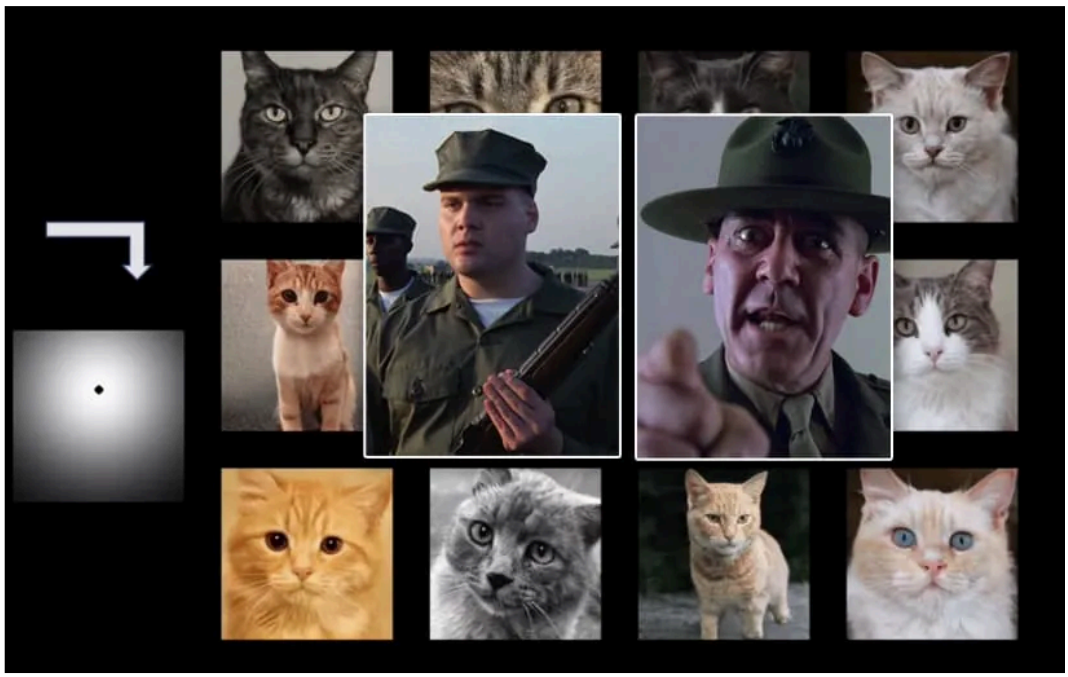
The Unintended Benefit of Mapping a GAN's Latent Space

Originally published at https://www.unite.ai/the-unintended-benefit-of-mapping-a-gans-latent-space/



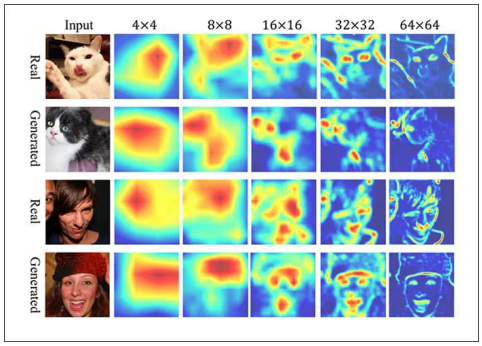
While trying to improve the quality and fidelity of AI-generated images, a group of researchers from China and Australia have inadvertently discovered a method to interactively control the latent space of a Generative Adversarial Network (GAN) – the mysterious calculative matrix behind the new wave of image synthesis techniques that are set to revolutionize movies, gaming, and social media, and many other sectors in entertainment and research.

Their discovery, a by-product of the project's central goal, allows a user to arbitrarily and interactively explore a GAN's latent space with a mouse, as if scrubbing through a video, or leafing through a book.

CLICK TO VIEW ANIMATED GIF

An excerpt from the researchers' accompanying video (see embed at end of article for many more examples). Note that the user is manipulating the transformations with a 'grab' cursor (top left).
Source: https://www.youtube.com/watch?v=k7sG4XY5rIc

The method uses 'heat maps' to indicate which areas of an image should be improved as the GAN runs through the same dataset thousands (or hundreds of thousands) of times. The heat maps are intended to improve image quality by telling the GAN where it's going wrong, so that its next attempt will be better; but, coincidentally, this also provides a 'map' of the entire latent space that can be browsed by moving a mouse.



Spatial visual attention emphasized via GradCAM, which indicates areas that need attention by imposing bright colors. Source: https://arxiv.org/pdf/2112.00718.pdf

The paper is called *Improving GAN Equilibrium by Raising Spatial Awareness,* and comes from researchers at the Chinese University of Hong Kong and the Australian National University. In addition to the paper, video and other material can be found at the project page.

The work is nascent, and currently limited to low resolution imagery (256×256), but is a proof of concept that promises to break open the 'black box' of the latent space, and comes at a time when multiple research projects are hammering at that door in pursuit of greater control over image synthesis.

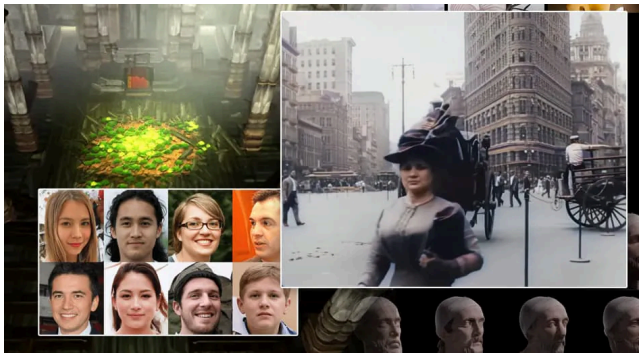
Though such images are engaging (and you can see more of them, in better resolution, in the video embedded at the end of this article), what's perhaps more significant is that the project has found a way to create improved image quality, and potentially to do it faster, by telling the GAN specifically where it's going wrong during the training.

But, as *Adversarial* indicates, a GAN is not a single entity, but instead an unequal conflict between authority and drudgery. To understand what improvements the researchers have made in this respect, let's look at how this war has been characterized until now.

The Piteous Plight of the Generator

If you've ever been haunted by the thought that some great new item of clothing you bought was produced in a sweatshop in an exploited country, or had a boss or client that kept telling you to 'Do it again!' without ever telling you what was wrong with your latest attempt, spare a mite of pity for the *Generator* part of a Generative Adversarial Network.

The Generator is the workhorse that has been delighting you for the past five or so years by helping GANs create photorealistic people that don't exist, upscale old video games to 4k resolution, and turn century-old footage into full-color HD output at 60fps, among other wondrous AI novelties.



From creating photoreal faces of unreal people to restoring ancient footage and revivifying archive video games, GAN has been busy in the last few years.

The Generator runs through all the training data again and again (such as pictures of faces, in order to make a GAN that can create photos of random, non-existent people), one photo at a time, for days, or even weeks, until it is able to create images that are as convincing as the genuine photos that it studied.

So how does the Generator know that it is making any progress, each time it tries to create an image that's better than its previous attempt?

The Generator has a boss from hell.

The Merciless Opacity of the Discriminator

The job of the *Discriminator* is to tell the Generator that it didn't do well enough in creating an image that's authentic to the original data, and to *Do it again*. The Discriminator doesn't tell the Generator *what* was wrong with the Generator's last attempt; it just takes a private look at it, compares the generated image to the source images (again, privately), and assigns the image a score.

The score is *never* good enough. The Discriminator won't stop saying *'Do it again'* until the research scientists turn it off (when they judge that the additional training will not improve the output any further).

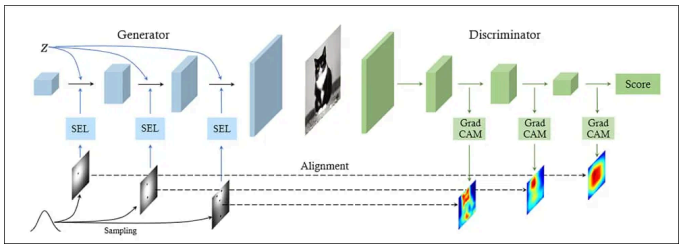
In this way, absent any constructive criticism, and armed only with a score whose metric is a mystery, the Generator must randomly guess which parts or aspects of the image caused a higher score than before. This will lead it down many further unsatisfactory routes before it changes something positively enough to get a higher score.

The Discriminator as Tutor and Mentor

The innovation provided by the new research is essentially that the Discriminator now indicates to the Generator *which parts of the image were unsatisfactory*, so that the Generator can focus on those areas in its next iteration, and not throw away the sections that were rated higher. The nature of the relationship has turned from combative to collaborative.

To remedy the disparity of insight between the Discriminator and the Generator, the researchers used GradCAM as a mechanism capable of formulating the Discriminator's insights into a visual feedback aid for the Generator's next attempt.

The new 'equilibrium' training method is called EqGAN. For maximum reproducibility, the researchers incorporated existing techniques and methods at default settings, including the use of the StyleGan2 architecture.



The architecture of EqGAN. The spatial encoding of the Generator is aligned to the spatial awareness of the Discriminator, with random samples of spatial heatmaps (see earlier image) encoded back into the generator via the spatial encoding layer (SEL). GradCAM is the mechanism by which the Discriminator's attention maps are made available to the generator.

GradCAM produces heatmaps (see above images) that reflect the Discriminator's criticism of the latest iteration, and make this available to the Generator.

Once the model is trained, the mapping remains as an artifact of this cooperative process, but can also be used to explore the final latent code in the interactive way demonstrated in the researchers' project video (see below).

EqGAN

The project used a number of popular datasets, including the LSUN Cat and Churches datasets, as well as the FFHQ dataset. The video below also features examples of facial and feline manipulation using EqGAN.

All images were resized to 256×256 prior to training EqGAN on the official implementation of StyleGAN2. The model was trained at a batch size of 64 over 8 GPUs until the Discriminator had been exposed to over 25 million images.

Testing the results of the system across selected samples with Frechet Inception Distance (FID), the authors established a metric called Disequilibrium Indicator (DI) – the degree to which the Discriminator retains its knowledge advantage over the Generator, with the objective of narrowing that gap.

Over the three datasets trained, the new metric showed a useful drop after encoding spatial awareness into the Generator, with improved equilibrium demonstrated by both FID and DI.

Method	Cat [29] 256 × 256		FFHQ [17] 256 × 256		Church [17] 256 × 256	
	FID ↓	DI ↓	FID ↓	DI ↓	FID ↓	DI ↓
Baseline	8.36	3.64	3.66	1.62	3.73	3.01
+ SEL	7.82	3.12	3.39	1.38	3.55	2.59
+ $\mathcal{L}_{align}$	6.81	2.39	2.96	0.73	3.11	2.07

The researchers conclude:

> *'We hope this work can inspire more works of revisiting the GAN equilibrium and develop more novel methods to improve the image synthesis quality through maneuvering the GAN equilibrium. We will also conduct more theoretical investigation on this issue in the future work.'*

And continue:

> *'Qualitative results show that our method successfully [forces the Generator] to concentrate on specific regions. Experiments on various datasets validate that our method mitigates the disequilibrium in GAN training and substantially improves the overall image synthesis quality. The resulting model with spatial awareness also enables the interactive manipulation of the output image.'*

Take a look at the video below for more details about the project, and further examples of dynamic and interactive exploration of the latent space in a GAN.

Improving GAN Equilibrium by Raising Spatial Awareness

11:12am 4th Dec 2021 – Corrected URL for GradCAM and tidied up surrounding reference.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai

Discover More