

# GANlapse Generative Photography

Simon Colton<sup>1,2</sup> and Blanca Pérez Ferrer<sup>3</sup>

<sup>1</sup> SensiLab, Faculty of IT, Monash University, Melbourne, Australia

<sup>2</sup> Game AI Group, EECS, Queen Mary University of London, UK

<sup>3</sup> Etopia Center for Arts and Technology, Zaragoza, Spain

## Abstract

We describe the incorporation of text-to-image generative deep learning techniques into an art practice for making video pieces akin to time-lapse photography. We show that the process can be suitably controlled to find a latent vector able to generate an appropriate image, construct nearby vectors for similar images and interpolate between them to produce video pieces. We describe the process, how this fits into the GAN-art movement, and the cultural impact of this work in terms of an online and physical art exhibition in the Etopia arts and technology centre in Spain.

## Introduction and Background

A generative adversarial network (GAN) is a pre-trained neural model able to generate content like images, audio or text. Training a GAN is done in conjunction with a critic model which rates how realistic the generated content is when compared to a dataset of target content (Goodfellow et al. 2014). The two networks start off being poor at their tasks, but force each other to improve until the generator can produce fake content that can (often) be passed off as real. GANs take a vector of floats known as a *latent vector* as input, propagate this forward through the network and produce numerical outputs interpretable as content, e.g., an image. Conditional GANs also take a class vector which enables them to produce different types of content. For instance, the GAN employed in the research described here is called *BigGAN* (Brock, Donahue, and Simonyan 2019), and normally takes a 1-hot vector which dictates the ImageNet (Deng et al. 2009) category to generate an image for.

There has been an explosion of GAN-generated visual art in the last five years, as digital artists have adopted the technique due to the quality, variety and often surprising nature of the output from GANs. As an indication of the impact GAN art is having, major auction house Sothebys held a recent sale<sup>1</sup> which included GAN-generated pieces from Anna Ridler and Mario Klingemann. GAN artists produce images and videos by using pre-trained models or training their own, then searching for latent vectors, often randomly or using non-standard techniques described as *active divergence* in (Berns and Colton 2020). To the best of our knowledge, GAN-generated art has yet to be formally studied or

categorised. The variety and quality of imagery has increased greatly over the years, but there are some commonalities in the majority of GAN-generated images, namely an abstract, dream-like quality in which the content doesn't bear scrutiny, i.e., is not particularly coherent on close inspection. Alternatively, GANs can produce photo-realistic images like faces (Karras et al. 2021), and these have also been used in artistic settings.

A new approach to finding latent vectors for GANs has been developed by members of an online tech/art community recently, sharing their code through Colab notebooks (Bisong 2019). The approach employs the CLIP pair of pre-trained models (Radford et al. 2021), which both calculate a latent vector encoding of a given input. For the first CLIP model, the input is a piece of text, and for the second, it is an image. CLIP has been trained so that the latent encoding of pairs (image, text) where the text appropriately describes the image – or vice versa, where the image is an appropriate response to the text – have a lower cosine distance than pairs where the match is not appropriate. CLIP was trained on 400 million (text, image) pairs scraped from the internet, and has captured a broad and deep understanding of the correlation between text and images.

The task of text-to-image generation (Agnese et al. 2019) involves producing images to somehow reflect a given text prompt  $T$ . One approach involves calculating a score for the appropriateness of a generated image,  $I$ , in terms of the cosine distance between the CLIP encodings for  $T$  and  $I$ . Hence, a search can be undertaken to find a latent vector input to a GAN which produces an image  $I$  with a CLIP encoding as close as possible to the encoding of  $T$ . Note that it is also possible to search for a generated image to match a given image or some combination of text and image. Various search techniques have been tried in the community, with the most successful being backpropagation, i.e., starting with a random latent vector and training it by propagating the value of a loss function based on CLIP. Such an approach was implemented into a Colab notebook called *The Big Sleep* by Ryan Murdock<sup>2</sup>, which uses CLIP to search for latent vector inputs to BigGAN.

We have found the Big Sleep technique remarkably good for a range of text prompts, ranging from specific content to vague moods, as described in (Colton et al. 2019), where de-

<sup>1</sup>[sothebys.com/en/buy/auction/2021/natively-digital-a-curated-nft-sale-2](https://sothebys.com/en/buy/auction/2021/natively-digital-a-curated-nft-sale-2)

<sup>2</sup>[colab.research.google.com/drive/1NCceX2mbiKOSlAd\\_o7IU7nA9UskKN5WR](https://colab.research.google.com/drive/1NCceX2mbiKOSlAd_o7IU7nA9UskKN5WR)

tails of a modified version of the approach we implemented are given. For our purposes here, we need to know that the process can be parameterised by (a) the text prompt (b) the starting point which can be either a random or a given latent vector and (c) the learning rate for optimising the latent vector. We describe below how we have used the text-to-image technique to produce examples of a novel type of generative art form we call GANlapse videos. We describe the process in terms of the usage of text-to-image generation, GAN interpolation and video post-production. Following this, we describe an art exhibition of generated still images and GANlapse videos at the Etopia Center for Arts and Technology. We conclude with some thoughts on GAN art, how CLIP can bring further autonomy to GAN generation, and future research and art projects we have planned.

## GANlapse Video Generation

We define a *GANlapse* video as akin to a time-lapse video but with the still images comprising it produced by a generative adversarial network rather than taken with a camera in a real-life scenario. By *akin* to a time-lapse, we mean that (a) the subject material is similar to that normally in time-lapse videos, i.e., looking as if it comes from reality, rather than an abstract or more artistic rendering and (b) the perceived time period covered is similar, for instance a construction site over the duration of a year, a skyline over the period of a day or the seasonal life and death of a plant, etc.

The production of GANlapse videos is in three parts: firstly a series of images is generated to reflect a given text prompt and then specialised versions of some of these images are further generated as keyframes for a video; secondly, multiple series of images animating the traversal from one image to another are generated, using interpolation over the keyframe latent vectors; and thirdly, a set of traversal image sequences is collated and post-processed into final video pieces. In this manner, we have produced 30 different architecture GANlapse videos and 3 featuring flora and fauna.

## Text to Image Generation

To start each GANlapse project, we chose a particular subject material which might traditionally have been the subject of a standard time-lapse photography piece, and then derived a text prompt with which to generate images. For instance, in one project, we employed the **content text prompt** “A beautiful modernist building in the countryside” and used the modified CLIP-guided BigGAN process mentioned above to generate 200 images, using four GPUs for around 2 hours, with a learning rate of 0.09 and a random latent vector as starting point. Some experimentation with the text prompt was required to produce suitable images for each project, e.g., replacing “beautiful” with “serene”.

We cherry-picked some suitably high-quality images and for each, we retrieved the BigGAN latent vector pair,  $l$ , responsible for its output. For each image, we then produced a series of *keyframe* images by running the text-to-image process again, but starting with  $l$  rather than randomly, using a learning rate of 0.01, so the image produced doesn’t stray too far from the original. Continuing the modernist build-



Figure 1: Keyframe images for (a) a modernist building project and (b) a flower life-cycle project.

ing example, we used the following **modifier prompts** to produce keyframes for a four-seasons GANlapse project:

- 
- (i) “Covered in snow” (ii) “Spring flowers”
  - (iii) “Summer sunshine” (iv) “Autumn leaves”
- 

Again, some experimentation with text prompts was required to find suitable keyframes for each project, and often we employed multiple small-yield sessions with slightly different prompts in parallel, to harvest good results.

As described below, the keyframe images become the basis for the interpolation videos which are ultimately the raw art material for the final productions. In other projects, the keyframe images were not produced by fine-tuning an image starting with its latent vector and using modifying prompts. Instead, multiple different content prompts were used, e.g., in a project to produce a GANlapse video of the life-cycle of flowers, we used seven content prompts as follows:

- 
- (i) “Foliage” (ii) “Beautiful small yellow flowers”
  - (iii) “Large yellow flowers” (iv) “Dying yellow flowers”
  - (v) “Dead yellow flowers” (vi) “Dead foliage”
- 

Once the keyframe images were generated and chosen, in order to produce suitable videos, we ordered them as a time-lapse storyboard, e.g., a building’s surroundings changing over four seasons, or a flower dying and being re-born. Example keyframe image orderings for both a modernist building and flower life-cycle project are given in figure 1

## GAN Interpolations

GAN interpolation is a common technique used for *inbetweening* two keyframes  $k_1$  and  $k_2$  to produce suitable intermediate frames. Here, a sequence of images is produced by generating an GAN image with latent vector pair  $((1-t) \times l_1) + (t \times l_2)$  as  $t$  varies via some increment from 0 to 1, where  $l_1$  and  $l_2$  are the latent vector pairs responsible for producing  $k_1$  and  $k_2$ . Such interpolation between keyframe images produces pleasing and tranquil animations similar to traditional time-lapse pieces, when the generated images are compiled into a video. We experimented with in-out easing functions to slow down the movement towards and away from the keyframes to make the video subtly smoother. Such easing involves applying a function to  $t$  before employing it to calculate the latent vector pair for the GAN, e.g., substituting  $t$  with  $16t^5$  for quintic easing.

We adopted the practice of making the last in the sequence the same as the first, so videos can loop, and we experimented with randomly perturbing the latent vector by small amounts (up to 0.04 standard deviations away from the original) before generating the images. For the architecture videos, such perturbations produced a jittery effect commonly seen in traditional time-lapse photography, due to slight camera movements. In the case of the flowers, this effect also gave an impression of them being delicately blown in the wind. We also found that pausing on a keyframe for a number of frames worked well with perturbation for the flowers projects. While maintaining 30 frames per second for smooth videos, we experimented with the number of frames between keyframes, in order for change to be visible at all times, yet the tranquility of the pieces to be upheld.

When inbetweening the second type of keyframe sequences, namely where a set of image generated from different content prompts are used, we had two difficulties. Firstly, the distance between the latent vectors was often quite large and so the visual change from one intermediate frame to another became disorientating. Secondly, the interpolation points occasionally went near to a one-hot class vector, which generated an image in the original ImageNet categories that BigGAN was trained on, usually depicting a dog or a bird, which ruined the animation. To solve both problems, we simply calculated in advance the cosine distance between the BigGAN latent vectors of two potential keyframes and avoided using any pairs where this distance was significantly larger than the average. For each project, we varied the usage of easing, perturbation and pausing and were able to produce (subjectively) aesthetically pleasing time-lapse style animations of between 30 and 60 seconds, which acted as the art materials for the final stage.

### Video Post-Processing

Due to the nature of the processing, GAN-generated images tend to be square, which is rarely the case for traditional time-lapse pieces. Moreover, as described below, we were commissioned to produce GANlapse videos for an exhibition involving screens with a 16:9 rectangular aspect ratio. For these reasons, and also to be a little more distinctive in the GAN-art world, we post-processed the videos to both change their shapes and make them more sophisticated. In particular, with the architecture images, we noticed that often the buildings depicted were cut by the edge of the image. This gave us the opportunity to mirror the images down the horizontal and produce a widescreen video, as depicted in figure 2(a) for the original architecture image of figure 1(a). For others, rather than mirroring the entire image, we mirrored one or both sides slightly in order to produce a wider, rectangular aspect. We also cropped where appropriate.

For two other projects, we collaged multiple interpolation videos into a more sophisticated piece. For instance, for a piece depicting the life-cycle of wildflowers called *Les Fleurs de Vie*, we combined five interpolation videos into a single one, as shown in figure 2(c). Each of the sub-windows of the piece provides a portal onto a different coloured flower at a different stage during its life-cycle. To complete the piece, we added a glassy border and table-top

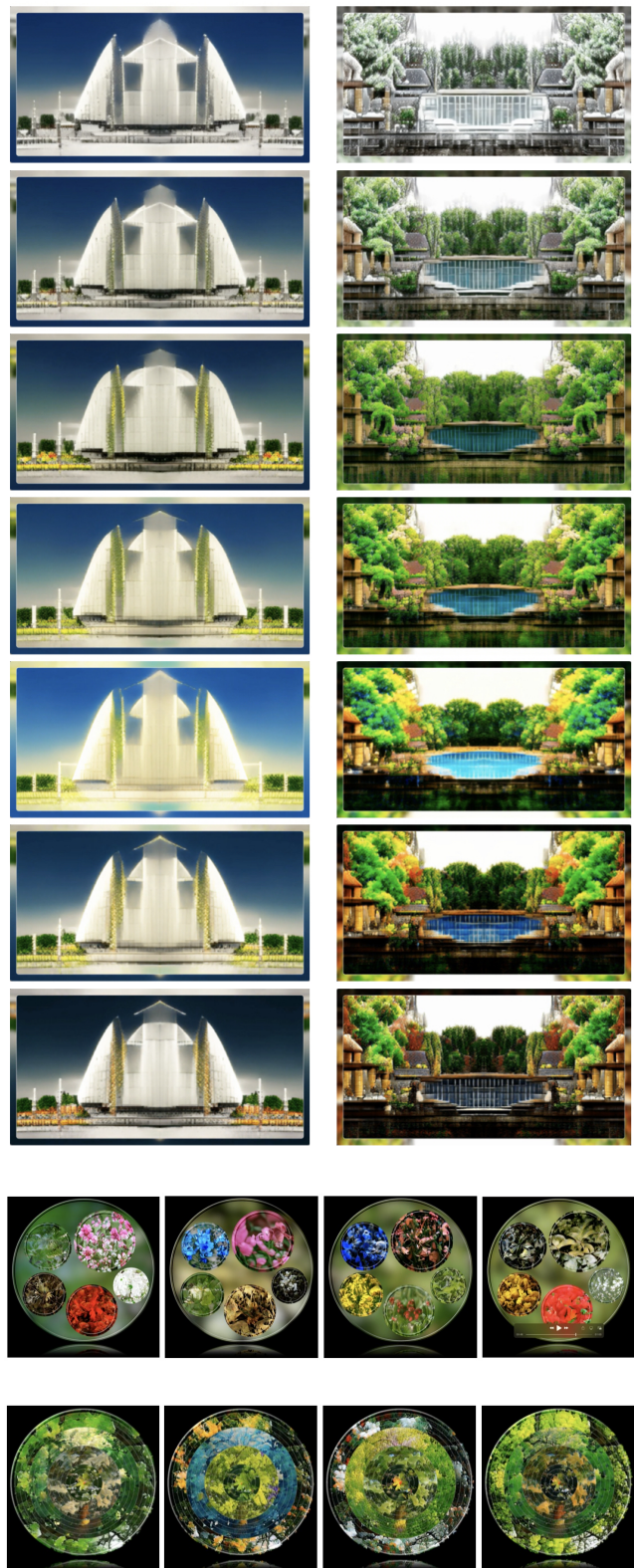


Figure 2: Still images from: the *Arquitecturas Imaginadas* GANlapse pieces (a) *The Museum* and (b) *The Glasshouse*. Still images from the (c) *Fleurs de Vie* and (d) *Mighty Oak* GANlapse video pieces, published on the Hic et Nunc NFT platform ([hicetnunc.xyz/simoncolton](http://hicetnunc.xyz/simoncolton)).

reflection effect. For another piece, entitled *Mighty Oak*, we again combined four different interpolation videos depicting different views of foliage during different seasons, again adding a border and reflection. For the post-production, we used our own image manipulation software and the ffmpeg software package ([ffmpeg.org](http://ffmpeg.org)) which, among many other things, can combine still images into MP4 videos. Still images from some of the GANlapse videos are given in figure 2, highlighting the mirroring, collaging, bordering and reflection techniques employed.

## Arquitecturas Imaginadas

The New European Bauhaus<sup>3</sup> is a cross-cultural initiative providing a platform for innovative thinking around living spaces, sustainable living and quality of life. One of the participating organisations is the Etopia Center for Art and Technology in Zaragoza, Spain. We were commissioned by Etopia to create an exhibition entitled “Arquitecturas Imaginadas” (Imagined Architectures) as part of the New European Bauhaus initiative. The first author of this paper provided the images/videos for the exhibition, for which the second author (working at Etopia) is the curator and mediator. The online version<sup>4</sup> of the exhibition runs from 1st June 2021 to October 2021 and comprises 13 GANlapse videos and five sets of 6 still images, as before generated via quotes from well-known female architects: Julia Morgan, Christine Lam, Pascale Sablan, Zaha Hadid and Marian Kamara. Screenshots of the online exhibition are given in figure 3.

The physical exhibition started on June 30th 2021 and runs until October 2021. It comprises the following:

- 16 printed images generated using the CLIP-guided BigGAN process above, using as text prompts quotations about architecture from 4 female architects.
- 21 different GANlapse videos split over four 24in screens, two 43in screens and a 2m by 1.5m video wall.<sup>5</sup>
- 27 different GANlapse videos on a two-wall 8m by 5m exterior media facade.

Photographs from the physical exhibition are given in figure 4, along with photographs from the media facade. 30 full resolution GANlapse videos, 54 still images and a blogpost further describing the making of the videos from the exhibition are provided here:

[imaginative.ai/wp/imagined-architectures](http://imaginative.ai/wp/imagined-architectures)

## Conclusions and Future Work

GAN artists have successfully cultivated their moderately abstract, dream-like aesthetic and promoted the process of serendipitous, often random usage of generative processes (Berns and Colton 2020). They regularly produce beautiful artworks which impact the art world. The work and resulting exhibition described here has been an attempt to broaden the usage of GANs in art beyond the current aesthetic. In particular, we wanted to see if it was possible to

<sup>3</sup>[europa.eu/new-european-bauhaus](http://europa.eu/new-european-bauhaus)

<sup>4</sup>[estoyenetopia.es/arquitecturas-imaginadas](http://estoyenetopia.es/arquitecturas-imaginadas)

<sup>5</sup>See YouTube channel: UCfCzCwiYAWBP aOYWTwbmDCA



Figure 3: Screenshots of the online version of *Arquitecturas Imaginadas*, featuring still imagery and GANlapse videos.

make pieces which are slower, more contemplative and traditionally beautiful in a representative way, than either the fast-moving, arresting, disturbing and often grotesque imagery or the moody, difficult-to-perceive dream-like images often afforded by GANs. We believe we have achieved this aim: while the architecture and floral videos are not entirely representational, they do bare scrutiny and look as if they may have been based on photographs rather than generated.

Artists and AI practitioners are increasingly using CLIP-guided GAN image generation in their projects, as exemplified by (Smith and Colton 2021). However, few GAN artists have so far shown much interest in software having creative responsibilities from a computational creativity perspective. However, the usage of CLIP as described above has provided an opportunity for a number of GAN artists to enable the automated guidance of image generation, and we are already seeing interesting results emerging from artists such as Mario Klingemann. In future work, we plan to automate text prompt engineering, image and keyframe cherry-picking (possibly using evolutionary techniques for latent-space search, as per (Fernandes et al. 2009)) and time-lapse story construction elements of the GANlapse production process. We also aim to branch out with more imaginative time-lapse scenarios, possibly employing automated fictional ideation such as in (Llano et al. 2016). Ultimately, we aim to bring computational creativity techniques and methodologies to bear on generative deep learning to drive forward truly autonomous machine creativity in the arts.

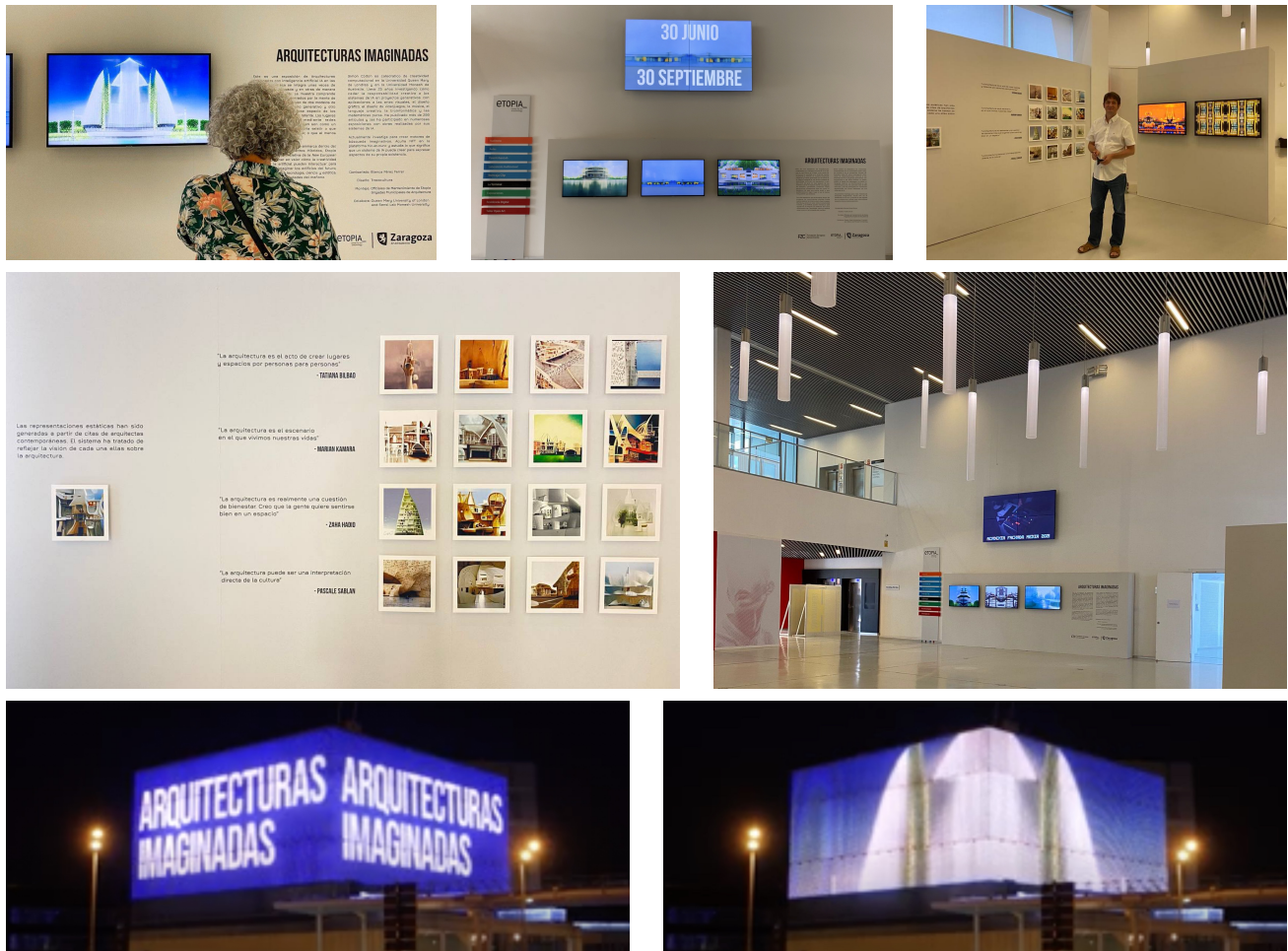


Figure 4: Physical installation of *Arquitecturas Imaginadas*, inside the Etopia centre and on the media facade outside it.

## Acknowledgements

We would like to thank the anonymous reviewers for their insightful comments. We would also like to thank the team at Etopia for their hard work and expertise in staging the *Arquitecturas Imaginadas* exhibition.

## References

- Agnese, J., Herrera, J., Tao, H. and Zhu, X. 2019. A survey and taxonomy of adversarial neural networks for text-to-image synthesis. *arXiv:1910.09399*.
- Berns, S. and Colton, S. 2020. Bridging generative deep learning and computational creativity. In *Proc. ICCV*.
- Bisong, E. 2019. Google collaborator. In *Building Deep Learning Models on Google Cloud Platform*. Springer.
- Brock, A., Donahue, J. and Simonyan, K. 2019. Large scale GAN training for high fidelity natural images synthesis. *arXiv:1809.11096*.
- Colton, S., Smith, A., Berns, S., Murdock, R. and Cook, S. 2021. Generative Search Engines: Initial Experiments. In *Proc. ICCV*.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. In *Proc. IEEE conference on CVPR*.
- Fernandes, P., Correia, J. and Machado, P. 2020. Evolutionary Latent Space Exploration of Generative Adversarial Networks. In *Proc. 9th International Conference on AI in Music, Sound, Art and Design*.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Ward-Farley, D., Ozair, S., Courville, A. and Bengio, Y. 2014. Generative adversarial networks. In *Proc. NIPS*.
- Karras, T., Laine, S. and Aila, T. 2018. A style-based generator architecture for generative adversarial networks. *arXiv:1812.04948*.
- Llano, T., Colton, S., Hepworth, R. and Gow, J. 2016. Automated fictional ideation via knowledge base manipulation. *Cognitive Computation* 8(2):153–174.
- Radford, A. et. al. 2021. Learning transferable visual models from natural language supervision. *arXiv:2103.00020*.
- Smith, A. and Colton, S. 2021. CLIP-Guided GAN Image Generation: An Artistic Exploration. In *Proc. 10th International Conference on AI in Music, Sound, Art and Design*.