the impressive capabilities of large generative models, this work attempts to revisit an older technique, CPPN-NEAT [Stanley, 2007, 2006, Ha, 2016], and augment it to be faster, scalable, and ultimately more practically expressive. In short, this work proposes a pipeline for automatically creating high quality generative art with minimal human supervision. This pipeline comprises three stages: initialization of the generating function as a random Watts-Strogatz (WS) neural field, generation of output images, and selection of a new generating function. The backbone of this method is a 2D neural field [Xie et al., 2022] that is used to generate a single image – that when randomly initialized, reflects the underlying architecture of the generator itself.

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

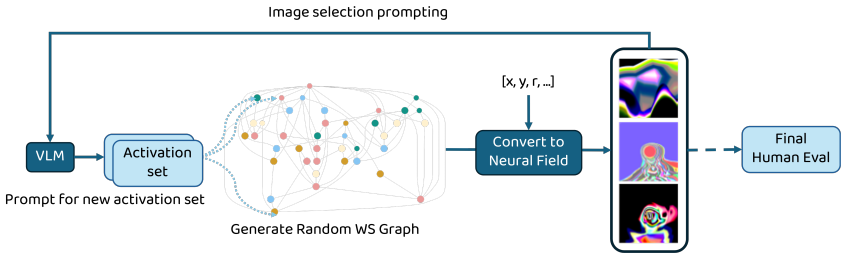


Figure 1: Overview of generation pipeline: activation selection, WS graph generation, converstion to neural field, to image generation.

2 Related Work

2.1 CPPN-NEAT

Compositional Pattern-Producing Networks (CPPNs) are neural networks that generate patterns by mapping a set of coordinates to colors or intensities, enabling the creation of complex and geometric structures. [Stanley, 2007, 2006] introduced CPPNs in conjunction with NeuroEvolution of Augmenting Topologies (NEAT), an evolutionary algorithm that evolves both the weights and connection patterns of neural networks. CPPN-NEAT generates arbitrarily complex patterns through its node activations like `sigmoid`, `tanh`, `gaussian`, `affine`, etc. Evolutionary [Secretan et al., 2008], and generative versions [Ha, 2016] of this method have been explored, but still require manual tuning of the CPPN parameters to get appealing outputs.

2.2 Neural fields

Also known as implicit neural representation functions, neural fields generalize CPPNs to real-valued output assignments to a coordinate space that can represent data such as physical quantities [Li et al., 2023], images [Sitzmann et al., 2020], and 3D shapes [Mildenhall et al., 2021] as continuous functions parameterized by neural networks. Instead of discrete representations like pixels or meshes, neural fields map spatial coordinates directly to signal values, enabling high-resolution and continuous representations. We make use of the neural field paradigm to infer pixel values, but also alpha and gamma values.

3 Method

In this work we begin with observing that generating images with randomly initialized neural fields can create interesting outputs by varying the activation functions [Ha, 2016], even with a small fixed architecture. We build on this approach by increasing the complexity of the generator, using a random Watts-Strogatz graph [Xie et al., 2019] with a large set of possible activation functions instead of a small MLP. This approach yields increased diversity, however, there exist a nontrivial set of graph topologies that yield images with extreme entropy (blank or noisy). This means that naively using this algorithm will result in extremely similar artwork, or else images where there is too little or too much information to be interesting. To acquire a more consistent set of interesting outputs without manually sifting through thousands of candidates, we design a semi-closed loop approach that uses a pretrained VLM [Liu et al., 2023, Li et al., 2024] to select the most interesting image from a batch of generations. The VLM also has access to a set of ground truth high-quality images and corresponding activation sets to use as references. After selecting the best image, the VLM is then prompted to generate a new set of activations that are applied in topological order to the next WS graph. In some sense the VLM is acting as an agent, trying to create interesting artwork with neural fields. Through this approach, images can go through K rounds of refinement, where K is a hyperparameter, before being chosen as a candidate for final human review. Additional samples can be found in the appendix.

References

J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Y. Bai, S. Kadavath, S. Kundu, A. Askell, J. Kernion, A. Jones, A. Chen, A. Goldie, A. Mirhoseini, C. McKinnon, et al. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*, 2022.

A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Yang, A. Fan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

D. Ha. Generating abstract patterns with tensorflow. *blog.otoro.net*, 2016. URL `https://blog.otoro.net/2016/03/25/generating-abstract-patterns-with-tensorflow/`.

F. Li, R. Zhang, H. Zhang, Y. Zhang, B. Li, W. Li, Z. Ma, and C. Li. Llava-next-interleave: Tackling multi-image, video, and 3d in large multimodal models. *arXiv preprint arXiv:2407.07895*, 2024.

X. Li, Y.-L. Qiao, P. Y. Chen, K. M. Jatavallabhula, M. Lin, C. Jiang, and C. Gan. Pac-nerf: Physics augmented continuum neural radiance fields for geometry-agnostic system identification. *arXiv preprint arXiv:2303.05512*, 2023.

H. Liu, C. Li, Q. Wu, and Y. J. Lee. Visual instruction tuning. In *NeurIPS*, 2023.

B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106, 2021.

W. Peebles and S. Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4195–4205, 2023.

R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, K. Ghasemipour, R. Gontijo Lopes, B. Karagol Ayan, T. Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, 35:36479–36494, 2022.

J. Secretan, N. Beato, D. B. D Ambrosio, A. Rodriguez, A. Campbell, and K. O. Stanley. Picbreeder: evolving pictures collaboratively online. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1759–1768, 2008.

V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

K. O. Stanley. Exploiting regularity without development. In *AAAI Fall Symposium: Developmental Systems*, volume 49, 2006.

K. O. Stanley. Compositional pattern producing networks: A novel abstraction of development. *Genetic programming and evolvable machines*, 8:131–162, 2007.

S. Xie, A. Kirillov, R. Girshick, and K. He. Exploring randomly wired neural networks for image recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1284–1293, 2019.

Y. Xie, T. Takikawa, S. Saito, O. Litany, S. Yan, N. Khan, F. Tombari, J. Tompkin, V. Sitzmann, and S. Sridhar. Neural fields in visual computing and beyond. In *Computer Graphics Forum*, volume 41, pages 641–676. Wiley Online Library, 2022.

A Additional Samples

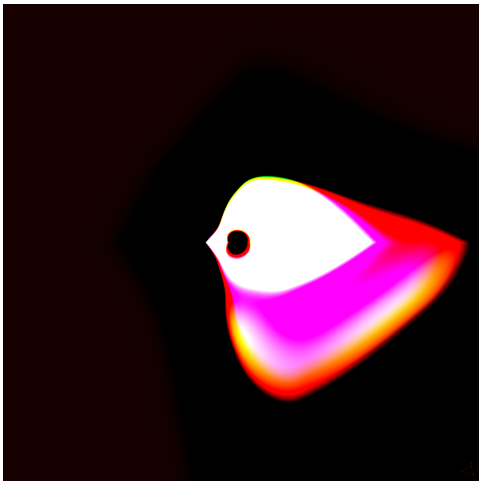


Figure 2

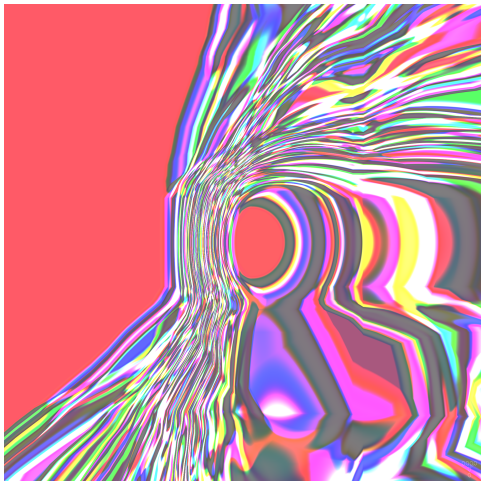


Figure 3

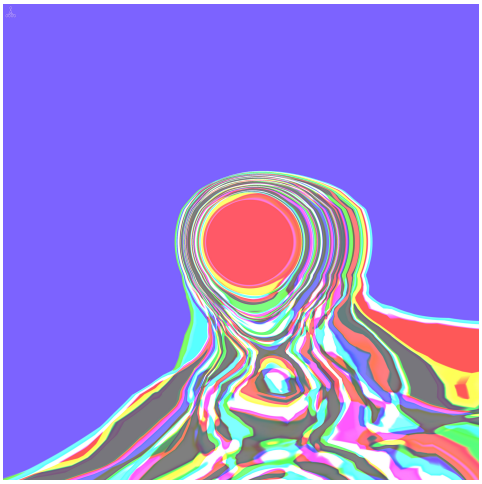


Figure 4

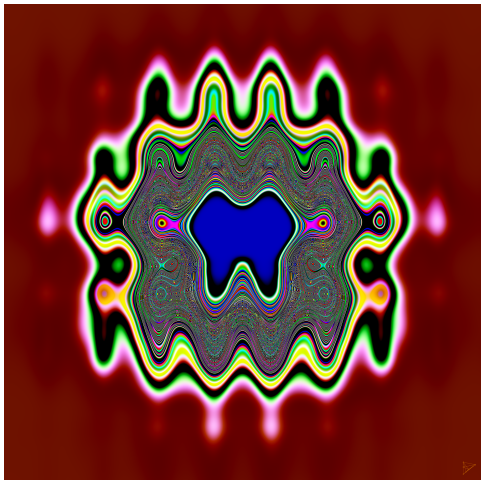


Figure 5

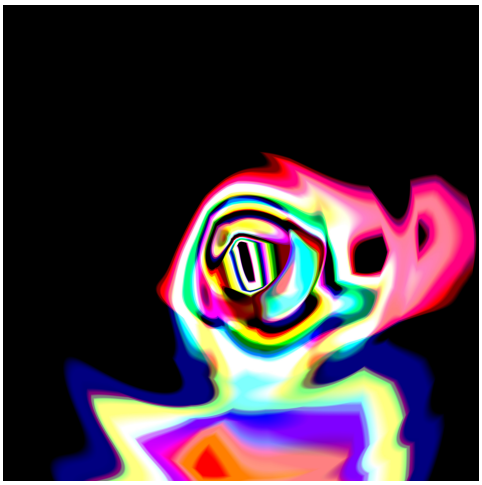


Figure 6

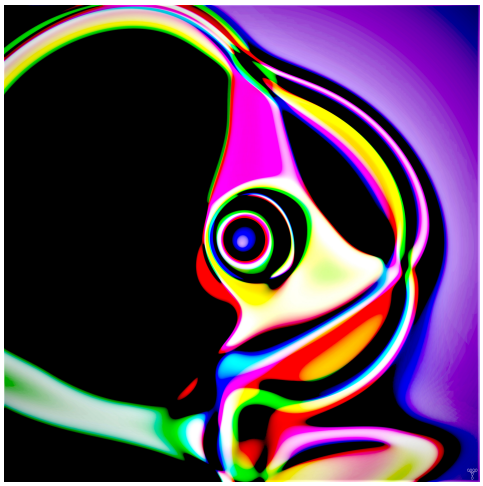


Figure 7

Figure 8

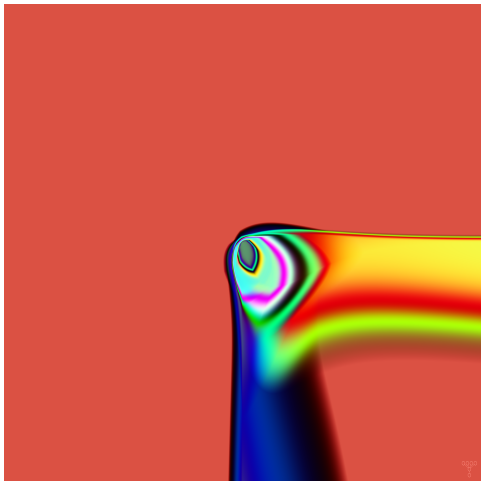


Figure 9

Figure 10

Figure 11

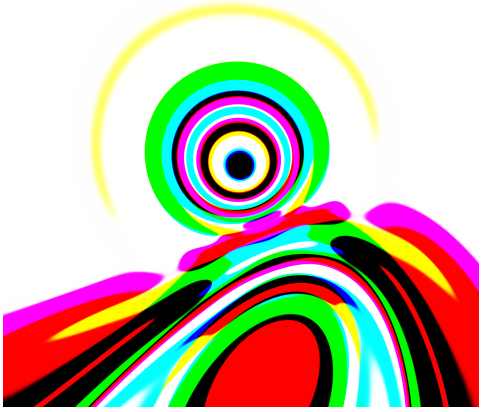


Figure 12

Figure 13