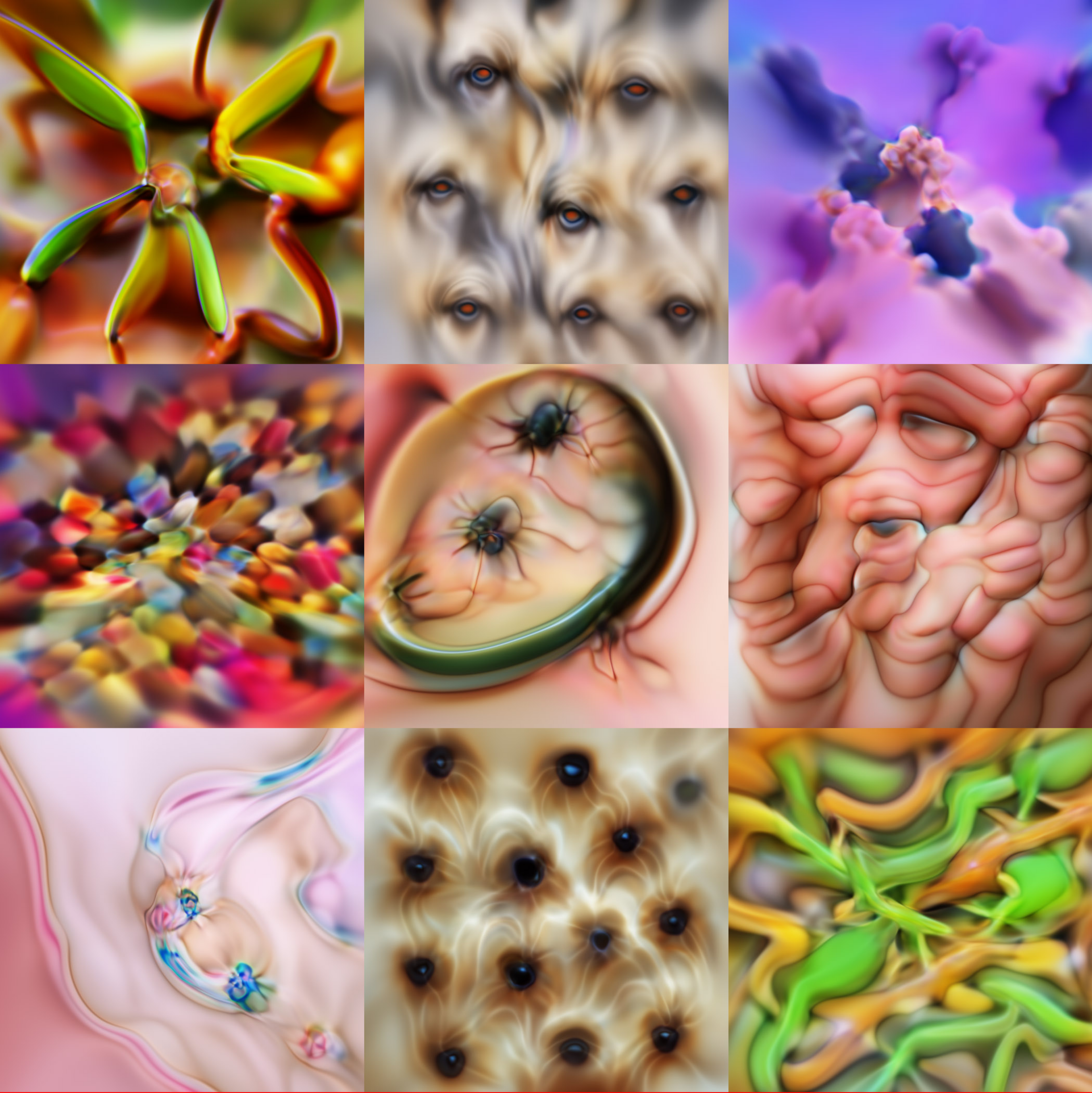
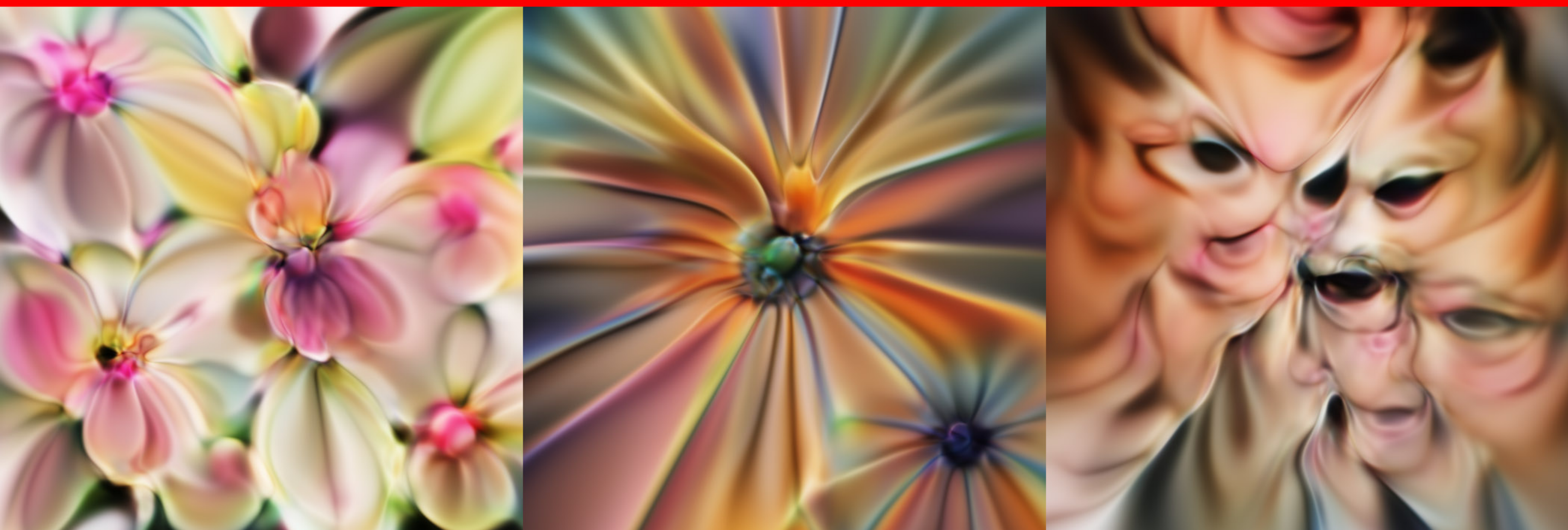


the neural image: an atlas of computer vision

project proposal



1. Introduction

Over the course of the last five years, artificial neural networks have become one of the most promising and popular technologies in the field of artificial intelligence. Machine learning with neural networks covers a huge variety of domains, like autonomous driving, board games or computer translation.[1] Most prominently, automated image recognition has made its way into video surveillance systems, consumer electronics and smartphone apps alike, and its results – but also new classes of unintended visual artifacts – have started to leak into popular culture.

In continuation of the trajectory of cinematic images that have shaped the visual imagination of the 20th century (*movement-images* giving way to *time-images*, if we want to follow Gilles Deleuze[2]), we may already be witnessing the advent of a new type of *neural image*[3]: one that is networked by definition, exists primarily as a relation to other images rather than as a specific type of content or form, and is always tied to a series of machines that either record, render, analyze, transform or display the image – but which are also, increasingly, the primary addressee, audience or target of neural images.

But what do neural networks actually "see"? In order to approach this question, new kinds of generative neural networks – trained to visualize features of image recognition networks[4] – have, in the last two years, received increased attention. This has led to the genesis of adversarial neural networks – specialized on creating images or objects that will be grossly misclassified by machine vision systems[5] – but has also brought into focus entirely new genres of artificial images that trace or animate the contours of the very building blocks of computer vision.

It is immediately obvious that these visual objects are not just of interest in the context of scientific research, but also ready to be appropriated for new forms of artistic creation. In fact, applications like DeepDream[6] and related artifacts like puppyslug[7] have become part of the internet's visual repertoire, and one can already make out a first generation of video art produced with generative neural networks.[8] In terms of artistic agenda, however, most of this output appears to remain more or less trapped in the neo-psychedelic and meme-ready aesthetic proposed by DeepDream et al. Similar constrains seem to exist in the field of neural style transfer, popularized by free online services like DeepArt[9], which rarely manages to transcend the aesthetic paradigms (van-gogh-ify, bruegel-ize, etc) of the first publicly available demo applications.[10] Generally, there is an abundance of eye-catching, mostly decorative material, but it's still rather rare to come across more analytical or long-form work.

This is, however, less of a lament about the short-lived or shallow-sighted preferences of the Instagram era, but more of a reminder that making media art with neural networks remains a "hard problem": without a relatively comprehensive understanding of the underlying technologies, including some halfway advanced math and programming skills, playing around with neural networks will

1. See, for example: David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, Demis Hassabis, *AlphaZero: Shedding new light on the grand games of chess, shogi and Go*, 2018, https://deepmind.com/blog/alphazero-shedding-new-light-grand-games-chess-shogi-and-go/

2. Gilles Deleuze, *L'image-mouvement. Cinéma 1*, 1983; Gilles Deleuze, *L'image-temps. Cinéma 2*, 1985

3. Sebastian Lütgert, *The Neural Image (and a few others)*, 2018, https://pad.ma/documents/AIH

4. Chris Olah, Alexander Mordvintsev, Ludwig Schubert, *Feature Visualization: How neural networks build up their understanding of images*, 2017, https://distill.pub/2017/feature-visualization/

5. LabSix, *Fooling Neural Networks in the Physical World with 3D Adversarial Objects*, 2017, https://www.labsix.org/physical-objects-that-fool-neural-nets/

6. DeepDream, https://en.wikipedia.org/wiki/DeepDream

7. #puppyslug hashtag on Twitter, https://twitter.com/hashtag/puppyslug

8. See, for example: Johan Nordberg, *Inside an Artificial Brain*, 2017, https://vimeo.com/132700334; Gene Kogan, *Submission for NIPS 2017*, 2017, https://vimeo.com/246047871

9. DeepArt, https://deepart.io/

10. For a recent proof of concept, see: Sebastian Lütgert, *Angelina (Neural Style)*, 2016, https://rolux.org/video/angelina/about/

inevitably be limited to modifying a few parameters of exisiting demos or tutorials. And without an often prohibitively expensive graphics processing unit, training or visualizing a neural network on consumer hardware remains an extremely slow process that doesn't lend itself well to artistic appoaches like experimental exploration or quick "sketching."[11] Most importantly though, what to do with neural networks – how to foreground the relational aspect of images no longer primarily defined by a singular type of form or content, how to render visible the full historical trajectory from early cinema and experimental video to artificial visual intelligence, and how to at least sightly demystify the domain of machine vision as part the process – remains a wide open question.

2. Project Outline

The proposed project – *The Neural Image: An Atlas of Computer Vision* – suggests a horizontal, enumerative and to some degree encyclopedic approach to neural networks (hence the concept of the "atlas"). The core idea is to exhaustively explore one or more image classification networks (like Google's InceptionV1[12] or the MIT's Places365[13]) layer-by-layer and channel-by-channel, as the role of individual neurons in the network evolves from edge and texture detection to pattern and object recognition.[14] Early layers tend to produce abstract geometric structures, often reminiscent of minimalist cinema and video works, while later layers usually feature aspects of real-world objects, sometimes in unexpected combinations and variations, evoking the visual vocabulary of experimental digital animation. The hope is that by employing a number of rather classical cinematic techniques – like parallel montage, but also simple zooms, pans and transitions – one can make these strange images look even stranger, and gradually shift the focus from the visual similarities with 20th century imagery, as exhibited by single neurons, to the complex relational architecture of the neural network itself.

The plan is to utilize several new visualization techniques, based on recent scientific papers about image parameterization and compositional neural networks.[15] In essence, these methods offer more degrees of freedom during the visualization process, can be adapted to produce a variety of different, largely unseen visual "styles," allow for renderings in arbitrarily high resolutions (including 4K video) and can be extended to communicate with other, more traditional algorithms for image manipulation. The visualization demos that accompany this proposal as a proof of concept[16] have been created with these techniques, and even though they stick relatively closely to a small set of suggested algorithms and only cover a tiny part of the InceptionV1 neural network, leaving a lot of room for further exploration, they already offer at least a glimpse of the potential of the proposed methodology.

The primary output of the project would be one or more multi-channel video installations and a series of digital prints. In addition, the nature of the material would lend itself well to the production of a digital catalogue, incorporating a larger number of stills, along

11. The 2 minutes of video linked in footnote 10 took over a month to render.

12. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, *Going Deeper with Convolutions*, 2014, https://arxiv.org/abs/1409.4842

13. Bolei Zhou, Agata Lapedriza, Aditya Khosla, Antonio Torralba, Aude Oliva, *Places*, 2017, http://places2.csail.mit.edu/

14. For a visual overview of InceptionV1's core layers, see: https://distill.pub/2017/feature-visualization/appendix/

15. Most notably: Alexander Mordvintsev, Nicola Pezzotti, Ludwig Schubert, Chris Olah, *Differentiable Image Parameterizations: A powerful, under-explored tool for neural network visualizations and art*, 2018, https://distill.pub/2018/differentiable-parameterizations/

16. Sebastian Lütgert, *Neural Network Visualization Demos*, 2019, https://rolux.org/video/nn

with the source code developed to render them. It may also turn out that there are subsets of the material that deserve further attention – be it in form of an “annotated encyclopedia of puppy slug”[17] or as a collection of fictional architectural elements, impossible landscape photography or yet-to-be-invented painting styles.

Needless to say, making art by way of machine learning consists, to a large extent, of learning by doing, and so it is entirely possible that in the course of the research and development phases of the project, new avenues open up that point beyond the outline of this proposal. Ideally, these earlier project phases would be accompanied by one or two public lectures or workshops – as an occasion to share concrete ideas and intermediate results, but also in order to develop a bit more of a shared vocabulary that would describe, critique or question the technological setup and its visual output. Even though most work in the field is carried out almost entirely with Open Source software,[18] neural network visualization has remained somewhat of a “dark art,” due to the fact that what works and what doesn’t – both technically and aesthetically – is hard to derive by means of theoretical reasoning, and usually arises from extensive practical experimentation. In this sense, this proposal is not just one to deliver one or more pieces of media art, but also a proposition to open up the process of artistic creation to an interested audience.

3. Possible Extensions

Almost all media art made with neural networks remains derivative, in the sense that it reproduces, however indirectly, the taxonomies that image classification systems are trained to match (like ImageNet[19] or the MIT Places Database[20]). In the context of the proposed project, it would be tempting to train a separate neural network on a custom corpus of images, and towards a distinctively different taxonomy.

The 16,000 movies collected on 0xDB[21] provide a corpus of around 1.6 billion single frames, with metadata and time-based information in form of subtitles, waiting to be analyzed by means of machine learning. Obviously, the nature of these images suggests new types of classification objectives, given that the visual vocabulary of cinema adheres to its own set of rules and exceptions,[22] and is much less concerned with the common targets of image recognition (like animals, plants, household objects, etc).[23]

The feasibility of such an undertaking remains to be determined. Proper training of a neural network requires a great amount of methodological rigor and, potentially, additional technical resources. Also, it may well turn out that the tagged or taggable portion of 0xDB is still too small to produce meaningful results. Still, this is definitely one of the areas to be examined more closely during the research phase of the project, and could evolve into another series of works – either as an extension of the project into Spring 2020, or as an independent follow-up.

17. Puppy slug (also see footnote 7): the frequent appearance of animal faces in artistic works that visualize neural networks – due to the fact that different species of dogs and birds take up an overproportional share of the vocabulary that the most common neural networks are trained to detect. See Sebastian Lütgert, *Softmax0 Pre Activation/Matmul*, 2018, https://pad.ma/documents/ANT

18. For example: TensorFlow, https://www.tensorflow.org/; Lucid, https://github.com/tensorflow/lucid

19. ImageNet, http://www.image-net.org/

20. MIT Places Database, http://places.csail.mit.edu/

21. Jan Gerber, Sebastian Lütgert, *0xDB*, https://0xdb.org

22. Again, Gilles Deleuze’s attempt of a classification of cinematic images comes to mind. See Deleuze 1983/1985

23. Of course, there is also a *grammar* of cinema – but as of today, there are no readily available neural networks that could easily be trained to classify time-based visual patterns.

4. Project Timeline

July-August 2019: Research

Reading scientific papers, setting up a workstation, experimenting with visualization frameworks and algorithms, exploring a number of matrix math and machine learning related python libraries, consulting with a few like-minded digital media practitioners.

September-October 2019: Development

Extending existing and developing new rendering toolchains, iterating over one or more neural networks and through several visualization algorithms, analyzing and sorting the output.

November-December 2019: Rendering

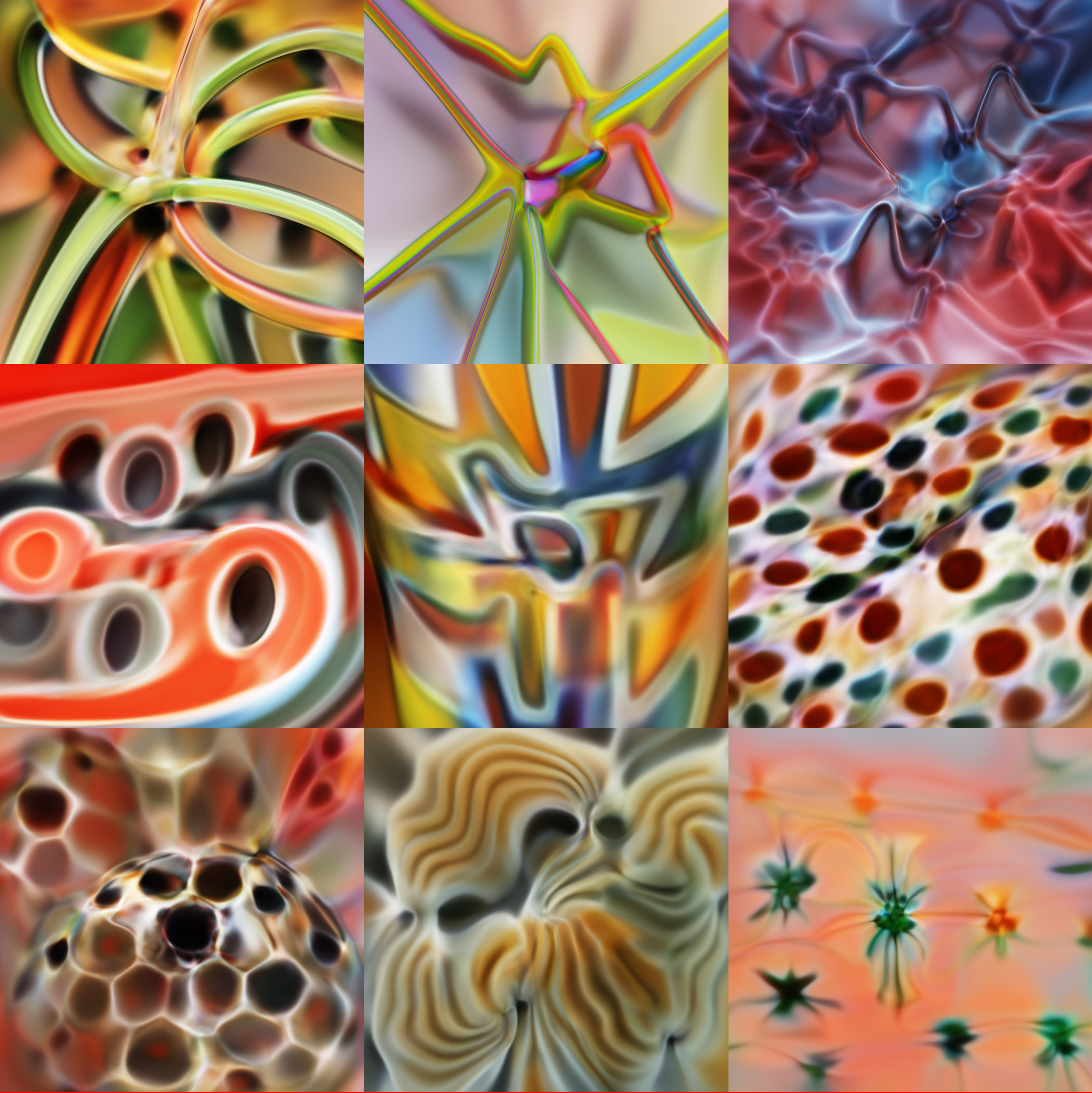
Setting up the final rendering architecture, narrowing down on the most promising sources for video and image output, letting the machines sing their songs.

5. Technical Requirements and Budget

GPU Workstation

Intel i7 CPU & CPU Cooler	EUR	600
Nvidia GeForce RTX 2080 Ti GPU	EUR	1,500
16GB DDR4 DRAM, WD Blue 1TB SSD	EUR	300
Corsair Carbide Case & Power Supply	EUR	100
Seagate 5TB HD, Monitor, Keyboard, Mouse	EUR	500
Subtotal	EUR	3,000
Electricity costs (4 months of rendering)	EUR	300
Printing costs (5 to 10 large digital prints)	EUR	800
Consulting fees (Jan Gerber, Robert M. Ochshorn)	EUR	500
Total	EUR	4,600

The above costs can be fully covered through the project grant, without the need for third-party funding.



sebastian lütgert

january 2019